THÈSE DE DOCTORAT DE
l'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité

Informatique

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

Dong Quan VU

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse:

Models and Solutions of Strategic Resource Allocation Problems: Approximate Equilibrium and Online Learning in Blotto Games

soutenue le 25 juin 2020

après avis de :

M. Jason M. Marden	Rapporteur
M. Vianney Perchet	Rapporteur

devant le jury composé de :

M. Patrick Loiseau	Directeur de thèse
M. Alonso Silva	Co-encadrant
M. Vianney Perchet	Rapporteur
M. Nicolò Cesa-Bianchi	Examinateur
Mme. Johanne Cohen	Examinatrice
M. Christoph Dürr	Examinateur
M. Rida Laraki	Examinateur

[This page intentionally left blank]

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms and Notations

	Acronyms	
APA	$\triangleq$	all-pay auction,
CB	$\triangleq$	Colonel Blotto game,
CDF	$\triangleq$	cumulative distribution function
CSF	$\triangleq$	contest success function,
Exp3	$\triangleq$	Exponential-weight algorithm for Exploration and Exploitation (Algorithm)
F-APA	$\triangleq$	all-pay auction with favoritism,
FPL	$\triangleq$	Follow-the-Perturbed-Leader (Algorithm)
GR-CB	$\triangleq$	generalized-rule Colonel Blotto game,
HS	$\triangleq$	Hide-and-Seek game
IU	$\triangleq$	independent uniform strategy,
LB	$\triangleq$	Lottery Blotto game,
MAB	$\triangleq$	Multi-armed bandit (with finitely many arms)
OComb	$\triangleq$	Online combinatorial optimization
OLO	$\triangleq$	Online learning optimization
OSMD	$\triangleq$	Online Stochastic Mirror Descent (Algorithm)
OSP	$\triangleq$	Online shortest path problem
OSPBand	$\triangleq$	Online shortest path problem with bandit feedback
SOComb	$\triangleq$	Online combinatorial optimization with side-observations
SOOSP	$\triangleq$	Online shortest path problem with side-observations
	Notation	
$\mathcal{CB}_n$	$\triangleq$	the generalized Colonel Blotto game with n battlefields,
$\mathcal{CB}_n^D$	$\triangleq$	the discrete Colonel Blotto game with n battlefields,
e	$\triangleq$	Euler's number,
$\mathbb{E}X$	$\triangleq$	expectated value of a random variable X,
F_X	$\triangleq$	cumulative density function of a random variable X,
$\mathcal{GR}-\mathcal{CB}_n$	$\triangleq$	the generalized-rule Colonel Blotto game with n battlefields,
$\mathcal{LB}_n$	$\triangleq$	the generalized Lottery Blotto game with n battlefields,
$\mathbb{M}_{m\times n}$	$\triangleq$	set of all matrices of the size $m \times n$,
$[n]$	$\triangleq$	the set $\{1, 2, \ldots, n\}$,
$\mathbb{P}(E)$	$\triangleq$	the probability that an event E happens.
$\mathbb{R}^n_{\geq 0}$	$\triangleq$	$\{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$,

$\mathcal{U}(a,b)$	$\triangleq$	the uniform distribution on $[a,b]$.
$x \sim F$	$\triangleq$	x is drawn from the random variable corresponding to the distribution F.

Other Convention

Throughout this thesis, we also often use the following convention in notation:

- For any $n \in \mathbb{N}, n \geq 1$, we denote by $[n]$ the set $\{1, 2, \ldots, n\}$.
- We use bold symbols (e.g., $\boldsymbol{x}$) to denote vectors and subscript indices to denote its coordinates, e.g., $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ or shorten by $\boldsymbol{x} = (x_i)_{i \in [n]}$; in some special cases, we also use the notation $\boldsymbol{x}(i)$ to denote the i-th coordinate of $\boldsymbol{x}$.
- We denote by $S_1 \bigtimes S_2$ the Cartesian product of sets S_1 and S_2.
- In game models, we often use the letter ϕ to denote a player and use $-\phi$ to indicate her opponent.
- We also use the asymptotic notations (Bachmann–Landau notations)[1] O, Ω by their standard definitions, i.e., for any (real-valued) functions f, g defined on an unbounded subset $S \subset \mathbb{R}^n_{>0}$, and $g(z) > 0$ for any $z \in S$, we write
  - $f(z) = O\left(g(z)\right)$ if $\exists M, C > 0 : |f(z)| \leq C g(z), \forall z \in S : z_i \geq M, \forall i \in [n]$;
  - $f(z) = \Omega\left(g(z)\right)$ if $\exists M, C > 0 : |f(z)| \geq C g(z), \forall z \in S : z_i \geq M, \forall i \in [n]$;

  Moreover, $\tilde{O}$ is a variant of O such that the logarithmic terms (in z_i) are ignored. We also write $h(z) \leq O(g(z))$ if there exists a term $f(z) = O(g(z))$ such that $h(z) \leq f(z), \forall z \in S$ (similar notations for $\tilde{O}$ and Ω can be deduced).

[1] These definitions are used by textbooks such as Brassard and Bratley (1996) and Cormen et al. (2009). Note that this definition of the Ω notation is from Knuth (1976).

Summary of the Thesis

Resource allocation problems, broadly defined as situations involving decisions on distributing a limited budget of resources in order to optimize an objective, ubiquitously appear in many real-life situations. These problems have attracted attention from a variety of scientific disciplines that study them from a wide range of perspectives. In particular, many of them involve interactions between competitive decision-makers and between them and their environment, which can be well captured by game-theoretic models that constitute the class of *resource allocation games*.

In this thesis, we choose to investigate resource allocation games. We primarily focus on the *Colonel Blotto game* (CB game)—one of the simplest and most well-known resource allocation games. In the CB game, two competitive players, each having a fixed budget of resources, simultaneously distribute their resources toward n battlefields. Each player evaluates each battlefield with a certain value. In each battlefield, the player who has the higher allocation wins and gains the corresponding value while the other loses and gains zero. Each player's payoff is her aggregate gains from all the battlefields. Despite its apparent simplicity and long-standing history, there still remain crucial and fundamental open questions in studying the CB game (and other resource allocation games sharing its basic structure).

We examine resource allocation games under the light of two perspectives coming from two different disciplines. First, *in the game-theoretic perspective*, we model them as one-shot complete-information games and analyze players' strategic behaviors. We conduct extensive analyses of several prominent variants of the CB game and their extensions in which, the leading open question is to find strategies guaranteeing good payoffs for players. Our first main contribution is a class of *approximate (Nash) equilibria* in these games for which we prove that the approximation error can be well-controlled. Moreover, we construct these approximate equilibria with simple and efficient methods; thus, these solutions are scalable and practical.

Second, we model resource allocation games as *online learning problems* to study situations involving sequential plays and incomplete information. In particular, we focus on the subclass of resource allocation games with combinatorial structures that poses interesting challenges. We exploit the particular structure of these games to make a connection between their online learning versions (with the online discrete CB game as a leading case study) and online shortest path problems. Our second main contribution is a set of novel *regret-minimization algorithms* for generic instances of online shortest path problems under several restricted feedback settings. These algorithms provide significant improvements in regret guarantees and running time in comparison with existing solutions. Finally, we apply these findings to several prominent online resource allocation games with combinatorial structures, including the online discrete CB game, showing how a player may use the structures of these games to improve performance and implementability.

Chapter 1

Introduction

1.1 Context and Motivation

1.1.1 Resource Allocation Problems and Games

Resource allocation problems, broadly defined as situations involving decisions on distributing a limited budget of resources in order to optimize an objective, are omnipresent and ubiquitous. We human beings and our society are impacted, either directly or indirectly, by such problems everyday: from mundane tasks such as personal time-management on a daily basis to more vital decisions such as rationing in wars; from applications like patrolling problems to more unnoticeable usages such as ads bidding for online search. A basic mathematical formulation of resource allocation problems is defined by Ibaraki and Katoh (1988) as follows (here, $[n]$ denotes the set $\{1, 2, \dots, n\}$):[1]

$$\begin{aligned} \min \quad & f(x_1, x_2, \dots, x_n) \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq X. \end{aligned}$$

Naturally, resource allocation problems have attracted interest from a variety of scientific disciplines that tackle them from a wide range of perspectives and with a diversity of techniques. Solutions of resource allocation models have also been implemented in real-world applications with positive results, validating the importance and relevance of this class of problems. A well-known example is the Koopman's problem in distributing search effort, introduced in the book series of Koopman (1956a,b, 1957), in which a searcher distributes his effort on a one-dimensional line (or toward n locations) trying to maximize the probability of detecting a target. Extensions of this problem are also studied in several works, e.g., by De Guenin (1961) and Stone (1976), and it is considered to be the original problem of the field of resource allocation optimization (e.g., see Ibaraki and Katoh (1988)). Another famous resource allocation

[1]Note that the objective function f in this formulation not only depend on the allocations $\{x_i\}_{i\in[n]}$ but also possibly depend on other variables and/or are subject to randomness.

problem is optimizing the portfolio selection in which one wants to choose a set of n investments (whose returns are random variables) in order to attain the maximum return (see e.g., Elton et al. (1976) and Markowitz (1952)). This problem essentially leads to the framework of portfolio theory, also known as the mean-variance theory (see e.g., El-Yaniv (1998) for a survey). Resource allocation is also one of the main focuses of the field of operational research; notably in (capacitated) lot-sizing problems (see e.g., the reviews of Bahl et al. (1987) and Karimi et al. (2003)) and supply chains (see e.g., the book by Ganesh et al. (2015)). Other notable applications of resource allocation problems include, but are not limited to, contest theory in economics and political science (see e.g., surveys of Fu and Wu (2019) and Kvasov (2007a)), scheduling and distributed manufacturing systems in operational research (see e.g., surveys of Lombardi and Milano (2012) and Tharumarajah (2001)), networking in telecommunications (see e.g., surveys of Altman et al. (2006) and Herrera and Botero (2016)) and high performance computing (see e.g., the survey of Hussain et al. (2013)).

Within the scope of resource allocation, a number of problems involve interactive situations between several competitive decision-makers, and between them and their environment. In these situations, the involved interactions are often essential for making the decisions. For instance, in (multi-prize) auctions (see e.g., Clark and Riis (1998a)), the payoff of each bidder—as a decision-maker—depends not only on her own bid but also on the decisions of other bidders and possibly on the reserve price of the auctioneer (i.e., the environment). Likewise, in security problems, the utility of a defender depends both on its actions as well as on the attackers' actions. Another type of such interaction can be found in situations requiring decision-makers to choose a sequence of actions based on their predictions about the environment (the decision-makers may not have complete information about the environment): the predictions can only be improved by repeatedly interacting with the environment and observing the outcomes. Examples include radio resources management problems where the systems' performance usually suffers unpredictable noises and online advertising problems where rewards (e.g., click-through-rate) are randomly drawn each time a customer visits a website and it is only known precisely when an advertisement campaign is run.

Among possible approaches, game theory is especially suitable to model problems involving interactions between competitive decision-makers; as Aumann (1985) famously stated: "Briefly put, game and economic theory are concerned with the interactive behavior of Homo rationalis—rational man." (p. 35). Moreover, game models can also serve as the motivation as well as a "testbed" to study the prediction problems with sequential decisions; we cite a statement from Cesa-Bianchi and Lugosi (2006) on this matter: "there exists an intimate connection between sequential prediction and some fundamental problems belonging to the theory of learning in games." (p. 180). For these reasons, in this thesis, we examine resource allocation problems that are modeled as games (but our work is not limited to game-theoretic analyses). There is a large collection of these games that are proposed and studied in the literature. However, it appears that there is currently no universal consensus on notation nor there exists a general framework for this kind of games. We commonly address them as *resource allocations games* hereinafter and present our definition of this class of games in Chapter 2. Simply put, a resource allocation game is any strategic game where players' strategies are n-tuples whose summation of components does not exceed pre-defined budgets (we call this the budget constraints). Some well-known examples of resource allocation games are the Von Neumann's hide-and-seek game (introduced by Von Neumann (1953) and extended by Flood (1972)), the Colonel Blotto game (introduced by Borel (1921)), the Tullock rent-seeking game (introduced by Friedman (1958) and Tullock (1980)), the minority game (introduced by Challet and Y.-C. Zhang (1998)). Resource allocation games have a large scope of applications, especially in economic competitions, networking planning and security problems. They also involve a large range of interesting challenges and provides a rich sets of open questions to study.

1.1.2 Colonel Blotto as a Resource Allocation Game

The Colonel Blotto game (hereinafter, the CB game) is one of the simplest and most well-known game-theoretic models in resource allocation. The CB game has a long-standing history; it was introduced by Borel (1921) which is considered (by e.g., Myerson (1991)) as one of the earliest works in modern game-theory. A general description of the CB game is given as follows:

> In the CB game, two competitive players, each having a fixed budget of resources, simultaneously distribute their resources toward n battlefields. Each player evaluates each battlefield with a certain value. In each battlefield, the player who has the higher allocation wins and gains the corresponding value while the other loses and gains zero. Each player's payoff is then her aggregate gains from all the battlefields.

The first impression that one might have about the CB game is its eccentric name and terminology—they are military-related. This is, in fact, a tradition adopted by the literature that can be traced all the way back to the fictional background story introduced by Gross and Wagner (1950): the players are two colonels, whose names are Blotto and Enemy, competing in a war campaign.[2] The term "Blotto" is also adopted to refer to a larger family of games (extended from the model of the CB game), called *Blotto games*—this terminology was first proposed by Blackett (1958) who gave a general description of a Blotto game as follows:

[2]In fact, "blotto" is an informal word used in the early 20th century to refer to drunkards (we do not know for certain if this is the reason inspiring Gross and Wagner (1950) to name in that way the colonels in their game).

In a Blotto game, players simultaneously distribute their forces toward n battlefields; if player A has an allocation plan of $(a_1, \ldots, a_n)$ (such that $\sum_{i \in [n]} a_i \leq X^A$) and player B has an allocation of $(b_1, \ldots, b_n)$ (such that $\sum_{i \in [n]} b_i \leq X^B$); then the gain of player $\phi \in \{A, B\}$ at location i is $R_i^{\phi}(a_i, b_i)$ and her total payoff is $\sum_{i \in [n]} R_i^{\phi}(a_i, b_i)$.

Intuitively, to formulate a (generic) Blotto game, one replaces the functions determining the gains that players receive in each battlefields in the CB game by a generic rule, denoted here as R_i^{ϕ}.[3] Several particular instances of Blotto games will be introduced in detail in Chapter 3 and throughout this thesis, we present results regarding these games. We note that the strategies of the players in the CB game (and other Blotto games) can be considered as n-tuples that satisfy the budget constraints (here, n is the number of battlefields); thus, the CB game (and other Blotto games) belongs to the class of resource allocation games as defined above.

Thanks to the simplicity of its framework, the CB game (and other Blotto games) possesses the elegance and generality that allow it to capture many situations in practice. We first review here several notable applications of the CB game; applications of Blotto games will be presented when we introduce each instance in particular (see Chapter 3). One of the original applications of the CB game is the military logistics, see e.g., Gross (1950) and Gross and Wagner (1950); in this setting, the forces might be soldiers, equipment or weapons. Another example is the use of the CB game in modeling problems in politics: players are the political parties who distribute their resources (time, people, money, etc.) to compete over voters (or states as in the US presidential election),[4] see e.g., Kovenock and Roberson (2012), Myerson (1993), and Roberson (2006). On the other hand, Chia (2012) and Schwartz et al. (2014) use the CB game to model problems in cybersecurity where the players are an attacker and a defender while battlefields are security targets (e.g., phishing sites) and resources are security forces or effort. The CB game is also used to model competitive contests in online advertising (see e.g., Masucci and Silva (2014, 2015)): two marketing campaigns allocate promotional gifts to compete for royalty of social networks' customers with high network value (i.e., high influence over peers). An example of applying the CB game into telecommunications is the following radio-spectrum management system (see e.g., Hajimirsaadeghi and Mandayam (2017)): two network service providers (NSP) (i.e., the players), each with a limited amount of bandwidth (resource), compete strategically by bidding on the users (battlefields), the user chooses to connect to the NSP that provides the larger bandwidth bid. A depiction of this radio-spectrum management system is given in Figure 1.1. The applications discussed above are

[3]Here, R_i^{ϕ} is a function that depends not only on the allocations but also on other parameters of the game (and/or possibly on some random variables).

[4]Applications of the CB game in politics often involve the majority-rule version; i.e., the player whose aggregate values (or the number) of battlefields won by her exceed a given threshold (often chosen to be 50%) wins the whole game and receive a positive payoff; the player who won less than that threshold receives a zero payoff.

summarized in Table 1.1.

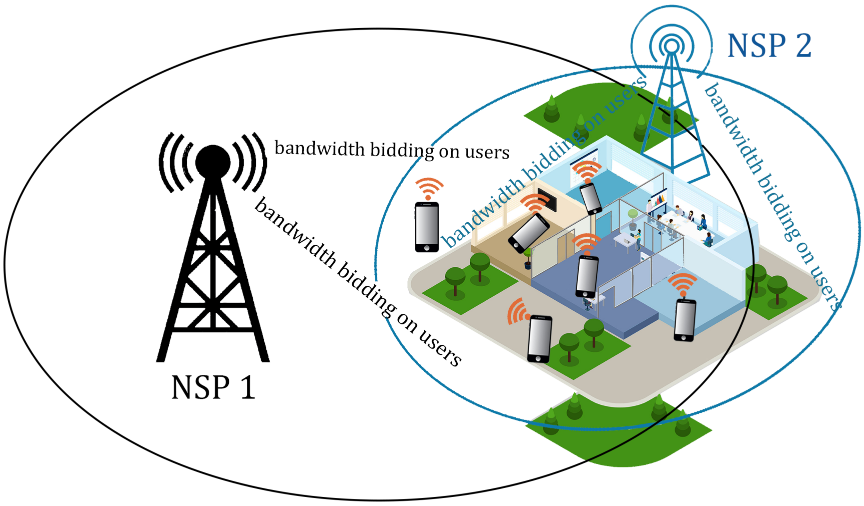


Figure 1.1: Illustration of modeling a radio spectrum allocation system as a CB game (a depiction of the system introduced by Hajimirsaadeghi and Mandayam (2017)).

Table 1.1: Several practical applications of the Colonel Blotto game.

Applications	Players	Resource	Battlefields	References
Politics	Political parties	Effort/ money	Voters/ states	Kovenock and Roberson (2012), Myerson (1993), and Roberson (2006).
Advertising	Marketing campaigns	Promotional gifts	Network users	Masucci and Silva (2014, 2015).
Security	Attackers/ defenders	Effort/ forces	Security targets	Chia (2012) and Schwartz et al. (2014).
Radio management	Network providers	Bandwidth	Mobile users	Hajimirsaadeghi and Mandayam (2017).

Beyond these direct applications, the CB game (and other Blotto games) is also analyzed under a large range of approaches from a diversity of communities and disciplines; it also connects to many other problems. These studies have provided a variety of important theoretical contributions. Because of this interdisciplinarity, instances of Blotto games are not always addressed under the same name and terminology and they are sometimes considered as instances of more general frameworks. For example, in the series of works by Laslier (2002, 2005) and Laslier and Picard (2002), the CB game, under the name "divide-a-dollar" problems,[5] is described in a slightly different manner: two players compete by distributing one unit of goods (one dollar) among a population of individuals that are partitioned into n groups (corresponding to n

[5]Note that this model differs from the two-player cooperative game with a similar name, proposed by Nash (1953).

battlefields). Analyses of the optimal strategies of the competitive players in this game are exploited to study coalition-forming strategies of the population (see also Baron and Ferejohn (1989) and Primo (2007)). On the other hand, the CB game can also be considered as an instance of multi-prize contests (see e.g., Fu and Wu (2019) and Konrad and Kovenock (2009) for surveys on contests). To be more precise, the CB game is a contest with budget constraints and use-it-or-lose-it costs (these terms are adopted from Kovenock and Roberson (2010)). In the contest theory literature, one of the most well-known models is the Tullock contest (introduced by Friedman (1958) and reintroduced by Tullock (1980)); it relates closely to the Lottery Blotto game—an instance of Blotto games (we discuss this with more details in Chapter 3). The study of the Lottery Blotto game extends further the literature on Tullock contests. Moreover, Blotto games also relates to the class of search games—introduced by Morgenstern and Von Neumann (1953) and later developed into the field of search theory (see Hohzaki (2016) for a survey on search games).

1.1.3 Perspectives of the Thesis and Main Challenges

In this thesis, we study resource allocation games—with the CB game as a case study—under two main settings: the offline setting and the online learning setting, each can be used to model a different set of applications.

The Offline Setting

We first consider resource allocation games involving situations where players have full access to information about the game before making decisions.[6] Here, the leading question is how to play strategically in order to optimize payoffs. The simplest and most natural approach to capture these situations of resource allocation games is to model them in the one-shot complete-information strategic form:[7] a game is played one time where players know all parameters and simultaneously choose their strategies. To analyze players' behaviors in this model, it is common in the game-theory literature to consider the well-known notion of Nash equilibria (see Section 2.1 for a formal definition). It turns out, however, that even in the CB game with its simple rule, this problem is non-trivial: despite the long-standing history of being studied intensively, *it still remains unknown how to completely characterize and compute Nash equilibria of the CB game in its most generalized parameters configurations* (see Section 3.3.1 for a more detailed discussion). The difficulty encountered here essentially comes from the budget constraints forcing a correlation between allocations of a player in different battlefields.[8] Note that partial results have been obtained in several restricted cases of the CB game (and they are applied to the corresponding practical situations—see Section 3.3.1) but the involved assumptions are not satisfied in many other applica-

[6]For example, in the CB game, each player knows the rule of the game, the number of battlefields and their values, as well as her budget and that of the opponent.

[7]A formal definition of the one-shot complete-information game is given in Section 2.1.

[8]Recall that budget constraints are essential elements found in all resource allocation games.

tions. Therefore, *it is crucial to find strategies that can guarantee good payoffs for players in the CB game with general configuration of parameters*. Moreover, in many applications of the CB game (and of other resource allocation games), the scale of the involved parameters can be very large while it often requires players to make decisions quickly; thus, we aim for strategies whose constructions are scalable. Besides that, the simplicity and interpretability of the proposed solutions are also important. By giving solutions satisfying these conditions, the scope of applications of the CB game model can be extended much further.

Our approach to solve this problem is to look for a special class of strategies in the CB game that guarantee near-optimal payoffs for players and such that the involved errors can be well-controlled. In other words, we *look for good approximate equilibria of the CB game with general configurations of parameters* (see Section 2.1 for a formal definition of approximate equilibria). Moreover, these *approximate equilibria are required to be constructed simply and efficiently*. Next, we study how this approach can be extended into other resource allocation games. In particular, we choose to first consider the family of Blotto games since they share a common basic structure with the CB game. In studying each instance of Blotto games, we determine new challenges and derive adjustments to overcome them. Showing extensions of results obtained in the CB game to other Blotto games, we aim to shed some light on the generalization and transferability of the CB game into other resource allocation games. In summary, **the first high-level challenge** that we consider in this thesis is the following:

> *In the one-shot complete-information CB game, how to find strategies that can be efficiently constructed and provably guarantee good payoffs for the players? In particular, even though optimality is not required for such strategies, the difference between the payoff derived from using these strategies and the optimal payoff is requires to be well-controlled and negligible, especially in large-scale instances. Can we extend this approach to other Blotto games?*

The Online Learning Setting

Second, we consider situations in which players need to repeatedly play a resource allocation game when they do not have complete information about the game's parameters; thus, they need to make predictions based on their (restricted and incomplete) observations about their payoffs in the past (several examples are given in Chapter 7). Hereinafter, we refer to these situations as the online setting (or online version) of resource allocation games. Among the cases of interest, we focus on resource allocation games where players' strategy sets have combinatorial structures; an example is the CB game where players' budgets and allocations are constrained to be integers (we call it the *online discrete CB game*, defined formally in Chapter 7). This set of games is particularly interesting to investigate because it is commonly found in practice and it poses considerable challenges in finding solutions guaranteeing good performance while maintaining implementability. It is natural to model the online versions of these games as online learning problems and to conduct a regret-minimization analysis.[9] In that case, the leading challenge is *how to find algorithms generating strategies (based on observed feedback) for a player to play at each stage in online resource allocation games (with combinatorial structures) such that it provides a good guarantee on the regret; moreover, these algorithms are required to be efficiently implementable.*[10]

Our approach to solve this problem is to cast online resource allocation games with combinatorial structures into online shortest path problems (OSP)—an important subclass of the online combinatorial optimization (OComb) framework, then use available tools from OComb and OSP as bases to develop better learning policies in these games.[11] We can carry out such conversions for several prominent online resource allocation games with combinatorial structures (including the online discrete CB game) by exploiting the games' structures.[12] The state-of-the-art algorithms in OComb and OSP still have issues in the performance's guarantees as well as the running time and there is still room for improvement (see also Section 7.3 for a literature review). We aim to first study and design improved algorithms for generic instances of OSP, then apply these findings into the instances corresponding to the online resource allocation games under consideration and show achievable benefits. Moreover, particular cases of online resource allocation games motivates us to study novel instances of OSP, e.g., the online discrete CB game with semi-bandit feedback motivate us to introduce and provide solutions for the OSP with side-observations model—an instance of OSP that has not been studied explicitly in the literature (see Section 7.2 and Section 8.1). In summary, **the second high-level challenge** that we consider in this thesis is the following:

> *How to play repeatedly online resource allocation games with combinatorial structures (e.g., the online discrete CB game) and guarantee a good payoff? In particular, can we design regret-minimization algorithms that run efficiently in the* OSP *instances corresponding to these games (if available) and provide improved regret guarantees in comparison with existing algorithms? Do these algorithms hold similar results in generic instances of* OSP*?*

[9] A formal definition of online learning problems can be found in Cesa-Bianchi and Lugosi (2006) while a formal definition of regret is given in Section 2.2. Intuitively, regret of a player is the difference between her cumulative payoff and the optimal payoff in the hypothetical scenarios where she knows all information and chooses a fixed strategy to play in all stages.

[10] We say that a regret-minimization algorithm in an online resource allocation game is efficient if it runs in polynomial time in terms of basic parameters of the game; e.g., in the online CB game, these parameters are the number of battlefields and the budgets.

[11] Briefly put, in an OSP, at each stage, edges on a graph are embedded with adversarially chosen losses; unknowing this, a learner has to choose a path and suffers the sum of losses of edges belonging to that path. See Section 2.2 for formal definitions of OComb and OSP.

[12] See Chapter 7 for a more detailed discussion on conversions of several online resource allocation games into OComb and OSP.

1.2 Contributions and Structure of the Thesis

In this thesis, we choose resource allocation games as the main objects of study and we focus on the CB game as a case study. Our results come mainly from applying methods and techniques from two different, but related, disciplines: game theory and online learning. They are used to address the two key challenges described above. Our *solution for the first high-level challenge* is to propose a class of simply-constructed approximate Nash equilibria of (one-shot and complete-information) Blotto games with well-controlled approximation errors. In the online setting, our *solution for the second high-level challenge* is to model the online discrete CB game (and several other online resource allocation games with combinatorial structures) as an online shortest path problem (OSP) and design efficiently implementable algorithms with good (expected) regret guarantees for OSP under several different feedback settings. This thesis is organized into three parts: Part I (Chapters 3, 4, 6 and 5) and Part II (Chapters 7, 8 and 9) respectively present the results in the two settings mentioned above, and Part III is dedicated to our conclusions on the obtained results and some discussions about future works. Note that we defer the in-depth reviews on the literature related to the CB game (and other related resource allocation game) and online learning problems to the first chapter of Part I and Part II respectively. A more detailed outline of the thesis and an overview of the results are presented below.

Chapter 2 serves as a background chapter in which for the sake of completeness, we revisit several preliminary definitions and results from the game theory and online learning literature. This includes the definition of strategic games and solution concepts such as Nash equilibrium and approximate (Nash) equilibrium; we also review definitions of online learning frameworks such as online linear optimization, online combinatorial optimizations, multi-armed bandits and online shortest path problems. Moreover, we introduce a general definition and several examples of resource allocation games.

Chapter 3 introduces formulations of the (one-shot complete-information) Blotto games that are our primary focus in Part I of this thesis. We start with the definition of the generalized CB game—the variant with the most generalized parameters configurations. Based on this basic model, we introduce two important variants: the constant-sum CB game (where players have the same evaluation of battlefields' values) and the discrete CB game (where budgets and all allocations are constrained to be integers). We also define two new extensions of the generalized CB game, called the generalized Lottery Blotto game (where the winner-determination rule is replaced by a generic contest success function) and the generalized-rule CB game (where the winner-determination rule is modified to capture situations involving pre-allocations and resources with asymmetric effectiveness). We conclude this chapter by discussing the literature and main challenges in characterizing Nash equilibria of the CB game and its variants/extensions.

In Chapter 4, we present our **first main contribution**: we propose a class of strategies (the independently uniform (IU) strategies) yielding approximate equilibria of the generalized CB game. To do this, we revisit the state-of-the-art results on optimal univariate distributions of players in this game and construct the IU strategies based on these distributions. We then characterize approximation errors of the IU strategies in terms of the game's parameters and prove that these errors are negligible when the number of battlefields is large. Recall that characterizing exact equilibria of the generalized CB game remains a challenging open question; moreover, even if an exact equilibrium is found, it is most likely that its construction would be complicated. Therefore, our simply-constructed approximate equilibria are useful and they extend further the scope of applications of the CB game model.

In Chapter 5, we consider the one-shot version of the discrete CB game and show an approximate equilibrium for this game with an approximation error that is negligible under a condition on the number of battlefields and the budgets (this is our **second main contribution**). We obtain this approximate equilibrium of the discrete CB game from modifying the IU strategies in the generalized CB game with a non-trivial round-up process. Although it can be proved that exact equilibria of the discrete CB game exist, the state-of-the-art methods for finding these equilibria still remain computationally impractical for large-size game instances. In contrast, the construction of the proposed approximate equilibria can be done by an efficient algorithm. We discuss the trade-off between the efficiency of this solution and the involved approximation error, in comparison with state-of-the-art results related to exact equilibria computations. Several numerical experiments are conducted to illustrate this trade-off.

Chapter 6 extends the results obtained in the previous chapter to other extensions of the CB game. More specifically, we prove that the class of IU strategies also yields an approximate equilibrium of the generalized Lottery Blotto game (LB game). Moreover, we can establish a connection between the CB game and some special cases of the LB game and give a characterization of approximation errors of IU strategies in these LB instances. This is our **third main contribution**; note that this extended result involves non-trivial analyses. Next, we also obtain several initial results in the generalized-rule CB game. We show that in this game, a modified version of IU strategies provides a similar approximate results (this is our **fourth main contribution**). To do this, we study the model of all-pay auctions with favoritism and completely characterize the exact equilibria of this game; this side-result fills in a gap in the all-pay auction literature.

Chapter 7 begins our analysis of online resource allocation games with combinatorial structures (i.e., the start of Part II). In particular, we define the online discrete CB game under three feedback settings:[13] *full-information* (i.e., the player observes all the game's parameters and her opponent's plays), *semi-bandit* (i.e., the player observes her gains in each battlefield but not the opponent's plays) and *bandit* (the player only observes her total payoff obtained from each stage and nothing else). We cast the online discrete CB game into the online shortest path problem (OSP) framework. We then discuss on similar conversions (into OSP) of several other online resource allocation games. To pinpoint the challenges and introduce bases for our studies in online resource allocation games, we give an in-depth literature review on regret-minimization analyses in

[13]These are information that a player may observe *after making decisions* at each stage.

bandit problems and other related works (including the literature of OComb and OSP). Finally, we also introduce a new instance of OSP, called OSP with side-observations (SOOSP), as a preparation for the following chapters.

In Chapter 8, we conduct an regret-minimization analysis of online shortest path problems with side-observations (SOOSP). This analysis is actually motivated by the fact that side-observations can be deduced in the online discrete CB game with semi-bandit feedback and that this game can be cast as an SOOSP. Existing algorithms (from the literature of OComb) that can be used for SOOSP have several drawbacks in the running-time (it lacks an overall guarantee on the efficiency) and regret-bound guarantees (a literature review is given in Section 7.3.3). We design a novel algorithm, Exp3-OE, that runs efficiently[14] in any instance of SOOSP and yields an improved regret bound in several cases of interest. This is our **fifth main contribution**. We then apply our novel algorithm to several online resource allocation games with combinatorial structures, including the online (discrete) semi-bandit CB game, the online hide-and-seek game and the online (discrete) CB game with full-information feedback. We end this chapter by discussing the benefits of using Exp3-OE in these games.

Chapter 9 presents initial results on the regret-minimization analysis of the online discrete CB game under the bandit feedback setting, being cast as an OSP with bandit feedback (OSPBand). State-of-the-art algorithms for OSPBand still have issues in implementation and there is still room for improvement in the performance guarantees. Our **sixth main contribution** is an algorithm, called Edge, that we obtain by introducing new modifications to the ComBand algorithm (introduced by Cesa-Bianchi and Lugosi (2012)). Unlike classical ComBand, the Edge algorithm always runs efficiently on any instance of OSPBand. Moreover, Edge allows more choices of inputs than other efficient algorithms in the literature, e.g., CombD by Sakaue et al. (2018); this upgrade eventually improves regret guarantees. We address the open question (posed by Cesa-Bianchi and Lugosi (2012)) on how to choose inputs for Edge (and ComBand) to run in OSPBand by providing a heuristic procedure that efficiently outputs an exploration distribution helping Edge improve its regret guarantee. We conduct several numerical experiments to illustrate these improvements in the case of the online CB bandit game.

Part III contains the conclusion of the thesis and our discussion on open directions for future research. In Appendix, we provide formal proofs and supplementary materials of all results presented in the previous chapters. Note that some of the ideas and results presented in this thesis have previously appeared in several of our publications; a full *list of publications* is given after Part III. In these publications, we conducted several computational studies and numerical experiments (that are also presented in this thesis) and published our codes of each result for the sake of reproducibility. We collect all these codes here at `https://github.com/dongquan11/CBGame_ApproxEqui_and_OnlLearning` and we also give the links (url) pointing toward related published codes where it is relevant in the thesis.

[14]Exp3-OE always runs efficiently in polynomial time in terms of the number of edges in the graph related to SOOSP.

Chapter 2

Preliminaries

For the sake of completeness, we review in this chapter several important concepts that serve as a basic for the results presented in this thesis. In particular, essential elements of game-theory and online combinatorial optimization are presented in Section 2.1 and 2.2 respectively. We note that this chapter only serves as a background chapter and that it is not a formal and exhaustive introduction to game theory or online learning; therefore, readers might find that the definitions that we give below are less abstract and contain less details than the ones often found in textbooks. We also omit several non-basic assumptions and concepts that lead to unnecessarily heavy notation.

2.1 Elements of Game Theory

Game theory is a collection of mathematical tools used to model strategic interactions which are situations where multiple agents try to make decisions with objectives or preferences that also depend on other agents' decisions. Modern game theory can be considered to be founded in the beginning of the last century with works by pioneers such as Borel (1921), Von Neumann (1928), and Zermelo (1913). Important breakthroughs come from the book by Morgenstern and Von Neumann (1953) and the series of works by Nash (1950, 1951). Some of the most basic concepts are the strategic game (also called the normal form game) and the Nash equilibrium. Below, we redefine these concepts in our notation, loosely based on the definitions in the books of Osborne and Rubinstein (1994) and Myerson (1991).

2.1.1 Strategic Game and Nash Equilibrium

A strategic game can be defined by a tuple $\Gamma = \left(\mathcal{N}, \left(S^j\right)_{j\in\mathcal{N}}, \left(u^j\right)_{j\in\mathcal{N}}\right)$, where:

$\mathcal{N}$ is the set of (selfish and rational)[1] players; $|\mathcal{N}| = N$;

S^j is the set of actions of player $j \in \mathcal{N}$;

[1]As often the case in the literature, here, "rational" players mean the ones satisfying the assumptions of Von Neumann (1953) and of Savage (1972) (see more details in e.g., Osborne and Rubinstein (1994) p.5).

$u^j : S \to \mathbb{R}$ is the utility function (i.e., payoff function) of player $j \in \mathcal{N}$; here we denote $[N] = \{1, 2, \dots, N\}$ and $S := \bigtimes_{j\in[N]} S^j$.

A **one-shot complete-information strategic game** $\Gamma = \left(\mathcal{N}, \left(S^j\right)_{j\in\mathcal{N}}, \left(u^j\right)_{j\in\mathcal{N}}\right)$ is interpreted as follows: it models an event that occurs only once where each player $j \in \mathcal{N}$ knows all the parameters and the details of the game (that is $\mathcal{N}, S^j$ and u^j for any $j \in \mathcal{N}$) and all players choose their actions simultaneously[2] and independently. Each element of S^j is called a *pure strategy* of player $j \in \mathcal{N}$; and each element of S is referred to as a *strategy profile*.

In the remainder of this chapter, unless stated differently, we refer to a one-shot complete-information strategic game as a strategic game (or simply, a game). We also note that a more general definition of a strategic game can be found in the literature, in which, one replaces the utility functions in the above definition by a set of preference relations $\succeq^j$ (for $j \in \mathcal{N}$) on the profiles set S.[3] Under a large range of circumstances, these definitions are equivalent; for the purposes of this thesis, we choose to present it with the above manner which we find to be a simpler and more intuitive definition. Throughout the thesis, we also often mention a class of games, called *constant-sum games*, defined as games where the summation of all players' payoffs is always the same regardless of players' actions. A well-known instance of the constant-sum game is the two-player zero-sum game where the gain or loss of utility of a player is exactly balanced by the loss or gain of the utility of the other player. We use the term *non-constant-sum* to refer to games that do not satisfy the condition of the constant-sum game.

Next, we present the definition of Nash equilibrium—one of the most important solution-concepts in game theory. Essentially, it is a steady state notion of the plays in strategic games.

Definition 2.1.1 (Nash equilibrium). *In a strategic game* $\Gamma = \left(\mathcal{N}, \left(S^j\right)_{j\in\mathcal{N}}, \left(u^j\right)_{j\in\mathcal{N}}\right)$, *a Nash equilibrium is a strategy profile* $s_* = (s_*^j)_{j\in\mathcal{N}} \in S$ *such that for every player* $j \in \mathcal{N}$, *we have:*

$$u^j(s_*) \geq u^j(s^j, s_*^{-j}), \forall s^j \in S^j.$$

Here, the notation s_*^{-j} denotes the collection of strategies of all players except j in the strategy profile s_* and (s^j, s_*^{-j}) denotes the strategy profile where one takes s_* and replaces s_*^j by s^j. Intuitively, a strategy profile s_* is a Nash equilibrium when there exists no player that can unilaterally deviate from s_* and gain a strictly larger utility.

On the other hand, given $s^{-j} \in \bigtimes_{j'\in\mathcal{N}\setminus\{j\}} S^{j'}$, the set of *best-response* of player j against s^{-j} is the following:

$$B^j(s^{-j}) = \{\tilde{s}^j \in S : u^j(\tilde{s}^j, s^{-j}) \geq u^j(s^j, s^{-j}), \forall s^j \in S^j\}.$$

[2]Here, "simultaneous" does not necessarily always mean that the actions are taken at the same time; it rather means that when a player chooses her action, she is unaware of the choices being made by the other players and that she does not have any extra information except from the game's parameters to use to interpret other players' behaviors.

[3]That indicates that player j prefers one strategy profile than the other.

Based on this notation, a Nash equilibrium of a game Γ can also be defined as any strategy profile s_* such that $s_*^j \in B^j(s_*^{-j})$ for any $j \in \mathcal{N}$.

In the special case of two-player constant-sum games (both players want to maximize the utility), Nash equilibrium is also equivalent to the following solution concept:

Definition 2.1.2 (Max-min strategy)**.** *Let* $\Gamma = \big(\{A, B\}, \{S^A, S^B\}, \{u^A, u^B\}\big)$ *be a two-player constant-sum game, a strategy profile* (s_*^A, s_*^B) *is a max-min strategy if for any strategy* $\tilde{s}^A \in S^A$ *and* $\tilde{s}^B \in S^B$ *of players A and B:*

$$\min_{s^B} u^A(s_*^A, s^B) \geq \min_{s^B} u^A(\tilde{s}^A, s^B), \tag{2.1}$$

$$\min_{s^A} u^B(s^A, s_*^B) \geq \min_{s^A} u^B(s^A, \tilde{s}^B). \tag{2.2}$$

Intuitively, if a player $\phi \in \{A, B\}$ plays a max-min strategy, she guarantees an optimal payoff even in the worst-case scenario when her opponent $-\phi$ plays the strategies that minimize ϕ's payoff (no matters how it affects $-\phi$'s payoff).

Now, let $\Gamma = \Big(\mathcal{N}, \big(S^j\big)_{j \in \mathcal{N}}, \big(u^j\big)_{j \in \mathcal{N}}\Big)$ be a strategic game, we denote the set of all probability distributions over S^j by $\Delta(S^j)$ and define the set $\Delta := \bigtimes_{j \in \mathcal{N}} \Delta(S^j)$. We refer to the elements of $\Delta(S^j)$ as the *mixed strategies* of player j, and the elements of Δ as mixed-strategy profiles. Hereinafter, we assume that in any game, the randomization in players' mixed strategies are independent. Trivially, any pure strategy of a game is also a mixed strategy (but not reversely).

Next, we define the *mixed extension* of the strategic game Γ: it is the strategic game $\tilde{\Gamma} = \Big(\mathcal{N}, \big(\Delta(S^j)\big)_{j \in \mathcal{N}}, \big(U^j\big)_{j \in \mathcal{N}}\Big)$ where $U^j : \Delta \to \mathbb{R}$ is the utility function that takes any mixed strategy profile $(\sigma^j)_{j \in \mathcal{N}} \in \Delta$ as input and returns the expected value of u^j with respect to the randomness of σ^j for all $j \in \mathcal{N}$.

Definition 2.1.3 (Mixed strategy Nash equilibrium)**.** *Given a strategic game* Γ*, a mixed strategy Nash equilibrium of* Γ *is a Nash equilibrium of its mixed extension.*

The concepts of best-response and max-min strategy are trivially extended to work with mixed-strategies. Moreover, for finite strategic games (that is games where the set of players and the players' strategy sets are all finite), we have two important propositions as follows:

Proposition 2.1.4 (extracted from Nash (1951))**.** *Every finite strategic game has a mixed strategy Nash equilibrium.*

Proposition 2.1.5. *Let* $\Gamma = \Big(\mathcal{N}, \big(S^j\big)_{j \in \mathcal{N}}, \big(u^j\big)_{j \in \mathcal{N}}\Big)$ *be a finite game, then* $\sigma_* \in \bigtimes_{j \in \mathcal{N}} \Delta(S^j)$ *is a mixed strategy Nash equilibrium of* Γ *if and only if for any player* $j \in \mathcal{N}$*, every pure strategy in the support of* σ_* *is a best response against* σ_*^{-j}*.*

Beside the Nash equilibrium, there exist several other stable-state notions; however, they do not relate to results in this thesis. Therefore, in the remaining chapters, unless stated otherwise, we refer to a (mixed strategy) Nash equilibrium simply as an equilibrium.

2.1.2 Approximate Equilibria

The concept of Nash equilibrium is elegant and useful in predicting players' behaviors. However, there exist games that have no equilibrium (Proposition 2.1.4 only applies to finite games). Moreover, even in the cases where equilibrium exists, it is not always easy to construct them. Therefore, it is desired to have another stable state notion for games. One approach is to relax the notion of equilibrium and allow some error margins.

Definition 2.1.6 (Approximate Equilibria). *Given $\varepsilon \geq 0$, an* ε-equilibrium *of a game* $\Gamma = \left(\mathcal{N}, \left(S^j\right)_{j\in\mathcal{N}}, \left(u^j\right)_{j\in\mathcal{N}}\right)$ *is any strategy profile $s_* \in S$ such that $u^j(s_*^j, s_*^{-j}) + \varepsilon \geq u^j(s^j, s_*^{-j})$ for any $j \in \mathcal{N}$ and $s^j \in S^j$. We use the generic term* approximate equilibrium *whenever the approximation error ε need not be emphasized.*

Trivially, a 0-equilibrium corresponds to an (exact) equilibrium defined in the previous section. Note also that approximate equilibria might exist even in games that have no equilibrium. It is meaningful only for us to consider the ε-equilibria that have small approximation error ε, relatively to the magnitude of players' payoffs. Finally, in the two-player constant-sum game, note also that the concept of *approximate max-min strategy* can be trivially extended from Definition 2.1.2 and Definition 2.1.6.

2.1.3 Resource Allocation Games

As discussed in Chapter 1, this thesis takes special interest in a class of games, namely resource allocation games. They are games that are used to model situations where players need to distribute a limited budget of resources in order to optimize their utility—these situations are commonly found in practice. Formally, in this thesis, by a resource allocation game, we refer to the following definition.

Definition 2.1.7 (Resource allocation game). *A strategic game $\Gamma = \left(\mathcal{N}, \left(S^j\right)_{j\in\mathcal{N}}, \left(u^j\right)_{j\in\mathcal{N}}\right)$ is a resource allocation game if for each player j, there exist $n^j, X^j \in (0, \infty)$ such that there is a one-to-one mapping between S^j and a subset of $\left\{\left(s_1^j, \ldots, s_{n^j}^j\right) : \sum_{i\in[n_j]} s_i^j \leq X^j\right\} \subseteq \mathbb{R}^n$.*

By this definition, the actions that players can choose are in the form of n-tuples and X^j is the limited budget of player j. We call the condition $\sum_{i\in[n_j]} s_i^j \leq X^j$ by the budget constraint of player j. Note that to the best of our knowledge, similar definitions have not been formally presented in the literature. The above definition is given in our notation, we try to define it in a manner that is as general and abstract as possible. This definition covers a variety of games models studied in the literature. Notably, according to Definition 2.1.7 the Colonel Blotto game (as presented in Chapter 1) is a resource allocation game. Several other notable examples are listed below.

Example 2.1.8 (Multi-looking hide-allocation game). *The game involves two players: one player hides a ball in one of n boxes and the other player searches for it. There are known probabilities that the searcher will overlook the ball if he searches the correct box. The hider wishes to minimize and the searcher wants to maximize the probability that the ball will be found in m or fewer search ($m \leq n$).*

This game is proposed by Subelman (1981). In this game, the hider's budget is 1 and the searcher's budget is m. This game relates to the class of *hide-and-seek game*, introduced by Von Neumann (1953), in which players play on a matrix: the hider chooses a cell, say (r, c), and the searcher chooses a row (or a column); if the chosen cell of the hider is in the row (or column) chosen by the searcher, the searcher gains a value, say q_{rc}; the searcher wants to maximize, the hider wants to minimize. There are many variants and extensions of this hide-and-seek game; e.g., Morris (1962) extends Von Neumann's game by allowing searchers to search for cells instead of rows or columns, Beck and Newman (1970) characterizes the equilibrium of the linear search game, Bostock (1984) works on hide-and-seek game embedded on a simple network with two nodes and three arcs, similar games on more complicated networks and trees can be found in Alpern et al. (2008, 2009) and Dagan and Gal (2008). We will revisit the hide-and-seek game in Chapter 8 with more details.

The hide-allocation game described above might also be considered as a search game (see e.g., Hohzaki (2016) for a survey of search games). Another search game that also belongs to the class of resource allocation game is the following:

Example 2.1.9 (Smuggling game)**.** *There are two players, an inspector and an inspectee. During n stages, the inspectee has at most a chance to violate a treaty in order to obtain a reward; while the inspector can choose to either "inspect" or "not inspect"; the inspector can only "inspect" m times ($m \leq n$). At each stage, in the cases when the inspectee violates the treaty: the inspector receives a reward of one if he inspected, and the inspectee loses the same amount of reward; players receive reserved rewards if no inspection was carried out. All decisions are made before the sequence of stages takes place.*

This game is studied by Avenhaus and Kilgour (2004) and Hohzaki (2007). There exist sequential and stochastic multi-stage versions of this game, known as the inspection game (see e.g., Avenhaus and Canty (1996), Dresher (1962), and Maschler (1966); these works are applied to make an inspection plan of the International Atomic Energy Agency (IAEA) for the Non-Proliferation Treaty for Nuclear Weapons (NPT) (see Hohzaki (2016) p.5)).

Example 2.1.10 (Search-search game)**.** *Two searchers have limited amounts of effort, denoted by X^A and X^B respectively. The searchers simultaneously assign their search effort to n targets. If the first searcher has allocation plan of $(a_1, \ldots, a_n)$ (such that $\sum_{i \in [n]} a_i \leq X^A$) and the second searcher has an allocation of $(b_1, \ldots, b_n)$ (such that $\sum_{i \in [n]} b_i \leq X^B$); then the payoff of searcher j at target i is $R_i^j(a_i, b_i)$ and her total payoff is $\sum_{i \in [n]} R_i^j(a_i, b_i)$.*

The generality of the utility functions R_i^j used in this game model provides the flexibility to cover a large set of game instances. We note that the Colonel Blotto game—the main focus of this thesis—can be also considered as an instance of this search-search game. The framework in Example 2.1.10 was actually proposed by Blackett (1958), under the name Blotto game, as an 3-page technical report on some necessary conditions for this game to have a pure equilibrium. Later on, it was extended and studied under the name search-search game (see e.g., Garnaev (2007), Hohzaki (2016), and Nakai (1986)).

2.2 Elements of Online Combinatorial Optimization

In this section, we review the model of online combinatorial optimization problems (hereinafter, OComb) which is an instance of the online linear optimization (henceforth, OLO)—a general framework that covers many classical problems, including the well-known multi-armed bandits (hereinafter, MAB). These three models will be sequentially redefined under our notation in Section 2.2.1 and Section 2.2.2. All of them belong to the class of prediction problems—each instance of this class can be considered to be a sequence of games played between a learner (also called a forecaster) and an adversary (can possibly be nature); see e.g., the book of Cesa-Bianchi and Lugosi (2006) for the relation between predictions and playing games. Note that the notion of the strategic game presented in the previous section fails to capture the prediction problem. This is due to the fact that predictions involve situations where there exist essential links between the sequential plays, i.e., the learner not only needs to care about her instantaneous payoff but also needs to think in a long-term manner because her current actions may affect future payoffs; moreover, when making the decisions, the learner may be *uninformed* about not only the plays of the adversary but also the payoffs of the actions in the past and/or even some parameters of the game.

The literature on OComb and other related models is rather exhaustive and diversified; in this thesis, we focus only on the regret-based learning algorithms. An important class of regret-minimization algorithms is Exp3; this is one of our main focuses in Part II of this thesis. In Section 2.2.2, we formally present the most basic version of Exp3, implemented in an MAB problem. Finally, in Section 2.2.3, we also present a special instance of OComb, called the online shortest path problem (henceforth, OSP), that involves making decisions on directed acyclic graphs. We also review a dynamic programming technique, called weight-pushing, used to efficiently sample a path from a special distribution; this can be applied to improve the efficiency of the Exp3-type algorithms when running in OSPs. Note finally that in this section, we only present the definitions and we delay the literature review on the regret-minimization algorithms used in OComb and OSP to Chapter 7.

2.2.1 Online Combinatorial Optimization

We begin with the model of online linear optimization problems. First introduced by Hannan (1957), this model covers many classical problems that were introduced and studied later on, e.g., the experts problem (see e.g., Cesa-Bianchi and Lugosi (2006)), the online shortest path problem (see e.g., György et al. (2007) and Takimoto and Warmuth (2003)) and the tree update problem (see e.g., Sleator and Tarjan (1985)).[4] The following definition is adapted from Kalai and Vempala (2005) and Dani et al. (2008), rewritten here in our notation:

[4]For a discussion on the relation between online linear optimization and these problems, see e.g., Kalai and Vempala (2005).

Definition 2.2.1. *An **online linear optimization problem** (*OLO*) is a T-round game between a learner and an adversary ($T \in \mathbb{N}\backslash\{0\}$), described as follows: let $S \subset \mathbb{R}^D$ be the action set of the learner, where $D \in \mathbb{N}\backslash\{0\}$ is fixed; at each stage $t \in [T]$, a loss vector $\boldsymbol{\ell}_t \in [0,1]^D$ is generated by the adversary; without knowing this, the learner chooses a vector $\tilde{\boldsymbol{p}}_t \in S$; then, a scalar loss $L\left(\tilde{\boldsymbol{p}}_t\right) = (\boldsymbol{\ell}_t)^\top \tilde{\boldsymbol{p}}_t$ is incurred; at the end of the stage, the learner receives some feedback about the losses. The learner's objective is to control the regret, defined as*

$$r_T := \sum_{t=1}^{T} L\left(\tilde{\boldsymbol{p}}_t\right) - \min_{\boldsymbol{p} \in S} \sum_{t=1}^{T} L\left(\boldsymbol{p}\right). \tag{2.3}$$

Intuitively, the regret is the difference between the cumulative losses that the learner suffers by playing $\tilde{\boldsymbol{p}}_t$ at stage t and that of playing the best action in hindsight. Note here that as a convention of notation, in OLO (and its instances including OComb and OSP defined below), we use the letter $\boldsymbol{p}$ to denote an arbitrary action[5] and use the *tilde notation* (i.e., $\tilde{\boldsymbol{p}}_t$) to emphasize that this is *the action chosen by the learner at stage t*. Additionally, we denote the i-th coordinate of $\boldsymbol{p}$ as $\boldsymbol{p}(i)$.

More importantly, in this thesis, we consider algorithms (telling the learner how to play in each stage) that involve randomization; and thus, we focus on analyzing the minimization of the *expected regret*, defined as follows:[6]

$$R_T := \max_{\boldsymbol{p} \in S} \mathbb{E}\left[\sum_{t=1}^{T} L\left(\tilde{\boldsymbol{p}}_t\right) - \sum_{t=1}^{T} L\left(\boldsymbol{p}\right)\right]. \tag{2.4}$$

We call an algorithm guaranteeing an expected regret that is sub-linear in terms of the time horizon T as a *no-regret algorithm* (in this case, $R_T/T \to 0$ as $T \to \infty$). Note also that besides the expected regret R_T defined in (2.4), another main focus of the literature is to have a guarantee on the regret with high probability (in this thesis, we do not focus on this approach).

In Definition 2.2.1, we choose the bounded condition on the loss vector $\boldsymbol{\ell}_t = (\boldsymbol{\ell}_t(1), \ldots, \boldsymbol{\ell}_t(D))$ generated by the adversary to be normalized into $[0,1]^D$; however, it can be relaxed such that $\left|(\boldsymbol{\ell}_t)^\top \boldsymbol{p}\right| \leq M, \forall \boldsymbol{p} \in S$ (for a given $0 < M < \infty$). Except from this condition, there is no other assumption on how the adversary generates the loss vector; that is, $\boldsymbol{\ell}_t$ can be an output of a function depending on the plays of the learner in the past, i.e., depending on $\tilde{\boldsymbol{p}}^s, \forall s \in [t-1]$; in that case, we call this an *non-oblivious adversary*; otherwise, we call it an *oblivious-adversary*. Note that in OLO problems with an oblivious adversary, the expected regret (given in (2.4)) can be rewritten as:

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} L\left(\tilde{\boldsymbol{p}}_t\right)\right] - \min_{\boldsymbol{p} \in S} \sum_{t=1}^{T} L\left(\boldsymbol{p}\right). \tag{2.5}$$

On the other hand, the feedback that the learner receives at the end of each stage has not been indicated precisely in Definition 2.2.1; in practice, these feedback depend on

[5] The reason is to make it consistent with the online shortest path problem (see Section 2.2.3) where each action is a path in a graph (here, $\boldsymbol{p}$ stands for "path").

[6] Here, the expectation is taken with respect to the randomization in the actions selection, i.e., internal randomization of the algorithms, and the randomization in the losses generation of the adversary.

the situations to be modeled. The type of feedback is also important in designing the algorithms running in OLO. Two basic feedback settings considered in the literature on linear optimization are as follows:

> The *full-information* setting: at the end of stage t, the learner observes the loss vector $\boldsymbol{\ell}_t$.

> The *bandit* setting: at the end of stage t, the learner observes the scalar (total) loss $L(\tilde{\boldsymbol{p}}_t) = (\boldsymbol{\ell}_t)^\top \tilde{\boldsymbol{p}}_t$.

Intuitively, with full-information feedback, the learner knows the loss that she would have suffered if she had chosen to play $\tilde{\boldsymbol{p}}_t := \boldsymbol{p}$ in stage t for any action $\boldsymbol{p} \in S$ (regardless of whether she really played it or not); on the other hand, in the bandit feedback setting, the learner only observes the loss incurred by an action if and only if she chose to play it. As a remark, the model of OLO with bandit feedback is often called as the *non-stochastic linear bandits* or *adversarial linear bandits*;[7] these terms are used to emphasize the fact that no assumption is made on how the losses are generated. On the other hand, it is common to use the term *stochastic linear bandit* to refer to the OLO with bandit feedback and an additional condition that for any stage $t \in [T]$, the loss vectors $\boldsymbol{\ell}_t$ is generated independently from a certain distribution.

A literature review on the algorithms in OLO will be given in Chapter 7. Now, one of the most important instances of OLO with many applications in practice is the problem where the action set of the learner has a combinatorial structure. We formally define it as follows:

Definition 2.2.2. *An* ***online combinatorial optimization*** (OComb) is an instance of OLO where the action set $S \subset \{0,1\}^D$.

In other words, OComb is an OLO problem where each action is a D-dimensional 0-1 vector. The OComb framework covers a variety of prediction problems, including the well-known multi-armed bandit (we revisit it in Section 2.2.2) and the online shortest path problem (we define it in Section 2.2.3); see e.g., Kalai and Vempala (2005) and Kocák et al. (2014) for other applications of the OComb model. Naturally, all the terminology of OLO will be transferred to OComb; particularly, the notions of regret, expected regret, oblivious/non-oblivious adversary and the two defined feedback settings. Note that the OComb with bandit feedback is also called a *combinatorial bandit* (this term is adopted from Cesa-Bianchi and Lugosi (2012)). Beside the full-information and bandit feedback described above, in OComb, another setting that is also considered as standard is:

> The *semi-bandit* feedback setting: at the end of stage t, the learner observes the (scalar) losses $\boldsymbol{\ell}_t(i)$ for any $i \in \{i \in [D] : \tilde{\boldsymbol{p}}_t(i) = 1\}$.

[7] Some works use these two terms interchangeably; however, there are works that use the term *adversarial linear bandit* to indicate specifically the OLO with a non-oblivious adversary.

Here, recall that $\boldsymbol{\ell}_t(i)$ and $\tilde{\boldsymbol{p}}_t(i)$ respectively denote the i-th coordinate of the loss vector $\boldsymbol{\ell}_t$ and the action vector $\tilde{\boldsymbol{p}}_t$. Intuitively, in the semi-bandit setting, the learner observes the coordinates of the loss vector $\boldsymbol{\ell}_t$ that correspond to the non-zero coordinates of the chosen action $\tilde{\boldsymbol{p}}_t$. An equivalent way to describe the semi-bandit feedback is that the learner observes all the products $\boldsymbol{\ell}_t(i) \cdot \tilde{\boldsymbol{p}}_t(i), \forall i \in [D]$ at the end of stage t.[8] Note that besides the three standard settings introduced above, there exist other feedback settings that are considered in the literature of OComb. One of them is the *semi-bandit feedback with side-observation* setting, proposed by Kocák et al. (2014), defined as follows:

> The *semi-bandit with side-observations* feedback setting: at the end of stage t, the learner observes the (scalar) losses $\boldsymbol{\ell}_t(i)$ for any $i \in O_t$ where O_t is a set such that $\{i \in [D] : \tilde{\boldsymbol{p}}_t(i) = 1\} \subseteq O_t \subseteq \{1, \ldots, D\}$.

Intuitively, in this setting, at the end of stage t, the learner observes the semi-bandit feedback (corresponding to her chosen action $\tilde{\boldsymbol{p}}_t$), then additionally, she observes some other elements of the loss vector (i.e., $\boldsymbol{\ell}_t(j)$ for *some* j such that $\tilde{\boldsymbol{p}}_t(j) = 0$). In this definition, O_t is the set of the observations of the learner at the end of stage t. This observation system can also be presented by a directed graph with D vertices, in which, if there is an edge from vertex i to vertex j and that $\tilde{\boldsymbol{p}}_t(i) = 1$, then the learner observes both $\boldsymbol{\ell}_t(i)$ and $\boldsymbol{\ell}_t(j)$. This graph is called the *observation graph at stage t*; if it is revealed to the learner before she makes the decision at each stage, it is called the informed setting; otherwise, when it is revealed only at the end of the stage, it is called the uninformed setting. We will review this feedback setting of OComb in more details in Chapter 8.

As a summary, by using the notation $X \succ Y$ to denote that with the same choice of policies, the learner observes more information at the end of each stage in OComb with the feedback setting X than in OComb with the feedback setting Y, we have an overview of all four feedback settings as follows:

Full information $\succ$ Semi-bandit with side-observation $\succ$ Semi-bandit $\succ$ Bandit.

2.2.2 The Exp3 Algorithm in Multi-armed Bandits and the ComBand Algorithm in OComb with Bandit Feedback

The Adversarial Multi-armed Bandit

In this section, we give the formulation of the well-known multi-armed bandit (MAB) model and Exp3—the basic variant of the class of algorithms that serves as the groundwork for our analyses in Part II of this thesis. We also introduce a variant of Exp3, called ComBand (introduced from Cesa-Bianchi and Lugosi (2012)), that is considered as a standard algorithm for OComb.

First, we define the MAB problem as an instance of the OComb framework:

[8] If $\tilde{\boldsymbol{p}}_t(i) = 0$, then $\boldsymbol{\ell}_t(i) \cdot \tilde{\boldsymbol{p}}_t(i) = 0, \forall \boldsymbol{\ell}_t(i)$; this means that the learner does not have any information on the loss coordinate $\boldsymbol{\ell}_t(i)$.

Definition 2.2.3. *An adversarial* ***multi-armed bandit*** *problem* (MAB*) (with D arms) is an* OComb *with bandit feedback*[9] *in which the learner's action set is* $S := \{e^1, e^2, \ldots, e^D\}$ *where* e^i *is the i-th standard basis vector in* $\mathbb{R}^D$*. Each vector in S is called as an arm.*

Although we define MAB here as an instance of OComb, the MAB model was actually invented long before the OComb and OLO problems (MAB is first considered by Thompson (1933)). The classical formulation of MAB is equivalent to Definition 2.2.3 but it is often given under a slightly different description: *the learner has D arms, labeled from* 1 *to D; at each stage t, each arm* $i \in [D]$ *is embedded with a loss* $L_t(i)$*; without knowing these losses, the learner chooses one arm, then observes and suffers the corresponding loss; the objective of the learner is also to minimize the regret*. The equivalence between these two definitions can be seen by observing that: in Definition 2.2.3, the action set S has D vectors corresponding to D arms in the classical description and by denoting the loss vector $\boldsymbol{\ell}_t = (L_t(1), \ldots, L_t(D))$, the loss corresponding to each arm i (i.e., the vector $e^i \in S$) is exactly equal to $(\boldsymbol{\ell}_t)^\top e^i = L_t(i)$. On the other hand, as a reversed perspective, any OComb problem with bandit feedback and an arbitrary action set $S \subset \{0,1\}^D$ can also be considered as an MAB problem with $|S|$ arms; where each action vector in S is mapped with an arm.

Note also that Definition 2.2.3 defines the *adversarial* MAB (also called the non-stochastic bandits) which is the MAB problem where we have no assumption on the losses (except from being bounded); this is used to distinguish with the *stochastic* MAB (also called the stochastic bandits) where the loss of each arm i at stage t is independently drawn from a distribution (unchanged through time) whose mean remains unknown to the learner. Finally, the notions of the (expected) regret, oblivious/non-oblivious adversary in MAB are naturally transferred from the corresponding notions in OComb and OLO.

The Exp3 Algorithm in Multi-armed Bandits

The literature on the MAB (both the stochastic and adversarial bandits) is extremely rich, this problem has been studied in many approaches and the list of obtained results is extensive. A literature review on the regret-minimization algorithms used in MAB will be given in Chapter 7. Among them, Exp3 (which stands for the "exponential-weight algorithm for exploration and exploitation") is one of the most important classes. In the following, we present a basic variant of Exp3, designed for the adversarial MAB by Auer, Cesa-Bianchi, Freund, et al. (2002). This is given in Algorithm 1, rewritten under our notations and in a slightly more general formulation.

The main idea of the Exp3 algorithm is to keep a weight for each arm at each stage t (denoted by $w_t(e^i)$ for arm e^i); then with probability $1-\gamma$, the learner samples randomly an arm from the normalized weights (called the exploitation) and with probability γ, samples an arm uniformly at random (called the exploration) (line 4—Algorithm 1);

[9] In fact, in the OComb with $S := \{e^1, e^2, \ldots, e^D\}$, semi-bandit feedback coincides with bandit feedback—the information that the learner observes in these cases are exactly the same.

Algorithm 1: The Exp3 Algorithm for MAB.

Input: Set of D arms $S = \{\boldsymbol{e}^1, \ldots, \boldsymbol{e}^D\}$; time horizon $T \in \mathbb{N}\backslash\{0\}$, parameters $\gamma \in [0,1]$, $\eta > 0$.

1 Initialize $w_1(\boldsymbol{e}^i) = 1, \forall i \in [D]$.
2 **for** $t = 1, 2, \ldots, T$ **do**
3 For any $i \in [D]$, the adversary chooses the loss $L_t(i) \in [0,1]$ embedded to the arm $\boldsymbol{e}^i$ (unobserved by the learner).
4 $\forall i \in [D]$, $x_t(\boldsymbol{e}^i) = (1-\gamma)\frac{w_t(\boldsymbol{e}^i)}{\sum_{\boldsymbol{e} \in S} w_t(\boldsymbol{e})} + \gamma \frac{1}{D}$
5 Sample an action $\boldsymbol{e}^I$ from the distribution $x_t(\boldsymbol{e}^1), \ldots, x_t(\boldsymbol{e}^D)$
6 Play the arm $\boldsymbol{e}^I$; then, suffer and observe the loss $L_t(I)$.
7 Compute the estimated loss $\hat{L}_t(i) = \frac{L_t(i)}{x_t(\boldsymbol{e}^i)} \mathbb{I}_{\{i=I\}}$, for any $i \in [D]$.
8 $\forall i \in [D]$, update $w_{t+1}(\boldsymbol{e}^i) := w_t(\boldsymbol{e}^i) \exp\left(-\eta \hat{L}_t(i)\right)$.

finally, based on the feedback, the learner estimates the loss of each arm[10] and updates the weights for the next stage (line 8—Algorithm 1). Importantly, *the running time of* Exp3 is $O(D)$ where D is the number of actions/arms. It can be proven that Exp3 is a no-regret algorithm for any MAB problem; formally, we have the following proposition.

Proposition 2.2.4 (Extracted from Cesa-Bianchi and Lugosi (2012))**.** *Running the* Exp3 *algorithm in the adversarial D-armed MAB (with appropriate parameters γ, η),*[11] *the expected regret of the learner is:*

$$R_T \leq O\left(\sqrt{TD\ln(D)}\right).$$

The ComBand Algorithm for OComb with Bandit Feedback

As discussed above, any OComb with bandit feedback and an arbitrary action set $S \subset \{0,1\}^D$ can be rewritten as an MAB problem with $|S|$ arms (each action corresponds to an arm); therefore, in principle, one can apply directly the Exp3 algorithm to any OComb problem (with bandit feedback). By doing this, the regret upper-bound provided by Exp3 is in $O(\sqrt{2T|S|\log|S|})$ and it runs in a polynomial time in terms of $|S|$. However, in most of (if not all) the cases of interest in practice, the number of actions in the OComb (i.e., $|S|$) is an exponential number in terms of D; therefore, the above performance of Exp3 in OComb with bandit feedback is poor and needed to be improved further.

An algorithm that improves the regret upper-bound in OComb is the ComBand algorithm; it is proposed by Cesa-Bianchi and Lugosi (2012) as an improved variant of the Exp3 algorithm. A pseudo-code of ComBand, written in our notation, is given as Algorithm 2.

[10] Note that $\hat{L}_t(i)$ in Algorithm 1 is an unbiased estimation of $L_t(i)$, i.e., $\mathbb{E}\left[\hat{L}_t(i)\right] = L_t(i)$.

[11] See Cesa-Bianchi and Lugosi (2012) for more details.

Algorithm 2: ComBand(μ) for OComb with bandit feedback.

Input: $S \subset \{0,1\}^D$, $T \in \mathbb{N}$, $\gamma \in [0,1]$, $\eta > 0$, distribution μ on S.

1 $\forall \boldsymbol{p} \in S$, $w_1(\boldsymbol{p}) := 1$.

2 **for** $t = 1, 2, \dots, T$ **do**

3 Loss vector $\boldsymbol{\ell}_t \in [0,1]^D$ is adversarially chosen (unobserved by the learner).

4 $\forall \boldsymbol{p} \in S$, $\nu_t(\boldsymbol{p}) := \frac{w_t(\boldsymbol{p})}{\sum_{\boldsymbol{q} \in S} w_t(\boldsymbol{q})}$ and $x_t(\boldsymbol{p}) = (1-\gamma)\nu_t(\boldsymbol{p}) + \gamma\mu(\boldsymbol{p})$.

5 Sample and play $\tilde{\boldsymbol{p}}_t$ according to $x_t(\tilde{\boldsymbol{p}})$.

6 Suffer and observe the loss $L(\tilde{\boldsymbol{p}}_t) = (\boldsymbol{\ell}_t)^\top \tilde{\boldsymbol{p}}_t \leq 1$.

7 Compute the matrix $C_t := \mathbb{E}_{\boldsymbol{p} \sim x_t(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top] \in \mathbb{M}_{D \times D}$.

8 Compute the estimated loss vector $\hat{\boldsymbol{\ell}}_t := L(\tilde{\boldsymbol{p}}_t)\left(C_t^{-1}\tilde{\boldsymbol{p}}_t\right) = \left(\boldsymbol{\ell}_t(\tilde{\boldsymbol{p}}_t)^\top\right) C_t^{-1}\tilde{\boldsymbol{p}}_t$.

9 $\forall \boldsymbol{p} \in S$, $w^{t+1}(\boldsymbol{p}) := w_t(\boldsymbol{p})e^{-\eta(\hat{\boldsymbol{\ell}}_t)^\top \boldsymbol{p}}$.

As in Exp3, at each stage t, ComBand keeps a weight $w_t(\boldsymbol{p})$ for each action $\boldsymbol{p} \in S$ and it samples an action (line 5—Algorithm 2) from a distribution, called x_t, mixing between the *exploitation distribution* ν_t (normalization of the action weights) and an *exploration distribution* μ (unchanged over time). An unbiased estimator $\hat{\boldsymbol{\ell}}_t \in [0,1]^E$, based on the *co-occurrence* matrix $C_t := \mathbb{E}_{\boldsymbol{p} \sim x_t(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top] \in \mathbb{M}_{E \times E}$, is used to estimate the loss vector $\boldsymbol{\ell}_t$. Then, the action weights are updated by the exponential rule using these estimated losses (line 9—Algorithm 2).

In ComBand, the exploration distribution μ is chosen a priori as inputs and it can be any arbitrary distribution on S such that S is spanned by the support of μ. Importantly, the performance guarantee of ComBand depends directly on the choice of μ; to highlight this, we parameterize ComBand with μ and use the notation ComBand(μ). Consider the matrix $M(\mu) = \mathbb{E}_{\boldsymbol{p} \sim \mu}\left[\boldsymbol{p}\boldsymbol{p}^\top\right]$, we denote by $\lambda^*[M(\mu)]$ the *smallest nonzero eigenvalue* of $M(\mu)$ and let $n := \max\{\|\boldsymbol{p}\|_1, \boldsymbol{p} \in S\}$. An upper-bound of the expected regret provided by ComBand is given as the following proposition, extracted from Cesa-Bianchi and Lugosi (2012) and rewritten here under our notations.

Proposition 2.2.5. *In any* OComb problem with bandit feedback and an arbitrary action set $S \subset \{0,1\}^D$, the ComBand(μ) algorithm with appropriate parameters γ and η,[12] yields an expected regret

$$R_T \leq 2\sqrt{\left[\frac{2n}{D\lambda^*[M(\mu)]} + 1\right] TD \log(|S|)}.$$

Here, the regret upper-bound provided by ComBand algorithm is in a logarithmic order of $|S|$; this has improves much further the bound provided by applying directly the classical Exp3 to OComb. Moreover, the larger $\lambda^*[M(\mu)]$ is, the better the regret bound that ComBand(μ) guarantees. The problem of optimizing μ and $\lambda^*[M(\mu)]$ in OComb with bandit feedback remains an open question in general (see Cesa-Bianchi and Lugosi (2012) for several positive examples).

[12]See Cesa-Bianchi and Lugosi (2012) for the choices on these parameters.

ComBand runs in $\Omega(|S| \cdot T)$ where T is the time horizon. Since $|S|$ is exponential in terms of D—the dimension of the action vector, it is inefficient to implement ComBand. This is due to the *weights-updating step* (line 9—Algorithm 2), the *sampling step* (line 5—Algorithm 2) and the computation of the *co-occurrence matrix* (line 7—Algorithm 2). We will revisit the ComBand algorithm in Chapter 9 and provide alternative procedures to improve these steps for an application of ComBand in the model of Colonel Blotto game under online setting with bandit feedback.

2.2.3 Online Shortest Path Problems and Weight Pushing

Now we review the online shortest path problem (OSP)—another important instance of the OComb framework. We will use the OSP model as the basis for our studies on the online discrete CB game (presented in Part II). Note here that the term "online shortest path" (adopted from György et al. (2007)) is often used interchangeably in the literature of bandit problems with the term "path planning problem" (proposed by Cesa-Bianchi and Lugosi (2006)). In this thesis, we choose to use the former because it provides a more direct intuition and because the latter may create unnecessary confusions with other similar terms in other communities.

Briefly put, OSP is an OComb problem where the action set of the learner is a set of paths (from a source to a destination) on a directed acyclic graph (henceforth, DAG). The original application of the OSP model (see György et al. (2007) and Takimoto and Warmuth (2003)) is the prediction problem in routing where at each stage, a decision-maker needs to choose a route for sending a packet in a network. Due to the traffic at each stage, each edge in the network has a different delay that is unknown when making the decisions. The objective is to minimize the aggregate delays incurred on the edges along the chosen route.

We now give a formal definition of OSPs with a different terminology from the ones used in OComb. First, we introduce several notations related to the DAGs. Each instance of OSP will be defined with a DAG, denoted by G, that has the following properties: there are two special vertices, a source and a destination, that are respectively called s and d; we denote by $\mathcal{P}$ the set of all *paths* starting from s and ending at d and let $P := |\mathcal{P}|$; the set of vertices and the set of edges of G will be respectively denoted by $\mathcal{V}$ and $\mathcal{E}$; let $V := |\mathcal{V}| \geq 2$ and $E := |\mathcal{E}| \geq 1$; we assume that each edge $e \in \mathcal{E}$ belongs to at least one path $\boldsymbol{p} \in \mathcal{P}$. Both the notations $e \in \boldsymbol{p}$ and $\boldsymbol{p} \ni e$ indicate that the edge $e \in \mathcal{E}$ belongs to the path $\boldsymbol{p} \in \mathcal{P}$. Finally, in G, we denote by n the length of the longest path in $\mathcal{P}$, that is $\|\boldsymbol{p}\|_1 \leq n, \forall \boldsymbol{p} \in \mathcal{P}$.

Definition 2.2.6. *Given a time horizon $T \in \mathbb{N}$, the* ***online shortest path problem*** *(OSP) on the graph G is described as follows: at stage $t \in [T]$, each edge $e \in \mathcal{E}$ is embedded with a scalar loss $\boldsymbol{\ell}_t(e) \in [0,1]$ that is generated by an adversary; without knowing these losses, the learner chooses a path $\tilde{\boldsymbol{p}}_t \in \mathcal{P}$; the learner's incurred loss at this stage, denoted by $L_t(\tilde{\boldsymbol{p}}_t)$, is the sum of the losses from the edges belonging to $\tilde{\boldsymbol{p}}_t$, i.e., $L_t(\tilde{\boldsymbol{p}}_t) = \sum_{e \in \tilde{\boldsymbol{p}}_t} \boldsymbol{\ell}_t(e)$. At the end of stage t, the learner observes some feedback. The learner's objective is to minimize the expected regret.*

The OSP defined here is an instance of OComb (with the dimension of the action vectors is $D := E$); this can be seen by mapping each path $\boldsymbol{p} \in \mathcal{P}$ to a vector in $\{0,1\}^E$ whose e-th coordinate is $\boldsymbol{p}(e) = 1$ if and only if the edge $e \in \boldsymbol{p}$ (thus, $\mathcal{P} \subset \{0,1\}^E$) and setting the loss vector to be $\boldsymbol{\ell}_t = (\boldsymbol{\ell}_t(e))_{e\in\mathcal{E}}$. All the concepts related to OComb will be transferred to OSPs, including the regret, the expected regret, the oblivious/non-oblivious adversary and the settings of the feedback observed by the learner at the end of each stage. The intuitions for the three basic feedback setting in OSPs are given as follows:

> The *full-information* setting: at the end of stage t, the learner observes the losses of all edges, i.e., $\boldsymbol{\ell}_t(e), \forall e \in \mathcal{E}$.
>
> The *semi-bandit* feedback setting: at the end of stage t, the learner only observes the losses of the edges belonging to the chosen path, i.e., $\boldsymbol{\ell}_t(e), \forall e \in \tilde{\boldsymbol{p}}_t$.
>
> The *bandit* setting: at the end of stage t, the learner only observes the aggregate loss on the chosen path; she does not knows precisely the loss embedded with any edge.

The semi-bandit with side-observation setting has not been considered in the literature of OSPs; we formally define it and give more details in Chapter 8.

Finally, we review an important technique that relates to the OSPs, called **weight pushing**. There exist variants of the Exp3 algorithm that are modified to run in the OSPs (we analyze them in detail in Chapters 8 and 9). In these algorithms, it is needed to have a sampling step as follows: given a DAG G, given a weight $w(e) > 0$ for each edge $e \in \mathcal{E}$; one needs to sample a path $\tilde{\boldsymbol{p}} \in \mathcal{P}$ with the probability:

$$x(\tilde{\boldsymbol{p}}) := \left[\prod\nolimits_{e\in\tilde{\boldsymbol{p}}} w(e)\right] \Big/ \left[\sum\nolimits_{\boldsymbol{p}\in\mathcal{P}} \prod\nolimits_{e'\in\boldsymbol{p}} w(e')\right]. \tag{2.6}$$

Note that the number of paths in a DAG G (denoted $P = |\mathcal{P}|$) is an exponential number in terms of the number of edges ($E = |\mathcal{E}|$) and the number of vertices ($V = \mathcal{V}$). Therefore, a direct computation and sampling from $x(\tilde{\boldsymbol{p}}), \forall \tilde{\boldsymbol{p}} \in \mathcal{P}$ takes $\mathcal{O}(P)$ time, which is very inefficient. It is desired to have a more efficient procedure (running in polynomial time in terms of E and V). To do this, the weight pushing technique was introduced by György et al. (2007) and Takimoto and Warmuth (2003); it is based on dynamic programming. We present this technique, under our notation, as a collection of two algorithms, called the WP and WPS algorithms, described below.

Let us respectively denote by $\mathbb{C}(u)$ and $\mathbb{F}(u)$ the set of the direct successors and the set of the direct predecessors of any vertex $u \in \mathcal{V}$. Moreover, let $e_{[u,v]}$ and $\mathcal{P}_{u,v}$ respectively denote the edge and the set of all paths from vertex u to vertex v. We then define the following terms for each pair of vertices $u, v \in \mathcal{V}$:

$$H(u,v) := \sum_{\boldsymbol{p}\in\mathcal{P}_{u,v}} \prod_{e\in\boldsymbol{p}} w(e).$$

Intuitively, $H(u,v)$ is the aggregate weight of all paths from vertex u to vertex v and $H(s,d)$ is exactly the denominator in (2.6). As a convention of notation, we let

$H(u,u) = 1, \forall u \in \mathcal{V}$ and $H(u,v) = 0$ if $\mathcal{P}_{u,v} = \emptyset$. Now, let us label the vertices set by $\mathcal{V} = \{s = u_0, u_1, \ldots, d = u_{V-1}\}$ such that if there exists an edge connecting u_i to u_j then $i < j$. The computation of all these terms $H(u_i, u_j)$ for any $i, j \in \{0, 1, \ldots, V-1\}$ can be done recursively by the following algorithm, called WP algorithm (i.e., Algorithm 3), that runs in $O(EV^2)$ time, through dynamic programming.

Algorithm 3: The WP Algorithm.

Input: Graph G, set of weights $\{w(e), e \in \mathcal{E}\}$.
Output: $H(u_i, u_j), \forall i, j \in [V-1]$ (that are $H(u,v), \forall u, v \in \mathcal{V}$).

```
for j ∈ {V − 1, V − 2, ..., 0} do
    Initialization H(u_j, u_j) := 1.
    for i ∈ {j − 1, ..., 0} do
        H(u_i, u_j) := Σ_{v∈ℂ(u_i)} w(e_[u_i,v]) H(v, u_j).
```

Algorithm 4: The WPS Algorithm.

Input: Graph G, set of weights $\{w(e), e \in \mathcal{E}\}$.
Output: $\tilde{\boldsymbol{p}} \in \mathcal{P}$ sampled from (2.6).

```
H(u, d), ∀u ∈ V are computed by Algorithm 3.
Initialize Q := {s}, vertex u := s.
while u ≠ d do
    Sample a vertex v from ℂ(u) with probability w(e_[u,v]) H(v, d) / H(u, d).
    Add v to the set Q and update u := v.
Set p̃ ∈ P to be the path going through all the vertices in Q.
```

Based on the WP algorithm (i.e., Algorithm 3), we construct the WPS algorithm (i.e., Algorithm 4) that uses the weights $w(e), e \in \mathcal{E}$ as inputs and randomly outputs a path in $\mathcal{P}$. Intuitively, starting from the source vertex $s = u_0$, Algorithm 4 sequentially samples vertices by vertices based on the terms $H(u,v)$ computed by Algorithm 3. It is noteworthy that Algorithm 4 also runs in $O(E)$ time and it is trivial to prove that the probability that a path $\boldsymbol{p}$ is sampled from Algorithm 4 matches exactly $x(\boldsymbol{p})$ defined in (2.6).

Part I

ONE-SHOT COMPLETE-INFORMATION RESOURCE ALLOCATION GAMES—APPROXIMATE EQUILIBRIA OF BLOTTO GAMES

Chapter 3

Blotto Games—Formulation and Related Works

Some of the ideas and definitions presented in this chapter have previously appeared in our publications Vu, Loiseau, and Silva (2018a,b) and in our pre-print article Vu, Loiseau, and Silva (2019a).

In this part (that includes Chapters 3, 4, 6 and 5), we tackle the first key question of this thesis: how to play strategically in a one-shot complete-information Colonel Blotto game (and other Blotto games) to obtain a good guarantee on payoffs? On this question, the main focus of the literature is to look for Nash equilibria of these games. In this thesis, we take a different perspective and tackle Blotto games from another angle. We look for *approximate equilibria* of Blotto games with simple (and efficient) constructions such that the involved approximation error is well-controlled and look for conditions under which this error is negligible. Our results are scalable and they extend the scope of applications of Blotto games to large-scale problems in practice. Besides these results, we also aim to investigate the generalizability of the obtained solutions into the general class of resource allocation games. Throughout this part of the thesis, we refer to one-shot complete-information strategic games simply as "games" (or "offline games" in order to distinguish with the model of "online learning in games" studied in Part II).

Before diving into our results, we first dedicate this chapter (Chapter 3) to formally define the model of the Colonel Blotto game and its variants/extensions and then give an overview on the state-of-the-art in studying these games. The detailed outline is as follows: in Section 3.1, we introduce the generalized Colonel Blotto game; in Section 3.2 we present several other variants and extensions, including the discrete Colonel Blotto game, the generalized Lottery Blotto game and the generalized-rule Colonel Blotto game; finally, in Section 3.3 we review the literature and the challenges encountered in characterizing equilibria of these games and discuss more broadly related works.

3.1 The Colonel Blotto Game

Colonel Blotto (CB game) is a famous game-theoretic model with a large range of applications (we invite the readers to see our discussions in Chapter 1 for a list of applications of this game). For an intuition before defining the game formally (in Section 3.1.1), we present in detail here a motivational example—an advertising competition that can be modeled as a CB game:[1]

Two marketing campaigns, each with a fixed budget (in the form of promotional gifts, coupons, etc.), want to optimize their direct marketing strategies on a common set of potential customers (i.e., how to distribute the gifts to the customers). To each marketing campaign, each customer has a certain value—the likeliness of buying the product. We assume that these values are known (or might be estimated precisely enough).[2] Simultaneously, the marketing campaigns distribute their promotional gifts towards the customers; and each customer prefers the product of the campaign giving him/her more valuable gifts. The total payoff gained by each marketing campaign is the aggregate values of customers choosing its products.

This is a typical application of the CB game. Two important features involved in this application that are also characteristics of the CB game are: (i) customers choose the campaign giving them the better gifts, regardless of the magnitude of the difference between values of the gifts; (ii) unused resources, i.e., gifts that are not given away, do not contribute to the final payoffs; this is the use-it-or-lose-it rule. Finally, to optimize its payoff, each marketing campaign needs to take into account the optimization of its opponent; therefore, knowing stable states of the game has a significant contribution in predicting players' behaviors in practice.

3.1.1 The Generalized Colonel Blotto Game

We consider the following one-shot, complete information game between two players A and B. Each player has a fixed amount of resources (called the *budgets*), denoted X^A and X^B, respectively. Without loss of generality, we assume that $0 < X^A \leq X^B$. Players simultaneously allocate their resources across n *battlefields* ($n \geq 3$). Each battlefield $i \in [n]$ is embedded with two parameters $w_i^A, w_i^B > 0$, corresponding to the *values* at which player A and player B respectively assess this battlefield. A *pure strategy* of player $\phi \in \{A, B\}$ is a vector $\boldsymbol{x}^\phi = \left(x_i^\phi\right)_{i \in [n]} \in \mathbb{R}^n_{\geq 0}$ that satisfies the budget constraint $\sum_{i=1}^n x_i^\phi \leq X^\phi$. In each battlefield i, when player ϕ allocates strictly more than her opponent, she gains her embedded value w_i^ϕ while the opponent gains 0. In cases of a tie, i.e., if $x_i^A = x_i^B$, then player A receives αw_i^A and player B receives $(1-\alpha) w_i^B$, where

[1] This motivational example is adopted from the problems studied by Masucci and Silva (2014, 2015) in advertising competitions on social networks (also modeled as CB games) where a customer's value is the aggregate between an intrinsic value and a network value (i.e., the influence on his/her peers in the network). Here, we simplify this model and do not include the network values to the customers' values.

[2] See e.g., Domingos and Richardson (2001) and Richardson and Domingos (2002) on how marketing campaigns may determine customers' values in the case of advertising on social networks.

$\alpha \in [0,1]$ is a fixed parameter. Each player's payoff is the summation of values she gains from all battlefields; formally, for any pure strategy profile $(\boldsymbol{x}^A, \boldsymbol{x}^B)$, the *payoffs* of players A and B are $\Pi^A(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum_{i=1}^n w_i^A \cdot \beta^A\left(x_i^A, x_i^B\right)$ and $\Pi^B(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum_{i=1}^n w_i^B \cdot \beta^B\left(x_i^A, x_i^B\right)$ respectively; here, β^A and β^B (henceforth, we called them the Blotto functions) are functions defined as follows:

$$\beta^A\left(x,y\right) = \begin{cases} 1 \text{, if } x > y \\ \alpha \text{, if } x = y \\ 0 \text{, if } x < y \end{cases} \quad \text{and} \quad \beta^B\left(x,y\right) = \begin{cases} 1 \text{, if } y > x \\ 1-\alpha \text{, if } y = x \\ 0 \text{, if } y < x \end{cases}, \text{ for all } x, y \in \mathbb{R}_{\geq 0}. \tag{3.1}$$

Definition 3.1.1. ***A generalized Colonel Blotto game***, *denoted by $\mathcal{CB}_n$, is the game described above; in particular, the action set of player $\phi \in \{A,B\}$ is $\{\boldsymbol{x}^\phi \in \mathbb{R}^n_{\geq 0} : \sum_{i=1}^n x_i^\phi \leq X^\phi\}$ and her payoff is $\Pi^\phi(\boldsymbol{x}^A, \boldsymbol{x}^B)$ when players A and B play the pure strategies $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$.*

In this game, a *mixed strategy* is a joint distribution on the allocations of all battlefields, such that any drawn pure strategy of a player is an n-tuple that satisfies her budget constraint. We reuse the notations $\Pi^A(\sigma_A, \sigma_B)$ and $\Pi^B(\sigma_A, \sigma_B)$ to denote the payoffs of players A and B when they play the mixed strategies σ_A and σ_B, respectively. It is convenient to sometimes work with the notion of the *normalized values* of the battlefields, defined as $v_i^A := w_i^A / W^A$ and $v_i^B := w_i^B / W^B$, where $W^A := \sum_{j=1}^n w_j^A$ and $W^B := \sum_{j=1}^n w_j^B$ for $i \in [n]$. Intuitively, W^A and W^B are the total values that players could access on battlefields; they are also upper-bounds of the maximum payoffs that players A and B can possibly obtain. We trivially observe that $v_i^\phi \in [0,1], \forall i \in [n]$ and that $\sum_{j=1}^n v_j^\phi = 1$.

We emphasize again that the definition of $\mathcal{CB}_n$ given above allows the asymmetry in the players' budgets and the heterogeneity in the battlefields values; moreover, it allows battlefield values to differ between the two players. Due to this asymmetry in players' assessment on battlefields' values, the summation of players' payoffs in $\mathcal{CB}_n$ may vary according to the chosen strategies. For instance, if player A wins all the battlefields, the total payoff of two players is W^A, while this total payoff is W^B if player B wins all the battlefields. In other words, without other assumptions, *the generalized* CB *game is a non-constant-sum game*. Furthermore, note that the term "generalized CB game" used in this definition is adopted from Kovenock and Roberson (2015); however, our formulation in Definition 3.1.1 is even more general than theirs: we include a general tie-breaking rule. The defined payoff functions (involving β^A and β^B) can be understood as if we randomly break the tie (if it happens) such that player A wins battlefield i with probability α while player B wins it with probability $(1-\alpha)$. This includes all the classical tie-breaking rules considered in the literature; for instance, the rule of giving the whole value to player B used by Roberson (2006) and Schwartz et al. (2014) corresponds to $\alpha = 0$; the 50-50 rule used by Ahmadinejad et al. (2016), Behnezhad, Dehghani, et al. (2017), and Kovenock and Roberson (2015) corresponds to $\alpha = 1/2$.

Now, we recall the advertising competition presented in the beginning of this section as an example on how one might use the generalized CB game to model practical situations. We see clearly that it can be modeled as a $\mathcal{CB}_n$ game: players correspond to the marketing campaigns; each customer is a battlefield and its value is the customer's value; the rule on how a customer chooses a product is given by the Blotto functions (β^A, β^B) and the marketing campaigns' payoffs are consistent with the winner-takes-all rule of the generalized CB game.

Hereinafter, in places with no ambiguity, we drop the term generalized and simply address the game $\mathcal{CB}_n$ as the CB game. Note finally that to lighten the notation, we only include the subscript n—the number of battlefields—in the notation $\mathcal{CB}_n$ and omit the other parameters; in particular the values X^A, X^B, α and w_i^A, w_i^B for $i \in [n]$. We will discuss the state-of-the-art results related to equilibrium characterization of the generalized CB game in Section 3.3.1. Before moving to definitions of other Blotto games, we first present an important special case of the $\mathcal{CB}_n$ game in Section 3.1.2.

3.1.2 The Constant-sum Colonel Blotto Game

As discussed, with general configurations of parameters, the game $\mathcal{CB}_n$ defined in Definition 3.1.1 is a non-constant-sum game. However, most works in the literature (a review is given in Section 3.3.1) focus only on the *constant-sum* variant of this game where players have the same evaluations on battlefields' values. Naturally, all our results for $\mathcal{CB}_n$ can be straightforwardly applied to this constant-sum version as well. However, for the purpose of comparing with the literature and because we can show stronger results in this special case, it is useful to also formally define the constant-sum game variant as follows:

Definition 3.1.2. ***A constant-sum Colonel Blotto game****, denoted by $\mathcal{CB}_n^C$, is a game that has the same formulation as the game $\mathcal{CB}_n$ but with the additional condition that $w_i^A = w_i^B, \forall i \in [n]$.*

Hereinafter, in the game $\mathcal{CB}_n^C$, we use the notation w_i to commonly address the values $w_i^A = w_i^B$. Similarly, we denote the common normalized valuation on battlefields by $v_i = v_i^A = v_i^B$ for all $i \in [n]$ and the value of the game (i.e., the total payoffs of players) by $W := W^A = W^B$.

3.2 Other Blotto Games

In this section, we introduce several variants and extensions of the generalized CB game presented in the previous section. They model a larger set of practical situations and provide additional open challenges to study. Henceforth, we often commonly refer to these games (including the generalized CB game) as Blotto games.

3.2.1 The Discrete Colonel Blotto Game

First, there exist real-world situations where the involved resources are indivisible. For example, in security problems, players (the attacker and the defender) need to commit their human forces to protect (or destroy) security targets; these allocations of resources must be integer numbers. In political competitions for voters, the budgets and the resource allocations are also sometimes required to be rounded up to integers (see e.g., Behnezhad, Blum, et al. (2018) and Behnezhad, Dehghani, et al. (2017)). Another example is the military logistic application (proposed by Gross (1950) and Gross and Wagner (1950)), in this case, if resources of players correspond to soldiers (i.e., troops), these integer constraints are also essential. These situations cannot be modeled directly by the $\mathcal{CB}_n$ game presented in Definition 3.1.1 in which the allocation of a player $\phi \in \{A, B\}$ to a battlefield can be an arbitrary real number in $[0, X^{\phi}]$. To capture situations involving indivisible resources, we introduce the discrete Colonel Blotto game (hereinafter, the DCB game), where players' budgets and allocations are constrained to be integers.

Definition 3.2.1. ***A discrete Colonel Blotto game*** *with n battlefields, denoted by $\mathcal{CB}_n^D$, is the game that has the same formulation as the game $\mathcal{CB}_n$ but with additional conditions that $X^A, X^B \in \mathbb{N}\backslash\{0\}$ and that the strategy set of player $\phi \in \{A, B\}$ is*

$$\left\{ (x_1^{\phi}, \ldots, x_n^{\phi}) : x_i^{\phi} \in \mathbb{N}, \forall i \in [n] \text{ and } \sum_{j \in [n]} x_j^{\phi} \leq X^{\phi} \right\}.$$

Naturally, other notations and related concepts of the generalized CB game ($\mathcal{CB}_n$), such as $\Pi^A, \Pi^B, v_i^A, v_i^B, W^A, W^B$, are transferred accordingly to the discrete CB game ($\mathcal{CB}_n^D$). Moreover, by adding the constraint requiring that players have the same evaluation on the battlefields' values into the game $\mathcal{CB}_n^D$, one can obtain the *constant-sum variant of DCB*. As in the case of the generalized CB game, most of works on the DCB game only focuses on this constant-sum variant. Unlike the $\mathcal{CB}_n$ game, the main challenge in the $\mathcal{CB}_n^D$ game lies in the complexity of its solution, this is due to the fact that the number of pure strategies of a player in $\mathcal{CB}_n^D$ is exponential in terms of the number of battlefields and the budgets. We give a more complete literature review on $\mathcal{CB}_n^D$ in Section 3.3.2.

3.2.2 The Generalized Lottery Blotto Game

In practice, there exist situations where the winner-takes-all rule of the CB game (i.e., the player who wins a battlefield gains totally the value) is too restrictive. In order to model these situations with more flexibility, in this thesis we also study an extension of the generalized CB game, called the *generalized Lottery Blotto game* (henceforth, LB game),[3] where each player only gains a part of her value in each battlefield. Alternatively, one can interpret the LB game as a version of the CB game

[3]Note that the generalized LB game is also a non-constant-sum game.

in which each player wins a battlefields' value with a certain probability depending on players' allocations on that battlefield and this probability can be non-zero even for the player with smaller allocations. Some examples where the LB game model are useful include online advertising competitions, political contests for voters' attention, research and development activities and radio-wave transmissions with noises.

We formulate the generalized LB game by presenting the players' payoffs based on the concept of contest success function (henceforth, CSF). CSFs, studied profoundly in the rent-seeking literature (see e.g., Corchón (2007) and Skaperdas (1996)) are functions that quantify the winning probability in *contests*, also called *rent-seeking* competitions, where several players compete for a single prize by exerting resources/efforts. CSFs can be defined for any number of players (see e.g., a general definition by Skaperdas (1996)), but in this work, we focus only on the case of two players.

Definition 3.2.2. *$\zeta_A : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ and $\zeta_B : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ are a pair of* ***contest success functions*** *(CSFs) if and only if the following two conditions are satisfied:*

(C1) $\zeta_A(x,y), \zeta_B(x,y) \geq 0$ *and* $\zeta_A(x,y) + \zeta_B(x,y) = 1, \forall x, y \geq 0.$

(C2) $\zeta_A(x,y)$ *(resp. $\zeta_B(x,y)$) is non-decreasing in x (resp. in y) and non-increasing in y (resp. in x).*

Intuitively, the function ζ_A (resp. ζ_B) maps any pair of players' invested resources to the probability that player A (resp. player B) might win the prize. Condition $(C1)$ indicates that the outputs of any pair of CSFs always satisfy the condition of a probability distribution. On the other hand, Condition $(C2)$ states that a player's winning probability increases (or at least stays the same) when she increases her effort and decreases (or at least stays the same) when her opponent increases her effort. We note that Definition 3.2.2 allows a more general definition of CSFs (in two-player cases) compared to the definitions given by Clark and Riis (1998b), Hirshleifer (1989), and Skaperdas (1996) that contain other assumptions.[4] While many of the CSFs considered in the literature are continuous functions, we do not include continuity requirement in Definition 3.2.2 to keep the generality. Importantly, the Blotto functions β_A, β_B of the game $\mathcal{CB}_n$ (i.e., the winner-takes-all rule defined in (3.1)) satisfy Conditions $(C1)$ and $(C2)$, hence β_A, β_B are CSFs. Besides these functions, some examples of other CSFs considered in the literature are:

1. The Tullock CSF, first proposed by Friedman (1958) and re-introduced later by Tullock (1980):

$$\zeta_A(x,y) = x/(x+y) \text{ and } \zeta_B(x,y) = y/(x+y); \tag{3.2}$$

[4]For example, Skaperdas (1996) defines ζ_A, ζ_B with an axiom of anonymity; they also require that any player who puts a strictly positive amount of resources has a strictly positive probability of winning the prize; Clark and Riis (1998b) considers the CSFs additionally satisfying the Choice Axiom. These are technical conditions needed for proving their results and we omit them here lest they unnecessarily limit our scope of study.

2. $\zeta_A(x,y) = \max\left\{\min\left\{\frac{1}{2}+C(x-y),1\right\},0\right\}$ and $\zeta_B(x,y) = 1-\zeta_A(x,y)$, proposed by Che and Gale (2000), where $C > 0$ is a fixed parameter;

3. $\zeta_A(x,y) = \frac{1}{2}-\frac{y-x}{2y}$ if $x \leq y$ and $\zeta_A(x,y) = \frac{1}{2}+\frac{x-y}{2x}$ if $x \geq y$; and $\zeta_B(x,y) = 1-\zeta_A(x,y)$, proposed by Alcalde and Dahm (2007).

Building on the notions of CSF and the Colonel Blotto game, we now define a new game model based on the following idea: in a game $\mathcal{CB}_n$, we view each battlefield as a contest between players where the prize is the battlefield's value and players' effort correspond to their allocations; by doing this, each pair of CSFs defines an instance of a new game where the probability of winning a battlefield follows them accordingly.

Definition 3.2.3. *Let $\zeta = (\zeta_A, \zeta_B)$ be a pair of CSFs.* ***A generalized Lottery Blotto game*** *with n battlefields, denoted $\mathcal{LB}_n(\zeta)$, is the game with the same players A and B and the same strategy sets as in $\mathcal{CB}_n$; but where payoffs are given, for any pure strategy profile $(\boldsymbol{x}^A, \boldsymbol{x}^B)$, by*

$$\Pi^A_\zeta(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum\nolimits_{i=1}^n w_i^A \cdot \zeta_A\left(x_i^A, x_i^B\right) \qquad \textit{and} \qquad \Pi^B_\zeta(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum\nolimits_{i=1}^n w_i^B \cdot \zeta_B\left(x_i^A, x_i^B\right).$$

The generalized Lottery Blotto game model is more flexible than that of the generalized Colonel Blotto game, as it allows choosing CSFs that define the players' payoffs for each specific practical situation. Intuitively, the players' payoffs in a Lottery Blotto game can be seen as the expected payoffs in the Colonel Blotto game with respect to the following random process determining the winner in any battlefield i: player A wins with probability $\zeta_A(x_i^A, x_i^B)$ and player B wins with probability $\zeta_B(x_i^A, x_i^B)$ if they allocate x_i^A and x_i^B respectively. As in the game $\mathcal{CB}_n$, players' payoffs in the $\mathcal{LB}_n(\zeta)$ game are monotonic with respect to the allocations in a battlefield (due to Condition (C2)). Note that the definition of the CSF (Definition 3.2.2) that we adopt here also includes the Blotto-rule functions β^A, β^B (see (3.1)) as a special case; therefore, *the CB game is a particular case of the LB game*. Throughout the thesis, for a generalized Lottery Blotto game $\mathcal{LB}_n(\zeta)$ (where ζ does not coincide with the Blotto-functions), we call a generalized Colonel Blotto game $\mathcal{CB}_n$ to be the corresponding game of $\mathcal{LB}_n(\zeta)$ (and vice versa) if they have the same parameters $n, X_A, X_B, w_i^A, w_i^B, \forall i \in [n]$.

In the literature, the terms "loterry Blotto" or "Blotto-type game with lottery CSF" are used in several works, but only to indicate the LB game with the Tullock CSF (defined as (3.2)). To avoid the confusion, we emphasize again that we use the term (generalized) Lottery Blotto game to indicate the game with any generic CSF. Henceforth, in places without ambiguity, we address a generalized LB game with n battlefields and a generic CSFs simply by the notation $\mathcal{LB}_n$ (that is we drop the notation of the involved CSFs). We also remark that a possible extension of the $\mathcal{LB}_n$ game is to allow the winner of each battlefield to be determined by a different pair of CSFs. In this thesis, we only consider the version, as defined in Definition 3.2.3, where one pair of CSFs is used for all the battlefields, because it is simpler and more tractable.

As in the generalized CB game, the equilibrium characterization of the generalized LB game is an open question (see also Section 3.3.3 for a literature review). In Chapter 6, we investigate the connection between the CB and LB games; then, based on this connection, we propose a class of approximate equilibria for the generalized LB game.

3.2.3 The Generalized-Rule Colonel Blotto Game

In this section, we introduce yet another extension of the Colonel Blotto game, called the generalized-rule Colonel Blotto game (henceforth, the GR-CB game), where the winner-determination rule is generalized in order to capture more complicated scenarios often found in practice. In several cases of applications, players may have resources committed to the battlefields before the CB game begins—we call them the *pre-allocations*. This can be found in R&D contests, e.g., companies can use their current products/technology to gain advantage when starting to develop new ones. Pre-allocations are also very common in lobbying, e.g., in the competition for the rights of operating 5G networks in Europe in 2020, Huawei Technologies Co., Ltd. received disadvantages due to political reasons (see e.g., the articles by Reichert (2020) and Stevis-Gridneff (2020)); here, the disadvantage of one player can be interpreted as either she has a negative pre-allocation, or as if her opponent has positive pre-allocations. On the other hand, the *effectiveness of players' resources* are not always symmetric, e.g., in airport-surveillance problems, it usually required more than one security agent to patrol one security target while it may only require one terrorist to make a successful attack. Effectiveness may also vary among battlefields; e.g., in the US presidential election, it is well-known that California is a reliable Democratic state; therefore, with the same budget, the Republican party can attract voters more effectively in the swing-states (e.g., Virginia) rather than in California. An example possibly involving both the pre-allocations and the asymmetric resource's effectiveness is the application of the CB game in military logistics: before a military operation commences, it is often the case that one side (or both sides) has already installed military forces in battlefields; moreover, the effectiveness of resources (equipment, soldiers, etc.) are different among the sides and they may also vary according to the landscapes/features of the battlefields.

A formal definition of the generalized-rule Colonel Blotto game (GR-CB game) is given as follows:

Definition 3.2.4. ***A generalized-rule Colonel Blotto game*** *with n battlefields (denoted $\mathcal{GR}-\mathcal{CB}_n$) is the game with the same players (A and B) and the same strategy sets as in the generalized CB game $\mathcal{CB}_n$ but players' payoffs when they play the pure strategies $\boldsymbol{x}^A = \left(x_i^A\right)_{i\in[n]}$ and $\boldsymbol{x}^B = \left(x_i^B\right)_{i\in[n]}$ are defined by:*

$$\Pi^A_{\textit{GR-CB}}\left(\boldsymbol{x}^A, \boldsymbol{x}^B\right) = \sum_{i\in[n]} w_i^A \cdot \beta^A\left(x_i^A, q_i x_i^B - p_i\right), \tag{3.3}$$

$$\Pi^B_{\textit{GR-CB}}\left(\boldsymbol{x}^A, \boldsymbol{x}^B\right) = \sum_{i\in[n]} w_i^B \cdot \beta^B\left(x_i^A, q_i x_i^B - p_i\right). \tag{3.4}$$

Here, $p_i \in \mathbb{R}$ and $q_i > 0$ are two parameters additionally embedded with each battlefield $i \in [n]$ (they are known by the players before making decisions), and β^A, β^B are the functions defined in (3.1).

Intuitively, the GR-CB game can be interpreted as a modified version of the generalized CB game where the rule determining the winner in each battlefield is replaced by the following new rule: assuming that players A and B respectively allocate x_i^A and x_i^B to battlefield i; if $x_i^A > q_i x_i^B - p_i$, player A wins and gains the value w_i^A and player B gains 0; reversely, if $x_i^A < q_i x_i^B - p_i$, player B wins and gains w_i^B while player A gains 0; finally, if $x_i^A = q_i x_i^B - p_i$, player A gains αw_i^A and player B gains $(1-\alpha) w_i^B$, where $\alpha \in [0,1]$ is a given parameter. As in CB game, the payoff of each player in this game is the summation of gains she obtains from all the battlefields. On the other hand, the parameters p_i and q_i indicated in the above definition of the $\mathcal{GR}-\mathcal{CB}_n$ game can be intuitively interpreted as follows:

(i) The parameters $p_i, \forall i \in [n]$ are used to model situations where players already have resources committed to the battlefields before the actual beginning of the CB game; we call these committed resources the *pre-allocations* of the players and they are not included in the players' budget X^A and X^B. Then, when the CB game begins, players simultaneously allocate their resources (with budgets X^A, X^B) toward the battlefields (as usual, we call these the *allocations*). After that, in each battlefield, the winner is determined as the player who has a larger total amount of resources at that battlefield, i.e., the summation of the pre-allocation and the allocation. Therefore, if we set p_i to be the difference between player A's allocation and that of player B, the winner-determination rule becomes comparing between the allocation of player A and the allocation of player B minus p_i (when we set $q_i = 1, \forall i \in [n]$). Note that by this interpretation, $p_i > 0$ implies that player A's pre-allocation at battlefield i is larger than that of player B; reversely, $p_i < 0$ implies that player B has a larger pre-allocation than player A.

(ii) On the other hand, the parameters $q_i, \forall i \in [n]$ are used to model the asymmetry in the effectiveness of players' resources. In particular, in the $\mathcal{GR}-\mathcal{CB}_n$ game, in battlefield i, each unit of player B's resource is worth q_i units of player A's resource. Therefore, the winner of battlefield i is player A if her allocation is larger than q_i times the allocation of player B; reversely, if player A's resource is less than q_i times player B's resource then player B is the winner of this battlefield (where we set $p_i = 0, \forall i \in [n]$). According to this interpretation, in battlefield i, if $0 < q_i < 1$, the resource of player A is more effective than that of player B; and reversely, if $q_i > 1$, player B's resource is more effective.

Finally, we note that if $p_i = 0$ and $q_i = 1, \forall i \in [n]$, the $\mathcal{GR}-\mathcal{CB}_n$ game coincides with the generalized CB game $\mathcal{CB}_n$ (see Definition 3.1.1). Moreover, the functions $\zeta_A(x,y) := \beta^A(x, q_i y - p_i)$ and $\zeta_B(x,y) := \beta^B(x, q_i y - p_i)$ satisfy the conditions to be a pair of CSFs; therefore, technically, the GR-CB game is an instance of the generalized LB game. However, due to its special winner-determination rule and the motivation for its formulation, we consider the GR-CB game separately from the class of LB games. Note also that this formulation of the GR-CB game (with both the pre-allocations and asymmetric effectiveness) is novel and we are not aware of any other work in the literature studying a similar model.

3.3 Literature Review on Blotto Games and Other Related Works

As discussed above, most of works in the literature study Blotto games by looking for their Nash equilibria in the one-shot complete information models. In this section, we will sequentially (from Section 3.3.1 to Section 3.3.4) review the state-of-the-art in equilibria characterization of the Blotto games defined in the previous sections. Besides this approach, there also exist other related research directions, either on game-theoretic solutions different from Nash equilibrium or on other game versions obtained from relaxing and modifying Blotto games' formulations; we review some worth-mentioning results in Section 3.3.5.

3.3.1 Equilibria Analyses of the Generalized CB Game

As previously discussed, the Colonel Blotto game has been studied profoundly in the literature; but it mostly focus on the constant-sum variant $\mathcal{CB}_n^C$. Even in this simpler version, the equilibria characterization is still not completely solved. Partial results are obtained in several restricted cases as follows:

- When players have symmetric budgets (i.e., $X^A = X^B$), the equilibria of $\mathcal{CB}_n^C$ are constructed by Borel and Ville (1938) in the game involving three battlefields, and by Gross (1950) and Gross and Wagner (1950) in the game containing any number of battlefields (see also Laslier (2002), Laslier and Picard (2002), and Thomas (2017) for a modern presentation of this solution).

- When players have asymmetric budgets, the equilibria characterization remains an open question in general; the exceptions are the following restricted cases: the game with only two battlefields (see Gross and Wagner (1950) and Macdonell and Mastronardi (2015)), the game with any number of battlefields but homogeneous values, i.e., $w_i = w_j, \forall i, j$ (see Roberson (2006)), and the game where there exists a sufficient number of battlefields of each possible value (Schwartz et al. (2014)).

At a high-level, all the previous works mentioned above are based on the following two-step scheme in order to find the equilibrium of the constant-sum CB game:

Step 1: Determine the optimal univariate distributions of players in each battlefield; i.e., relax the budget constraints to be held only in expectation and seek for the optimal allocation in each battlefield (given that the opponent is doing the same). A formal definition of the concept of optimal univariate distributions is given in Definition 4.1.1.

Step 2: Construct an n-variate joint distribution of the univariate distributions found in Step 1 such that any strategy drawn from this joint distribution satisfies the budget constraint (i.e., it is a mixed strategy).

A toy example is the simple case of the constant-sum CB game with 3 battlefields where players have symmetric budgets (Example 3.3.1).

Example 3.3.1 (Equilibrium of $\mathcal{CB}_n^C$ where $n = 3$, $X^A = X^B$ and $\alpha = 1/2$)**.** *This is clearly a symmetric game; therefore, in equilibrium, players' strategies are the same. First, if there exists a battlefield having the value strictly larger than the sum of the other two battlefields' values, there exists a trivial pure equilibrium: both players allocate all their resource on the battlefield having the maximum value. Each player's equilibrium payoff is* $W/2$.[5]

Otherwise, one can follow the scheme mentioned above:

Step 1: if player $\phi \in \{A, B\}$ *draws her allocation toward battlefield* $i \in \{1, 2, 3\}$ *from the uniform distribution* $\mathcal{U}\left(0, \frac{2X^\phi w_i}{W}\right)$*, then it is optimal for player* $-\phi$ *to do the same; moreover, by doing that, the budget constraint of player* ϕ *holds in expectation (i.e.,* $\mathcal{U}\left(0, \frac{2X^\phi w_i}{W}\right)$ *are the optimal univariate distributions in this game).*

Step 2: If player ϕ *draws her allocation toward battlefield* i*, say* x_i*, independently from* $\mathcal{U}\left(0, \frac{2X^\phi w_i}{W}\right)$*, the summation* $x_1 + x_2 + x_3$ *may exceed the budget. We now describe a procedure for the players to guarantee that their drawn allocations at each battlefield* i *follow* $\mathcal{U}\left(0, \frac{2X^\phi w_i}{W}\right)$ *and that the budget constraints are satisfied. They construct a non-degenerate triangle, called Triangle* T*, such that its sides have the lengths of* w_1, w_2 *and* w_3 *(it can be constructed due to the conditions on values' battlefields). They then inscribe a circle within this triangle and erect a hemisphere upon it (Figure 3.1). Player* $\phi \in \{A, B\}$ *chooses a point on this hemisphere uniformly at random and projects it to a point, namely* M^ϕ*, within the triangle. She plays the strategy* (x_1, x_2, x_3) *such that* $x_1 + x_2 + x_3 = X^\phi$ *and that* $x_1 : x_2 : x_3 = A_1 : A_2 : A_3$ *where* A_1, A_2 *and* A_3 *are the area of the triangles constituted by* M^ϕ *and the sides (with lengths* w_1, w_2, w_3 *respectively) of the Triangle* T*.*

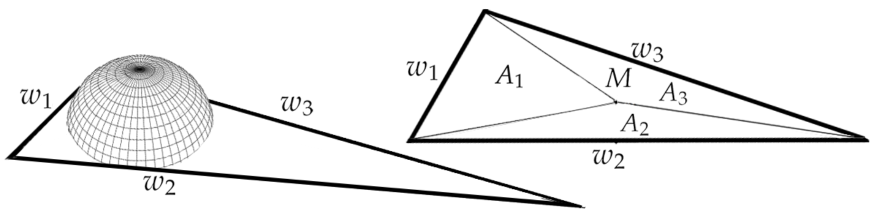


Figure 3.1: Illustration of an equilibrium construction in the $\mathcal{CB}_n^C$ game with three battlefields and $X^A = X^B$.

The equilibrium presented in Example 3.3.1 is called the Disk Solution, proposed by Gross and Wagner (1950); it is illustrated in Figure 3.1. As shown here, even in this extremely simple setting, Step 2 is non-trivial. The main challenges in characterizing

[5]If a player deviate from this, he only gain at most the sum of values of the two small battlefields, that is strictly less than $W/2$.

equilibria of the CB game is how to construct mixed strategies (guaranteeing budget constraints) having the marginals that are optimal in each battlefield. In the cases that are more complicated than Example 3.3.1, the optimal univariate distributions are not simply the uniform distributions and Step 2 is even more challenging. For example, for the constant-sum CB game with homogeneous battlefields (i.e., $w_i = w_j, \forall i, j \in [n]$), Roberson (2006) propose a solution based on Copula theory that has a complicated representation and non-trivial implementation. We encounter similar problems in studying the generalized CB game: we can compute the players' optimal univariate distributions in each battlefield separately but it is unknown whether there exists a mixed strategy constructed from these distributions. It is unlikely that one can extend the solutions from previous works to the generalized $\mathcal{CB}_n$ game because these solutions depend heavily either on the symmetry assumption of the budgets (e.g., Gross (1950), Gross and Wagner (1950), Laslier (2002), and Thomas (2017)) or on the homogeneity of battlefields' values (Roberson (2006) and Schwartz et al. (2014)). We will revisit this discussion and give more details on the univariate optimal distributions in the game $\mathcal{CB}_n$ in Chapter 4.

Besides the results mentioned above for the constant-sum variant, there also exist a few works considering the *non constant-sum* CB game. Kovenock and Roberson (2015) define and study the generalized CB game (as given in Definition 3.1.1) but only with a fixed tie-breaking rule. They provide a set of optimal univariate distributions of this game. They then indicate a sufficient condition for these distributions to become the marginals of an equilibrium[6]—which is identical to that of Schwartz et al. (2014) for the constant-sum case—that only covers a restricted range of games. This leads to the complete characterization of equilibria of the General Lotto game (i.e., a version of the CB game where the budget constraints are relaxed to be held only in expectation—see also Section 3.3.5). Kovenock and Roberson (2015) also show a necessary condition where there is no equilibrium satisfying such a set of marginals in the generalized CB game. On the other hand, several other works also consider restricted variants of the generalized CB game that are *non constant-sum* games, though with a significantly different flavor than our perspective. For example, Hortala-Vallve and Llorente-Saguer (2012) consider a discrete version of the $\mathcal{CB}_n$ game[7] and identify conditions under which a pure Nash equilibrium exists while Kvasov (2007b) and Roberson and Kvasov (2012) consider the relaxation of the use-it-or-lose-it rule that changes the payoffs. As for *approximate equilibria* of the generalized CB game, to the best of our knowledge, there exists no work that explicitly considers this research direction. The only work that mentions approximate results is from Weinstein (2005) who discusses approximate equilibria (with a fixed approximation error) of a variant of the constant-sum CB game with the majority objective (see Section 3.3.5 for a definition) and only 3 battlefields.

[6]Briefly put, this sufficient condition on the attainability of equilibria is that if the set of battlefields are partitioned such that two battlefields are in the same partition if they have the same (normalized) values; then, there exists a sufficient number of battlefields in each partition.

[7]In discrete CB games, there always exists at least one mixed equilibrium because they are finite games.

A characterization of equilibria in the generalized CB game with a general configuration of parameters remains an open question and alternative stable-state solutions, such as an approximate equilibrium, are important for analyzing players' behaviors. Following the latter approach, the leading question is as follows: how to construct approximate equilibria of the generalized CB game such that the involved approximation errors are negligible (especially in large-scale applications)?

Finally, it is worth mentioning that there exist a variety of problems related to the generalized CB game and the *all-pay auction* (APA) is one of the most important classes.[8] Equilibria of APA are often used as tools to study equilibria of the CB game (more specifically, for constructing the optimal univariate distributions in Step 1 of the two-step scheme mentioned above). In an all-pay auction, a number of players secretly decide their bids to compete for a common item; the highest bidder wins the item and gains its value; then, all players pay their bids (regardless of the winner). Nash equilibrium is one of the main focuses of the APA literature and its equilibria have been completely characterized by Baye, Kovenock, and De Vries (1994) and Hillman and Riley (1989) (in the game with any number of bidders). In Section 6.2.1, we will revisit the (single-item) APA in the case of two players and present its formal definition and particular results. We then consider an extension of APA, called the APA with favoritism model, to use as tools to study the generalized-rule CB game. A definition of APA with favoritism is given in Section 6.2.1 and a literature review is presented in Section 3.3.4.

3.3.2 Equilibria Analyses of the Discrete CB Game

First, note that although we present the discrete Colonel Blotto game as a variant of the generalized CB game (see Definition 3.2.1), it is not too uncommon to find in the literature that the DCB game is considered as the primary model and the variant without integer constraints on players' allocations (as in $\mathcal{CB}_n$) is an extension from it. Moreover, in several works, $\mathcal{CB}_n^D$ is simply referred to as the Colonel Blotto game (e.g., Ahmadinejad et al. (2016) and Behnezhad, Dehghani, et al. (2017)) and $\mathcal{CB}_n$ is called the continuous CB game. DCB can also be described quite differently from Definition 3.2.1; one of the most well-known descriptions of the DCB game is from Hart (2008) which is given in a slightly different formulation than our $\mathcal{CB}_n^D$ game: after the players finish their allocation, a battlefield is chosen uniformly at random, the winner of this battlefield wins the whole game (the winner's payoff is 1 and the loser's payoff is -1). This game is equivalent to our definition of $\mathcal{CB}_n^D$ where the battlefields' values are all equal to $1/n$ (and the payoffs are normalized), i.e., a constant-sum discrete CB game with homogeneous battlefields.

Unlike the generalized CB game, the discrete CB is a finite game with the additional integer constraints on the players' allocations and budgets. Therefore, in principle, its equilibrium exists and can be solved numerically in general cases through linear

[8] We defer the literature review on other problems related to the generalized CB game to Section 3.3.5.

programming. However, standard solutions to compute the Nash equilibria in the DCB game face the issue that the strategy space of the players grows exponentially with the number of battlefields and the budgets (see Table 3.1 for some examples).[9] Finding a more efficient computation of its equilibrium is the main challenge in studying the DCB game. This problem has gained traction recently in the algorithmic game theory community and partial results are obtained for the constant-sum variant of the DCB game (in fact, as in $\mathcal{CB}_n$, most of works in the literature only consider the constant-sum variant of $\mathcal{CB}_n^D$). In particular, two algorithms were proposed for this variant, relying on transforming players' strategy sets into linear programming (LP) formulations, that significantly improve the complexity: Ahmadinejad et al. (2016) propose an algorithm based on a reduction to an exponential-size LP and a clever use of the Ellipsoid method to solve it in polynomial time, and Behnezhad, Dehghani, et al. (2017) propose another algorithm that obtains a polynomial-size LP and solves it using the simplex method.[10] Yet, these algorithms still become computationally impractical when the number of battlefields and/or the budgets are large. Applications such as security or politics frequently involve large-scale parameters.

Table 3.1: Number of strategies in several instances of the discrete CB game

Number of battlefields	Budgets	Number of strategies
5	5	126
10	20	10015005
10	50	$> 1.26 \times 10^{10}$
50	10	$> 6.28 \times 10^{10}$

Besides the research direction mentioned above, there are several other results in the literature that also relate to the DCB game. Hart (2008) studies the constant-sum discrete CB game where battlefields' values are homogeneous; however, the author focuses on a special set of uniform distributions and look for the conditions (on the game's parameters) such that these distributions can be obtained; these results are then used to study the bounds of the value of the game in several cases. Similar results can also be found in C. Cohen and Sela (2007) and extended results to the General Lotto game (i.e., where budget constraints are only required to be satisfied in expectation) can be found in Dziubiński (2013). Another approach is proposed by Behnezhad, Blum, et al. (2018) where the main focus is to characterize the (u, p)-maxmin strategy[11] in the constant-sum discrete CB game. Notably, the discrete CB game has also attracted the attention from the behavior game theory community with results using experimental

[9] The number of strategies of a player who has a budget equal k, in a discrete CB game with n battlefields is in the order of $\Omega\left(2^{\min\{n-1,k\}}\right)$.

[10] In a DCB game with n battlefields and players have at most m troops, the complexity of the algorithm from Ahmadinejad et al. (2016) is $\Omega(m^{12}n^4)$ while the algorithm from Behnezhad, Dehghani, et al. (2017) requires to solve an LP with $\Omega(m^2 n)$ constraints and $\Omega(m^2 n)$ variables.

[11] An (u, p)-maxmin strategy of player ϕ is the one that guarantees for ϕ a payoff of at least u with probability at least p, regardless of the strategy of player $-\phi$.

data; e.g., the simulated round-robin experiments from Wittman (2011) and the real plays by humans from Project Waterloo[12]—an application on Facebook that allows users to invite both friends and strangers to play the DCB game against them (the results were collected and analyzed by Kohli et al. (2012)).

There exist algorithms to compute an equilibrium of a DCB game; however, they are still impractical to be implemented for medium and large size instances in practice. The open question is how to efficiently compute a strategy for a player in the DCB game with good guarantees on her payoff? In particular, can we give a fast procedure to construct an approximate equilibrium of DCB and what is the trade-off between the error suffered from using this approximate equilibrium and the obtained improvements in running time?

3.3.3 The Generalized LB Game and Contest Success Functions

Among the instances of the generalized LB game, the game with the Tullock CSF (this CSF is defined in (3.2)) is the one that attracts the most attention from the literature: Friedman (1958) and Kovenock and Roberson (2010) investigate the pure equilibrium of the constant-sum variant; these results are extended into the case with more than two players (Duffy and Matros (2015)), to the games where players asymmetrically assess battlefields' values (G. J. Kim et al. (2018)), and to the case where the surplus is allowed, i.e., unused resources are included into the payoffs (J. Kim and B. Kim (2017)). A similar function to define the winning probabilities is also used by Rinott et al. (2012) to study a variant of the CB game involving sequential tournaments. Moreover, another version of the LB game with Tullock CSFs and the weighted majority objective (i.e., a player wins the game if the aggregate values of battlefields won by her exceeds 50% of the total value) is also studied by Duffy and Matros (2015) and B. Kim and J. Kim (2019); in this case, the equilibrium is partially characterized. An extension of this model with a generalization of the Tullock CSF[13]—belonging to the broader class of ratio-form CSFs—has actually been introduced previously by Shubik and Weber (1981); however, no explicit result on its equilibrium has been given (we also address this game instance in Chapter 6 and call it the LB game with Power-form CSF). The same model is studied by Osório (2013) (coincidentally, it is also called there as the lottery Blotto game); however, only numerically computed approximate-results of the equilibrium are proposed and no tractable close-form solution is provided in the general cases where battlefields' values are asymmetric across players.

To the best of our knowledge, the formulation of the generalized LB game (with generic CSFs) that we define in Definition 3.2.3 is novel and has not appeared in any previous work. However, the idea of considering an extension of the CB game with a CSF in the place of the winner-determination rule can also be found in Kovenock

[12]It used to be available at `http://apps.facebook.com/msrwaterloo/`; however, at the time when this thesis is written, this link is temporarily inaccessible.

[13]In particular, for a given $R > 0$, these CSFs are defined as $\zeta_A(x, y) = x^R/(x^R + y^R)$ and $\zeta_B(x, y) = y^R/(x^R + y^R)$.

and Roberson (2010); note importantly that such an extension has not been explicitly defined there. Kovenock and Roberson (2010) actually consider a larger class of problem, called conflicts with multiple-battlefields (that covers both the LB and CB games); following their terminology, the generalized LB game can be categorized as a conflict with multiple battlefields having the budget-constraint and use-it-or-lose-it cost. Unlike Blotto games, in a conflict, players' strategies may or may not have the structural linkage (e.g., the budget-constraint), moreover, battlefields' outcomes can be defined by a generic CSF and players' payoffs are the difference between their objective functions (not necessary the summation of the gains from the battlefields) and their cost functions. Kovenock and Roberson (2010) summarize the results on the equilibrium of a variety of instances of this generic framework of conflicts. Concerning conflicts with budget constraints, the only results given by Kovenock and Roberson (2010) are either in the CB game (see the related works in Section 3.3.1) or in the LB game with the Tullock CSF that we have discussed above. Additionally, there are many results in other kinds of conflicts defined with these kinds of CSFs, but they are for games where players allocate *without* the budget constraints. In these cases, the equilibrium are partially characterized; however, due to the lacking of budget constraints, these results do not straightforwardly solve the problem of equilibria characterization in the generalized LB game. Some notable results are the multi-item all-pay auction[14] (see also Baye, Kovenock, and Vries (1996) and Hillman and Riley (1989)) and the multi-item contests using the Tullock CSF to determine the outcomes of the battlefields[15] (see e.g., Klumpp and Polborn (2006), Robson (2005), and Snyder (1989)). The idea of Blotto-type games with a generic CSF also appears in Klumpp, Konrad, et al. (2019); however, it is a sequential variant of CB game with the majority rule (see Section 3.3.5 for a definition of this game variant); moreover, only results under a sufficiently-concave assumption on the CSFs are given. The idea of using the generic CSF is also mentioned by Snyder (1989) but without an explicit formulation.

A method analyzing the equilibria of the generalized Lottery Blotto with generic CSFs has not been studied in the literature. Even in the LB game with well-used CSFs, such as ratio-form CSFs,[a] equilibria are also unknown in the case with general configurations of parameters. Alternatively, one can look for an approximate equilibrium of the generalized LB game; in this case, the question becomes how to control the involving approximation error.

[a]See Section 6.1.2 for more details

[14]In the terminology used by Kovenock and Roberson (2010), this is a conflict without budget constraints where battlefields' outcomes are determined by the Blotto-rules and payoffs involve the linear cost.

[15]In the terminology used by Kovenock and Roberson (2010), this is a conflict without budget constraints played on multiple battlefields; battlefields' outcomes are determined by the lottery CSF and the cost function is the linear cost.

3.3.4 The Generalized-Rule CB Game and the All-pay Auction with Favoritism

The generalized-rule Colonel Blotto game (GR-CB game) is obtained by allowing the players in the CB game to have pre-allocations of their resources and to have asymmetric effectiveness. Although there is no work in the literature considering precisely the GR-CB model, an idea that is similar to the sets of additional parameters p_i and q_i ($i \in [n]$) used in the GR-CB game can be found in all-pay auctions with favoritism problems (henceforth, F-APA)—an extension of the all-pay auction framework.[16] In the F-APA, a bidder may have an additive favoritism to include into its bids and a multiplicative favoritism that handicaps its opponents' bids. A formal definition of F-APA is given in Section 6.2.1.

In the following, we give a literature review on the model of all-pay auction with favoritism (also called the APA with head-starts and handicaps; or the APA with incumbency advantages). An equilibrium of the two-player F-APA is characterized by Konrad (2002) but only in the case where players assess the item with the same value, the tie-breaking rule is to share the value equally among the bidders and that both kinds of favoritism are in favor of one player. We will revisit these results in Section 6.2.2. On the other hand, Clark and Riis (2000), Kirkegaard (2012), and Kitahara and Ogawa (2010) focus on the incomplete information game where the item's values to each player is drawn from a distribution and kept as private information. Siegel (2009, 2014) studies the APA with a general transformation in the winner-determination rule where the additive and multiplicative favoritism are functions of players' bids; however, these works focus on an axiomatic approach; particularly, defining a set of assumptions such that an equilibrium may be constructed from an algorithm. Besides the results considering the favoritism as exogenous factors; the model where an auctioneer needs to decide a favoritism (mostly, only with the additive favoritism) to maximize the revenue has also attracted attention, see e.g., Fu (2006) and Li and Yu (2012). The APA with favoritism between a continuum of bidders is also studied by I. Pastine and T. Pastine (2012). For surveys on APA and APA with favoritism, see e.g., Corchón (2007), Fu and Wu (2019), and Konrad and Kovenock (2009).

On the other hand, while we are not aware of any work in the CB game literature with the generalization on the resource's asymmetric effectiveness, there are some works allowing players to have pre-allocations, although with a different taste than our perspective here. For example, Paarporn et al. (2019) focus on an incomplete information game where battlefields' values are randomly generated and players are informed asymmetrically about the battlefields; the relation between the information and the equilibrium value (and other games' parameters) is characterized in the three-battlefield constant-sum CB game. A three-stage CB game model that allows players to pre-allocate their resources is studied by Chandan et al. (2020) (and the pre-allocations are publicly known); the conditions where pre-allocating is advantageous are indicated in two-player games; these results are extended into the three-player games where two

[16]The all-pay auction has been briefly introduced at the end of Section 3.3.1.

players fight against an opponent on separate battlefields.

The generalized-rule CB game has not been studied in previous works and a game-theoretic solution of this game is important for applying it to real-world situations. Sharing the same issues with the generalized CB game, it is important to determine the set of optimal univariate distributions and an approximate equilibrium of the generalized-rule CB game—these problems remain open questions.

3.3.5 Broader Views on Blotto games: Other Extensions, Variants and Results

Finally, the literature on Blotto games and other related models is much broader than what we discussed in the previous sections. Moreover, the Nash equilibrium is not the only game-theoretic concept that is considered in studying these games. Although they are not studied in this thesis, to have a more complete overview on Blotto games, we mention several notable works on some other versions of these games as follows:

- The Colonel Blotto game with the majority rule is an important class, its main applications are situations in politics; in this game, a player wins the game (and receive a positive payoff) only if the aggregate values (or the number) of battlefields won by her exceed a given threshold (often chosen to be 50%). To distinguish with this version, in the literature, the Blotto games that we defined previously in this chapter are sometimes referred to as the plurality Blotto game. Partial results on the equilibrium of the majority rule Colonel Blotto game are provided in several works, e.g., Kvasov (2007a), Laslier (2005), Roberson and Kvasov (2012), and Weinstein (2005). The LB game with the Tullock CSF and the majority rule is also studied by Klumpp and Polborn (2006) and Snyder (1989) with results on pure equilibria in several special cases.

- A sequential version of the generalized CB game is proposed by Powell (2009) as a Stackelberg's leadership model; the existence of subgame perfect equilibria is proven in this game. Another sequential version is the Colonel Blotto Gladiator game studied by Rinott et al. (2012) where players distribute their troops into n groups and let these groups fight sequentially; the group with a larger number of troops wins the fight and moves on to fight the next group of the opponent.

- The class of General Lotto games is also studied intensively; briefly put, it is the CB game where budget constraints are relaxed to hold in expectation. In this game, players do not have the issue of constructing the mixed strategy from the optimal univariate distributions as in the CB game (simply drawing independently the allocations from these distributions is a feasible strategy in the General Lotto game) and its equilibria are completely characterized (see Myerson (1993) for the case of homogeneous battlefields, see Hart (2008) for a discrete variant and Kovenock and Roberson (2015) for a non-constant sum version). Another version is the Captain Lotto where players choose non-negative random variable values bounded from above by a cap and its sum's expectation is equal to a given budget. This game is formulated by Hart (2016) and extended by Amir (2018).

- The CB game with incomplete information has also attracted the attention of the literature. Adamo and Matros (2009) and Kovenock and Roberson (2011) study the CB game where players are equally uninformed about either the opponent budgets and/or battlefields' values; symmetric Bayes-Nash equilibria are characterized in these games. As mentioned, Paarporn et al. (2019) also concerns an incomplete information game.

To keep track of all the different variants mentioned previously, we summarize the names of the Blotto games in Table 3.2. In this table, the games whose names are written in bold letters are studied with details in this thesis and the results regarding their approximate equilibria will be presented sequentially in Chapters 4, 5 and 6. The position of Blotto games in the broader context of conflicts and their relations with other problems are presented as a diagram in Figure 3.2.

Table 3.2: Summary of variants and extensions in the class of Blotto Games

<table>
<tr><th></th><th colspan="2">Blotto functions
(β^A, β^B)</th><th>General CSFs
(ζ^A, ζ^B)</th><th>Generalized-rule</th></tr>
<tr><td>No additional constraint</td><td colspan="2">Generalized CB ($\mathcal{CB}_n$)
(Chapter 4)</td><td>Generalized LB ($\mathcal{LB}_n$)
(Chapter 6)</td><td>Generalized-rule CB (GR-CB)</td></tr>
<tr><td>$w_i^A = w_i^B, \forall i$</td><td>Constant-sum CB ($\mathcal{CB}_n^C$)</td><td rowspan="2">Constant-sum discrete CB ($\mathcal{DCB}_n^{m,p}$)
(Chapter 5)</td><td>Constant-sum LB game</td><td>Constant-sum GR-CB
(Chapter 6)</td></tr>
<tr><td>$X^A, X^B \in \mathbb{N}$ and $S^A, S^B \subseteq \mathbb{N}^n$</td><td>Discrete CB</td><td>Discrete LB</td><td>Discrete GR-CB</td></tr>
</table>

Summary: In this chapter, we introduced the class of Blotto games including the generalized CB game, its variants (the constant-sum CB game and the discrete CB game) and its extensions (the generalized LB game and the generalized-rule CB game). We presented the formulations of these games, discussed the motivation for studying them and gave a literature review on the equilibria characterization problem in each game.

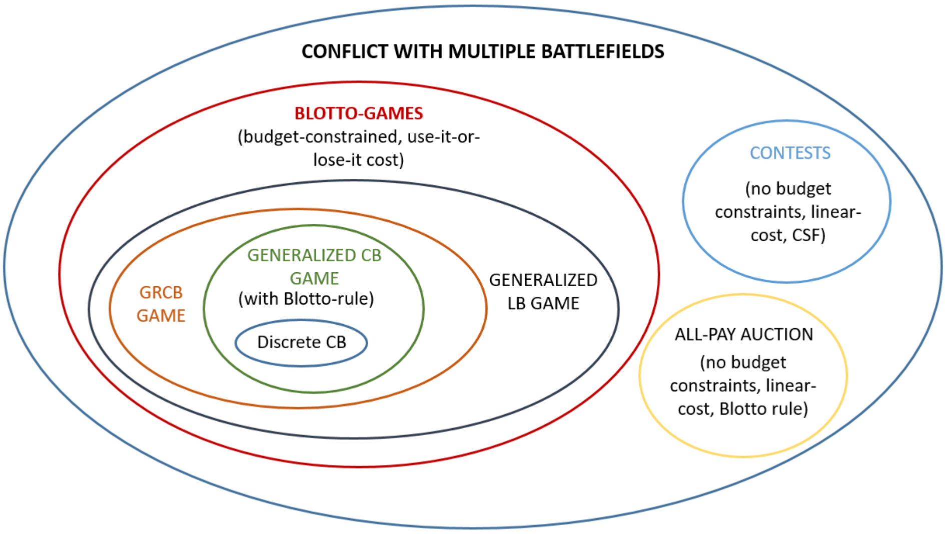


Figure 3.2: The Venn diagram showing the high-level perspective of Blotto games in the context of the contests framework.

Chapter 4

Approximate Equilibria of the Generalized Colonel Blotto Game

Some of the ideas and results presented in this chapter have previously appeared in our pre-print article Vu, Loiseau, and Silva (2019a).

The generalized Colonel Blotto game (see Definition 3.1.1), as the name suggests, is the game with the most generalization of parameters' configuration among the variants of the CB game. It still remains an open question to prove (or disprove) the existence of a Nash equilibrium of the generalized CB game and to construct an equilibrium if it exists. In Section 3.3.1, we discussed the challenges and state-of-the-art results in equilibria characterization of various restricted variants of the generalized CB game. In this chapter, we change our perspective and do not focus on this approach; instead we provide solutions for the following question: *how to construct an approximate equilibrium of the generalized CB game such that the involved approximation error is negligible (especially in large-scale applications)?*[1]

For the ease of reading, we recall several important notations in the definition of a generalized CB game $\mathcal{CB}_n$ (Definition 3.1.1) as follows: n denotes the number of battlefields; X^A, X^B denote players' budgets; w_i^A, w_i^B are the values and v_i^A, v_i^B are the normalized values that players A and B assign to battlefield $i \in [n]$; α is the tie-breaking parameter and we often use $\boldsymbol{x}^\phi$ to denote a pure strategy (i.e., an allocation) of a player ϕ. Moreover, in this chapter, we often work with an additional assumption that battlefields' values are bounded away from zero and infinity (see Assumption $(A0)$ below). This is a fairly mild assumption that is satisfied in most of (if not all) practical applications.

$$(A0) \quad \exists \underline{w}, \bar{w} > 0 : \underline{w} \leq w_i^\phi \leq \bar{w}, \forall i \in [n], \forall \phi \in \{A, B\}.$$

As a direct consequence, the normalized values satisfy

$$\frac{\underline{w}}{n\bar{w}} \leq v_i^\phi \leq \frac{\bar{w}}{n\underline{w}}, \quad \forall i \in [n], \forall \phi \in \{A, B\}. \tag{4.1}$$

[1]See Definition 2.1.6 for a definition of approximate equilibria.

The outline of this chapter is as follows: in Section 4.1, we revisit several preliminary results on optimal univariate distributions of players in the generalized CB game; then, in Section 4.2, we propose a simply-constructed class of approximate equilibria of the generalized CB game (Section 4.2.1), study the relation between the involved approximation error and the game's parameters (Section 4.2.2) and analyze the particular case of the constant-sum CB game (Section 4.2.3).

4.1 Preliminaries on Optimal Univariate Distributions

In this section, we briefly review some results from the literature that are useful for our analyses of the generalized CB game; and we show new bounds on the involved parameters, based on Assumption ($A0$), that are essential for the asymptotic analysis in the next sections. First, we introduce an important terminology that will be used regularly throughout the thesis:

Definition 4.1.1. *In the $\mathcal{CB}_n$ game, a set of univariate distributions $\left\{F_i^A, F_i^B\right\}_{i\in[n]}$ is called **optimal univariate distributions**[2] if they satisfy the following two conditions:*

- *If player ϕ draws the allocation toward battlefield i from F_i^{ϕ} for any $i \in [n]$, then her budget constraint holds in expectation; formally,*[3]

$$\sum_{i\in[n]} \left[\mathbb{E}_{x_i^{\phi}\sim F_i^{\phi}} x_i^{\phi}\right] \leq X^{\phi}. \tag{4.2}$$

- *It is optimal*[4] *for player $\phi \in \{A, B\}$ to draw her allocation toward battlefield i from F_i^{ϕ} when player $-\phi$'s allocation toward battlefield i follows $F_i^{-\phi}$; in other words, for any pure strategy $\tilde{\boldsymbol{x}}^{\phi}$ of player ϕ, the sum of expected gains from all the battlefields satisfies the following inequality:*

$$\sum_{i\in[n]} \mathbb{E}_{x\sim F_i^A, y\sim F_i^B}\left[w_i^A \beta^A(x,y)\right] \geq \sum_{i\in[n]} \mathbb{E}_{y\sim F_i^B}\left[w_i^A \beta^A(\tilde{x}_i^A, y)\right], \tag{4.3}$$

$$\sum_{i\in[n]} \mathbb{E}_{x\sim F_i^A, y\sim F_i^B}\left[w_i^B \beta^B(x,y)\right] \geq \sum_{i\in[n]} \mathbb{E}_{x\sim F_i^A}\left[w_i^B \beta^B(x, \tilde{x}_i^B)\right]. \tag{4.4}$$

Importantly, we recall that the payoff of player ϕ in the $\mathcal{CB}_n$ game is defined by $\Pi^{\phi}(\boldsymbol{x}^A, \boldsymbol{x}^B) := \sum_{i\in[n]} w_i^{\phi} \beta^{\phi}(x_i^A, x_i^B)$ when players play the pure strategies $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$. Therefore, if there exists a mixed-strategy profile (that is a pair of n-variate distributions on the players' strategy sets) that yields the marginals matching the optimal univariate distributions, then the left-hand-side of (4.3) and (4.4) becomes player ϕ's payoff when

[2]We adopt the terminology from Roberson (2006) who calls them the "optimal univariate marginal distributions" (they are only implicitly defined by Roberson (2006)) but we drop the word "marginal" since our definition does not involve a joint distribution.

[3]Here, the notation $x \sim F_i^{\phi}$ denotes that x is drawn from the random variable corresponding to F_i^{ϕ}.

[4]We only consider sets of distributions satisfying (4.2) (this is implied by (4.3) and (4.4)).

both players follow this mixed-strategy profile and the right-hand-side of (4.3) and (4.4) becomes player ϕ's payoff when she plays the pure strategy $\tilde{x}^{\phi}$ and player $-\phi$ plays by the mixed-strategy profile. Note that in Definition 4.1.1, we avoid to call these involved terms by the payoffs since the budget constraints might not hold here and drawing (independently) the allocations from F_i^{ϕ} is not necessarily a feasible mixed-strategy.[5]

From the notion of optimal univariate distributions, we have the following important proposition:

Proposition 4.1.2. *In any $\mathcal{CB}_n$ game, let $\{F_i^A, F_i^B\}_{i\in[n]}$ be optimal univariate distributions of the players; if for any $\phi \in \{A, B\}$, there exists an n-variate joint distribution of $\left\{F_i^{\phi}\right\}_{i\in[n]}$ such that any realization of this joint distribution satisfies the budget constraint of player ϕ, then the strategy profile constituted from these joint distributions is an equilibrium of the $\mathcal{CB}_n$ game.*

Proposition 4.1.2 can be trivially proved based on the definition of optimal univariate distributions. This proposition straightforwardly implies that if equilibria exist in the $\mathcal{CB}_n$ game, then we can find an equilibrium by following two steps: (i) search for a set of optimal univariate distributions of $\mathcal{CB}_n$; (ii) construct the joint distributions from these optimal univariate distributions satisfying the budget constraint. In several *restricted cases* of $\mathcal{CB}_n$, equilibrium existence is proved and an equilibrium is characterized by following the two-step scheme mentioned above (see Section 3.3.1 for an extensive review on these results). However, in the $\mathcal{CB}_n$ game with generic configurations of parameters (that is our main focus of this chapter), despite many studies, it still remains an open question to prove (or disprove) the existence of equilibria and to characterize them.

A class of optimal univariate distributions of the game $\mathcal{CB}_n$ is proposed by Kovenock and Roberson (2015) (i.e., the Step (i) is solved). To find these distributions, observe that we can break down the problem of finding the best-response of a player against a fixed strategy of her opponent into solving n all-pay auctions involving the Lagrange multipliers corresponding to the budget constraints (see e.g., Kovenock and Roberson (2015), Roberson (2006), and Schwartz et al. (2014)). Equilibria of two-player all-pay auctions are well-known (see e.g., Baye, Kovenock, and Vries (1996) and Hillman and Riley (1989) and we also rewrite these results in Proposition 6.2.3 in Chapter 6). Based on these results, Kovenock and Roberson (2015) present a class of optimal univariate distributions parameterized by positive solutions[6] of a special equation (we analyze these results in detail below). Note that since this equation can have multiple solutions, there might be more than one set of optimal univariate distributions in the $\mathcal{CB}_n$ game.

As for Step (ii) in the scheme mentioned previously, unfortunately, it still remains unknown how to construct joint distributions of the sets of optimal univariate distributions found by Kovenock and Roberson (2015) such that any realization from these joint

[5] On the other hand, in the General Lotto game—a relaxed version of the CB game where budget constraints are only required to hold in expectation—, an equilibrium is where player $\phi \in \{A, B\}$ draws independently his allocation to battlefield i from F_i^{ϕ} for any $i \in [n]$, where $\{F_i^A, F_i^B\}_{i\in[n]}$ satisfy Definition 4.1.1.

[6] Kovenock and Roberson (2015) also prove the existence of such solutions of this equation.

distributions satisfies the budget constraints. Therefore, equilibria characterization in $\mathcal{CB}_n$ game is still an open question.[7]

Although in this thesis we do not attempt to solve the open question of the equilibria characterization of the generalized Colonel Blotto game, we still use several preliminary results from this approach to construct an approximate equilibrium of the games. We present below the results on optimal univariate distributions of the game $\mathcal{CB}_n$ obtained by Kovenock and Roberson (2015) using a notation similar to that of these authors.

For each instance of the game $\mathcal{CB}_n$, for any $\gamma \in (0, \infty)$, we define

$$\Omega_A(\gamma) := \left\{ i \in [n] : v_i^A / v_i^B > \gamma \right\},$$

and consider the following equation with the variable γ (other coefficients are the parameters of $\mathcal{CB}_n$):

$$\frac{X^B \gamma}{X^A} = \frac{\gamma^2 \sum_{i \in \Omega_A(\gamma)} \frac{(v_i^B)^2}{v_i^A} + \sum_{i \notin \Omega_A(\gamma)} v_i^A}{\sum_{i \in \Omega_A(\gamma)} v_i^B + \frac{1}{\gamma^2} \sum_{i \notin \Omega_A(\gamma)} \frac{(v_i^A)^2}{v_i^B}}. \tag{4.5}$$

Let us denote by $\mathcal{S}_n^{(4.5)}$ *the set containing all positive solutions* of Equation (4.5) corresponding to the game $\mathcal{CB}_n$ (or $\mathcal{LB}_n$).[8] Based on Brouwer's fixed-point theorem, the following lemma is proved by Kovenock and Roberson (2015).

Lemma 4.1.3. *For any game $\mathcal{CB}_n$ (or $\mathcal{LB}_n$), Equation* (4.5) *has at least one positive solution; i.e., $\mathcal{S}_n^{(4.5)} \neq \emptyset$.*

Equation (4.5) may have more than one solution and it can be solved in $O(n \ln(n))$ time.[9] Now, corresponding to each positive solution $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we define two constants,[10] namely λ_A^* and λ_B^* as follows:

$$\lambda_A^* := \frac{(\gamma^*)^2}{2X^B} \sum_{i \in \Omega_A(\gamma^*)} \frac{\left(v_i^B\right)^2}{v_i^A} + \frac{1}{2X^B} \sum_{i \notin \Omega_A(\gamma^*)} v_i^A, \tag{4.6}$$

$$\lambda_B^* := \frac{1}{2X^A} \sum_{i \in \Omega_A(\gamma^*)} v_i^B + \frac{1}{2(\gamma^*)^2 X^A} \sum_{i \notin \Omega_A(\gamma^*)} \frac{\left(v_i^A\right)^2}{v_i^B}. \tag{4.7}$$

[7] We refer the interested readers to Section 3.3.1 for our discussion on results obtained by Kovenock and Roberson (2015) in the $\mathcal{CB}_n$ under several (restricted) assumptions.

[8] Note that (4.5) and $\mathcal{S}_n^{(4.5)}$ also depend on other parameters of the game $\mathcal{CB}_n$ but we use the notation with only the subscript n and omit other parameters to lighten the notation.

[9] To solve this equation algebraically, we first sort out all ratios v_i^A / v_i^B in a non-decreasing order (which can be done in $O(n \ln(n))$), then there are three possible cases: $\gamma^* < \min\{v_i^A / v_i^B, i \in [n]\}$ or $\gamma^* \geq \max\{v_i^A / v_i^B, i \in [n]\}$ or $\exists j : \gamma^* \in \left[v_j^A / v_j^B, v_{j+1}^A / v_{j+1}^B\right)$. In all of these cases, Equation (4.5) becomes a cubic equation. Finding numerically the solutions of Equation (4.5) and which one of them is positive is even more costly.

[10] These constants are the Lagrange multipliers corresponding to the budget constraints in finding players' best-response; see Kovenock and Roberson (2015) for more details.

Note importantly that we have $\gamma^* = \lambda^*_A/\lambda^*_B$ (see Lemma A.1 in Appendix A.1 for a proof). We now use these constants λ^*_A and λ^*_B to define several important distributions.

Definition 4.1.4. *Given a game $\mathcal{CB}_n$, for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and the corresponding constants λ^*_A, λ^*_B, we define the following random variables and distributions,*[11] *for each $i \in [n]$:*

(a) If $i \in \Omega_A(\gamma^*)$ (i.e., $\frac{v_i^A}{\lambda^*_A} > \frac{v_i^B}{\lambda^*_B}$), we define $A^S_{\gamma^*,i}$ and $B^W_{\gamma^*,i}$ as the random variables whose distributions are

$$F_{A^S_{\gamma^*,i}}(x) := \frac{x\lambda^*_B}{v_i^B}, \forall x \in \left[0, \frac{v_i^B}{\lambda^*_B}\right], \tag{4.8}$$

$$F_{B^W_{\gamma^*,i}}(x) := \frac{\frac{v_i^A}{\lambda^*_A} - \frac{v_i^B}{\lambda^*_B}}{\frac{v_i^A}{\lambda^*_A}} + \frac{x\lambda^*_A}{v_i^A}, \forall x \in \left[0, \frac{v_i^B}{\lambda^*_B}\right]. \tag{4.9}$$

(b) If $i \notin \Omega_A(\gamma^*)$ (i.e., $\frac{v_i^A}{\lambda^*_A} \leq \frac{v_i^B}{\lambda^*_B}$), we define $A^W_{\gamma^*,i}$ and $B^S_{\gamma^*,i}$ as the random variables whose distributions are

$$F_{A^W_{\gamma^*,i}}(x) := \frac{\frac{v_i^B}{\lambda^*_B} - \frac{v_i^A}{\lambda^*_A}}{\frac{v_i^B}{\lambda^*_B}} + \frac{x\lambda^*_B}{v_i^B}, \forall x \in \left[0, \frac{v_i^A}{\lambda^*_A}\right], \tag{4.10}$$

$$F_{B^S_{\gamma^*,i}}(x) := \frac{x\lambda^*_A}{v_i^A}, \forall x \in \left[0, \frac{v_i^A}{\lambda^*_A}\right]. \tag{4.11}$$

To lighten the notation, hereinafter, we often commonly denote these random variables as follows (the corresponding distributions are denoted by $F_{A_i^*}$ and $F_{B_i^*}$):

$$A_i^* := \begin{cases} A^S_{\gamma^*,i} & \text{if } i \in \Omega_A(\gamma^*) \\ A^W_{\gamma^*,i} & \text{if } i \notin \Omega_A(\gamma^*) \end{cases} \quad \text{and} \quad B_i^* := \begin{cases} B^S_{\gamma^*,i} & \text{if } i \notin \Omega_A(\gamma^*) \\ B^W_{\gamma^*,i} & \text{if } i \in \Omega_A(\gamma^*) \end{cases}. \tag{4.12}$$

We term these distributions the *uniform-type distributions*: $F_{A^S_{\gamma^*,i}}(x)$ is the continuous uniform distribution on $\left[0, v_i^B/\lambda^*_B\right]$ and $F_{B^W_{\gamma^*,i}}(x)$ is the distribution placing a positive mass $\left(\frac{v_i^A}{\lambda^*_A} - \frac{v_i^B}{\lambda^*_B}\right)\Big/\frac{v_i^A}{\lambda^*_A}$ at 0 and uniformly distributing the remaining mass on $\left(0, v_i^B/\lambda^*_B\right]$; similarly, $F_{B^S_{\gamma^*,i}}$ is the uniform distribution on $[0, v_i^A/\lambda^*_A]$ and $F_{B^W_{\gamma^*,i}}$ is uniform on $(0, v_i^A/\lambda^*_A]$ with a positive mass at 0. More importantly, we have the following proposition (see Lemma A.1-(iii) and Lemma A.5 in Appendix A for a proof):

[11] Here, the superscripts S and W, standing for strong and weak, are used to emphasize the intuition on players' incentive to play according to these distributions in the CB games: if $i \in \Omega_A(\gamma^*) := \left\{i : v_i^A/\lambda^*_A > v_i^B/\lambda^*_B\right\}$, player A has a "stronger" incentive to win battlefield i and player B has a "weaker" incentive; if $i \notin \Omega_A(\gamma^*)$, the roles of players are exchanged.

Proposition 4.1.5. *In the $\mathcal{CB}_n$ game, for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the distributions $\left\{F_{A_i^*}, F_{B_i^*}\right\}_{i\in[n]}$ are optimal univariate distributions of the players.*

We note also that although A_i^* and B_i^* have finite upper-bounds,[12] and that among these random variables, some may (with strictly positive probability) exceed the budgets X^A, X^B for certain parameters' configuration of the game; therefore, allocating according to $F_{A_i^*}, F_{B_i^*}$ may violate the budget constraints and it is then trivial that there exists no equilibrium yielding $F_{A_i^*}, F_{B_i^*}, \forall i \in [n]$ as marginals. On the other hand, given fixed X^A, X^B, if n is large enough, we can guarantee that A_i^*, B_i^* do not exceed the budgets for each i; however, even in this case, we still do not have guarantees on the summation of allocations sampled from all $A_i^*, B_i^*, i \in [n]$, i.e., it is still unknown if there exists an equilibrium yielding $F_{A_i^*}, F_{B_i^*}, i \in [n]$ as marginals. Note importantly that the budget-constraints violation of A_i^*, B_i^* does not affect our work and our results hold for any parameters' configuration of the games.

Finally, under Assumption ($A0$), we obtain a novel result, presented below as Proposition 4.1.6, stating that the parameters γ^*, λ_A^* and λ_B^* are all bounded. The main results of this chapter are based on asymptotic analyses in terms of the number of battlefields of the game; therefore, it is noteworthy that the bounds of these parameters do not depend on n. The proof of this proposition is given in Appendix A. From the proof of Proposition 4.1.6, we observe that as the ratios $\bar{w}/\underline{w}$ and (or) X^B/X^A increase, the ranges to which γ^* and λ_A^*, λ_B^* belong also become larger (i.e., the ratios $\bar{\gamma}/\underline{\gamma}$ and $\bar{\lambda}/\underline{\lambda}$ also increase).

Proposition 4.1.6. *Under Assumption ($A0$), for any game $\mathcal{CB}_n$, there exist positive constants $\underline{\gamma}, \bar{\gamma}, \underline{\lambda}, \bar{\lambda}$, that do not depend on n, such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and its corresponding λ_A^*, λ_B^*, we have $\underline{\gamma} \leq \gamma^* \leq \bar{\gamma}$ and $\underline{\lambda} \leq \lambda_A^*, \lambda_B^* \leq \bar{\lambda}$.*

4.2 Approximate Equilibria of the Generalized CB Game

In this section, we propose a class of strategies in the generalized Colonel Blotto game $\mathcal{CB}_n$, called the independently uniform strategies, and we show that it is an approximate Nash equilibrium (and an approximate max-min strategy in the constant-sum case)—see Section 2.1 for the definitions of approximate equilibria (and approximate max-min strategy).

4.2.1 The Independently Uniform Strategies

Given a game $\mathcal{CB}_n$, consider the corresponding Equation (4.5) and its positive-solutions set $\mathcal{S}_n^{(4.5)}$. For any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we define in Definition 4.2.1 a mixed strategy via an algorithm: Algorithm 5. We term this strategy as the *independently uniform* strategy (or IU^{γ^*} strategy), parameterized by γ^*. Intuitively, this strategy is constructed by a simple

[12]Trivially from Proposition 4.1.6, the random variables $A_i^*, B_i^*, \forall n, \forall i \in [n]$ are upper-bounded by $\bar{w}/(\underline{w} n \underline{\lambda})$. In the remainder of the chapter, we sometimes need an upper-bound of these random variables that does not depend on n: we can prove that they are bounded by $2X^B$ (see Lemma A.1 in Appendix A.1).

procedure: players start by *independently* drawing n numbers from the *uniform-type* distributions defined in Definition 4.1.4, then they re-scale these numbers to guarantee the budget constraints.

Definition 4.2.1 (The independently uniform strategy)**.** *For any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and any player $\phi \in \{A, B\}$, $\mathrm{IU}_{\phi}^{\gamma^*}$ is the **mixed** strategy of player ϕ where her allocation $\boldsymbol{x}^{\phi}$ is randomly generated from Algorithm 5.*

Algorithm 5: IU^{γ^*} strategy-generation algorithm.

Input: $n \in \mathbb{N}$, $w_i^A, w_i^B \in [\underline{w}, \bar{w}], \forall i \in [n]$, budgets X^A, X^B, $\gamma^* \in \mathcal{S}_n^{(4.5)}$

Output: $\boldsymbol{x}^A, \boldsymbol{x}^B \in \mathbb{R}^n_{\geq 0}$

1 Draw $a_i \sim F_{A_i^*}, b_i \sim F_{B_i^*}, \forall i \in [n]$ independently
2 **if** $\sum_{j\in[n]} a_j = 0$ **then**
3 $\quad x_i^A := 0, \forall i \in [n]$
4 **else**
5 $\quad x_i^A := \frac{a_i}{\sum_{j\in[n]} a_j} X^A, \forall i \in [n]$
6 **if** $\sum_{j\in[n]} b_j = 0$ **then**
7 $\quad x_i^B := 0, \forall i \in [n]$
8 **else**
9 $\quad x_i^B := \frac{b_i}{\sum_{j\in[n]} b_j} X^B, \forall i \in [n]$

Henceforth, we use the term IU^{γ^*} strategy to denote the strategy profile $(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})$. We also simply use the notation IU^{γ^*} in some places to commonly address either $\mathrm{IU}_A^{\gamma^*}$ or $\mathrm{IU}_B^{\gamma^*}$ strategy in case the name of the player need not be specified. Observe that for any player $\phi \in \{A, B\}$, any output $\boldsymbol{x}^{\phi}$ from Algorithm 5 is an n-tuple that satisfies her budget constraint. In other words, $\mathrm{IU}_{\phi}^{\gamma^*}$ is a mixed strategy that is implicitly defined by Algorithm 5 and each run of this algorithm outputs a feasible pure strategy sampled from $\mathrm{IU}_{\phi}^{\gamma^*}$. Note that the marginals of the IU^{γ^*} strategy are *not* the uniform-type distributions $F_{A_i^*}, F_{B_i^*}, i \in [n]$ defined in Section 4.1. In terms of computational complexity, Algorithm 5 terminates in $O(n)$ time. Below we discuss the specificity of the outputs of Algorithm 5 in the cases where $\sum_{j\in[n]} a_j = 0$ or $\sum_{j\in[n]} b_j = 0$.

Remark 4.2.2. *If $\sum_{j\in[n]} a_j = 0$ or $\sum_{j\in[n]} b_j = 0$, the $\mathrm{IU}_p^{\gamma^*}$ strategy allocates zero resource to all battlefields for the corresponding player (line 3 and line 7 of Algorithm 5). It may seem more natural that, if $\sum_{j\in[n]} a_j = 0$, player A allocates equally on all battlefields, i.e., set $x_i^A := X^A/n, \forall i \in [n]$ in line 3 of Algorithm 5 (and similarly for player B). In reality though, these assignments can be chosen to be any arbitrary n-tuple $\boldsymbol{x}^p$ in $\mathbb{R}^n_{\geq 0}$ as long as $\sum_{i\in[n]} x_i^p \leq X_p$ without affecting the results in our work. This comes from the fact that in most cases, the conditions in line 2 and 6 hold with probability zero. They can happen with a positive probability only when one player is the "weak player" and the other is the "strong player" on all of the battlefields (i.e., either $\Omega_A(\gamma^*) = \emptyset$ or $\Omega_A(\gamma^*) = [n]$), e.g., in a constant-sum* *game $\mathcal{CB}_n^C$. Yet, even in this case, this probability decreases exponentially as the number of battlefields increases (see (29) in A.2). The asymptotic order of the approximation error in all of our results is larger than this probability; therefore, it does not matter which assignment we choose in lines 3 and 7 of Algorithm 5. Here, we choose to assign $x_i^A = 0, \forall i$ and $x_i^B = 0, \forall i$ to ease the notation in the proofs of the results in the following sections; in particular, it avoids creating a discontinuity outside 0 in the CDF of the effective allocation in each battlefield (see also Lemma A.3 in Appendix A.2).*

4.2.2 Approximate Equilibria of the Generalized CB Game $\mathcal{CB}_n$

We now present the main result of this chapter, stating that the IU^{γ^*} strategy is an approximate equilibrium with an error that is only a negligible fraction of the maximum payoffs that the players can achieve, quickly decreasing as n increases. In the following results, note that since we focus on the setting of games with a large number of battlefields, we now focus on characterizing the approximation error according to n and treat other parameters of the $\mathcal{CB}_n$ games, including $X^A, X^B, \underline{w}, \bar{w}$ and α, as constants (but not the values $w_i^\phi, v_i^\phi, \forall i \in [n], \phi \in \{A, B\}$). We recall the notation $\tilde{O}$: it is a variant of the big-O notation that ignores the logarithmic factor.

Theorem 4.2.3.

(i) In any game $\mathcal{CB}_n$, there exists a positive number $\varepsilon = \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the following inequalities hold for any pure strategy $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ of players A and B:

$$\Pi^A(\boldsymbol{x}^A, \text{IU}_B^{\gamma^*}) \leq \Pi^A(\text{IU}_A^{\gamma^*}, \text{IU}_B^{\gamma^*}) + \varepsilon W^A, \tag{4.13}$$

$$\Pi^B(\text{IU}_A^{\gamma^*}, \boldsymbol{x}^B) \leq \Pi^B(\text{IU}_A^{\gamma^*}, \text{IU}_B^{\gamma^*}) + \varepsilon W^B. \tag{4.14}$$

(ii) There exists a constant $C^* > 0$ such that for any $\varepsilon \in (0, 1]$ and in any game $\mathcal{CB}_n$ with $n \geq C^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, (4.13) and (4.14) hold for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, any pure strategy $\boldsymbol{x}^A$, $\boldsymbol{x}^B$ of players A and B.

A proof of this theorem is presented in Appendix A.2. The two results given in Theorem 4.2.3 are two equivalent statements that can be interpreted from different perspectives. Result (i) states that given a priori a game $\mathcal{CB}_n$, there exists no unilateral deviation from the IU^{γ^*} strategy that can profit any player $\phi \in \{A, B\}$ more than a small portion of her maximum payoff W^ϕ. As a trivial corollary, the IU^{γ^*} strategy is an approximate equilibrium of the game $\mathcal{CB}_n$; this is formally stated as follows:

Corollary 4.2.4. *In any game $\mathcal{CB}_n$, there exists a positive number $\varepsilon = \tilde{O}\left(n^{-1/2}\right)$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the IU^{γ^*} strategy is an εW-equilibrium where $W := \max\{W^A, W^B\}$.*

Now, we interpret the above results in a sequence of games $\mathcal{CB}_n$ having larger and larger numbers of battlefields (i.e., n increases) to see the relation between the approximation error of the IU^{γ^*} strategies and the game's parameters in a clearer way.

The first question that might arise is: *does the approximation error always decreases as n increases?* Our answer for this question is negative. We note that as n increases, the error εW of the approximate equilibrium might not decrease to 0. This is due to the fact that although ε decreases as n increases, W might not: we recall that W^A and W^B are the total values that players A and B assess on the battlefields; therefore, as we add more battlefields, it is inevitable that W^A, W^B and $W = \max\{W^A, W^B\}$ do not decrease. This, however, does *not* reduce the contribution of our results and the applicability of the IU$^{\gamma^*}$ strategies. *In fact, we have not asked the right question*. It is not meaningful to compare the magnitude of approximation errors of IU$^{\gamma^*}$ strategies in two $\mathcal{CB}_n$ games having different sizes. Instead, we should compare the ratio between the approximation error and W—an upper-bound on players' payoffs—which is relative to the scale of the considered games. This ratio is exactly ε (we call this *the level of error*). As discussed above, as we consider the CB games with larger and larger number of battlefields, ε indeed quickly tends to 0. The bound $\tilde{O}(n^{-1/2})$ indicates the order of and how fast ε decreases as n increases.[13]

Moreover, note that this upper-bound on ε also depends on other parameters of the game $\mathcal{CB}_n$, including $X^A, X^B, \underline{w}, \bar{w}$ and α.[14] We can extract from the proof of Theorem 4.2.3 that as $\bar{w}/\underline{w}$ and/or X^B/X^A increases, ε also increases, i.e., for games with higher heterogeneity of the battlefields values and/or higher asymmetry in players' budgets, the IU$^{\gamma^*}$ strategy yields higher errors. Additionally, we note that to keep the generality, Result (i) is presented such that the approximation error ε is commonly addressed for any IU$^{\gamma^*}$ strategy with any $\gamma^* \in \mathcal{S}_n^{(4.5)}$. For each specific solution γ^* of Equation (4.5) (implying λ^*_A and λ^*_B), the corresponding IU$^{\gamma^*}$ strategy is an approximate equilibrium of $\mathcal{CB}_n$ with an approximation error that is at most (and it might be strictly smaller than) ε.

On the other hand, Result (ii)-Theorem 4.2.3 indicates the number of battlefields that a $\mathcal{CB}_n$ game needs to contain in order to guarantee a desired level of the approximation error by using the IU$^{\gamma^*}$ strategy as an approximate equilibrium. Hence, in practical situations involving large instances of the Colonel Blotto game, the IU$^{\gamma^*}$ strategy (simply and efficiently constructed by Algorithm 5) can be used as a safe replacement for any Nash equilibrium whose construction may be unknown or too complicated. Now, let us introduce an important notation:

Definition 4.2.5. *Corresponding to the players' allocations toward each battlefield $i \in [n]$, let $F_{A_i^n}$ and $F_{B_i^n}$ denote the univariate marginal distributions of the* IU$_A^{\gamma^*}$ *and* IU$_B^{\gamma^*}$ *strategies (see* (12) *and* (13) *in Appendix A.2 for a more explicit formulation of $F_{A_i^n}$ and $F_{B_i^n}$).*

Intuitively, Result (ii) can be proved by showing the two following results: *(a)* when player B's allocation to the battlefield $i \in [n]$ follows $F_{B_i^*}$, the best response

[13]Note that an alternative perspective is to consider a sequence of $\mathcal{CB}_n$ games (where n increases) whose the battlefields values are re-scaled such that $W = 1$ in all games. In this case, Corollary 4.2.4 indicates that the IU$^{\gamma^*}$ strategy is an ε-equilibrium of $\mathcal{CB}_n$ and $\varepsilon \to 0$ when $n \to \infty$.

[14]This dependency is implicitly presented in the asymptotic notation $\tilde{O}$ in Result (i) and the constant C^* in Result (ii).

of player A is to play such that her allocation to i follows the distribution $F_{A_i^*}$ (and vice versa); *(b)* as n—the number of battlefields—increases, $F_{A_i^n}$ and $F_{B_i^n}$ uniformly converge toward the distributions $F_{A_i^*}$ and $F_{B_i^*}$, i.e., the marginal distributions of the IU$^{\gamma^*}$ strategy approximate the distributions $F_{A_i^*}$ and $F_{B_i^*}$. This convergence can be proved by applying concentration inequalities on the random variables $\sum_{j \in [n]} A_j^*$ and $\sum_{j \in [n]} B_j^*$ (see Lemma A.6 in A.2); moreover, the relation between ε and n in the results of Theorem 4.2.3 depends directly on the rate of this convergence. In this work, we use the Hoeffding's inequality (see Hoeffding (1963)) that yields a better convergence rate than working with other types of concentration inequalities (e.g., Chebyshev's inequality). To complete the proof of Result (ii), we finally show that as n increases, when player $-\phi \in \{A, B\}$ plays the IU$_{-\phi}^{\gamma^*}$ strategy, the IU$_{\phi}^{\gamma^*}$'s payoff of player ϕ converges toward her best-response payoff. Note that these payoffs can be written as expectations with respect to different measures (see (14), (15) and Lemma A.4 in Appendix A.2). To prove the convergence of payoffs, we use a variant of the portmanteau theorem (see Lemma A.7 in Appendix A.2) regarding the equivalent definitions of the weak convergence of a sequence of measures. Note importantly that a direct application of the portmanteau theorem leads to a slow convergence rate (notably, (4.13) and (4.14) only hold when $n = \Omega(\varepsilon^{-4})$). This is due to the fact that the players' payoffs involve the bounded Lipschitz functions $F_{A_i^*}$ and $F_{B_i^*}$ and that these functions depend on n, particularly, their Lipschitz constants (that are either λ_A^*/v_i^A or λ_B^*/v_i^B) increase as n increases. In order to obtain the convergence rate as indicated in Theorem 4.2.3, we exploit the special relation between $F_{A_i^n}$ and $F_{A_i^*}$, and between $F_{B_i^n}$ and $F_{B_i^*}$; then we apply a telescoping-sum trick allowing us to avoid the need of using the Lipschitz properties (for more details, see the proof of Lemma A.7 in Appendix A.7).

4.2.3 Approximate Equilibria of the Constant-sum CB Game $\mathcal{CB}_n^C$

In this section, we discuss the constant-sum variant $\mathcal{CB}_n^C$ of the Colonel Blotto game, defined in Definition 3.1.2. As an instance of the generalized game $\mathcal{CB}_n$, the game $\mathcal{CB}_n^C$ satisfies all results presented in Sections 4.2.1 and 4.2.2. Additionally, we show that any IU$^{\gamma^*}$ strategy is an approximate max-min strategy of the game $\mathcal{CB}_n^C$.

In any game $\mathcal{CB}_n^C$, Equation (4.5) has a unique solution $\gamma^* = X^B/X^A \geq 1$; this γ^* uniquely induces $\lambda_A^* = 1/(2X^B)$ and $\lambda_B^* = X^A/(2{X^B}^2)$. Moreover, in $\mathcal{CB}_n^C$, we have that $v_i^A/v_i^B = 1 \leq X^B/X^A = \gamma_A^*/\gamma_B^*, \forall i \in [n]$; therefore, we have $\Omega_A(\gamma^*) = \emptyset$; intuitively, player A is the "weak player" (and B the "strong player") in *all* battlefields. Recall the notation $W := \max\{W^A, W^B\}$, in the constant-sum game $\mathcal{CB}_n^n$, we have $W = W^A = W^B$. Applying Theorem 4.2.3, we obtain the following result:

Corollary 4.2.6. *In any game $\mathcal{CB}_n^C$, there exists a positive number $\varepsilon \leq \tilde{O}(n^{-1/2})$ such that the* IU$^{\gamma^*}$ *strategy is an εW-equilibrium with $\gamma^* \in \mathcal{S}_n^{(4.5)} = \{X^B/X^A\}$.*

Note that if a Nash equilibrium exists in $\mathcal{CB}_n^C$, then the set of equilibrium univariate marginal distributions is unique (see e.g., Corollary 1 of Kovenock and Roberson (2015)) and they correspond to the distributions $F_{A_{\gamma^*,i}^W}$ and $F_{B_{\gamma^*,i}^S}$, defined in (4.10) and (4.11), where λ_A^* and λ_B^* are respectively replaced by $1/(2X^B)$ and $X^A/(2{X^B}^2)$. The marginals of the IU^{γ^*} strategy with $\gamma^* = X^B/X^A$ converge toward these unique equilibrium marginals.

Finally, we also deduce that the IU^{γ^*} strategy is an approximate max-min strategy of the game $\mathcal{CB}_n^C$; formally stated as follows:

Corollary 4.2.7. *In any game $\mathcal{CB}_n^C$, there exists a positive number $\varepsilon \leq \tilde{O}(n^{-1/2})$ such that the following inequalities hold for $\gamma^* \in \mathcal{S}_n^{(4.5)} = \left\{X^B/X^A\right\}$ and any strategy $\tilde{s}$ and $\tilde{t}$ of players A and B:*

$$\min_t \Pi^A(\tilde{s}, t) \leq \min_t \Pi^A(\mathrm{IU}_A^{\gamma^*}, t) + \varepsilon W, \tag{4.15}$$

$$\min_s \Pi^B(s, \tilde{t}) \leq \min_s \Pi^B(s, \mathrm{IU}_B^{\gamma^*}) + \varepsilon W. \tag{4.16}$$

Intuitively, if player $\phi \in \{A, B\}$ plays the $\mathrm{IU}_p^{\gamma^*}$ strategy, she guarantees a near-optimal payoff even in the worst-case scenario when her opponent $-\phi$ plays strategies that minimize ϕ's payoff (no matters how it affects $-\phi$'s payoff). The proofs of Corollary 4.2.6 and Corollary 4.2.7 can be trivially deduced by applying specifically Theorem 4.2.3 to the constant-sum Colonel Blotto games and thus are omitted in this work.

. . . .

Summary: In this chapter, we considered the generalized Colonel Blotto game. While most of (if not all) works in the literature attempt (but do not completely succeed) to construct an exact equilibrium of (variants of) this game, we took a different angle: We proposed a class of strategies called the IU^{γ^*} strategies that is simply constructed by an efficient algorithm; the IU^{γ^*} strategies guarantee the budget constraints but their marginals are not the uniform-type distributions. Yet, we proved that the IU^{γ^*} strategies are approximate equilibria of the generalized CB game with errors that are negligible relative to the magnitude of players' payoffs when the number of battlefields is large.

Chapter 5

Approximate Equilibria of the Discrete Colonel Blotto Game

Some of the ideas and results presented in this chapter have previously appeared in our publications Vu, Loiseau, and Silva (2018a,b).[a] *The numerical experiments presented in this chapter have also appeared in these publications and the corresponding codes are available at* `https://github.com/dongquan11/Approx_discrete_Blotto`.

[a] A note on the terminology: in Vu, Loiseau, and Silva (2018a,b), the constant-sum variant of the game DCB is the main focus and it is simply called the Colonel Blotto game there.

We now turn our focus to the discrete Colonel Blotto game (henceforth, DCB game) a variant of the generalized CB game that is used to model situations involving indivisible resources (see Section 3.2.1 for a formal definition of the DCB game and a discussion on its motivation). Briefly put, the DCB game has the same formulation as the generalized CB game (Definition 3.1.1) with additional constraints requiring players' budgets and allocations to be integers. In this chapter, *we focus on the constant-sum version of* DCB (i.e., the DCB game where players have the same evaluation of battlefields' values) as this version is simpler, more tractable and it has also been studied in the literature with notable results that we can use as benchmarks. The results we obtained in this constant-sum version can be extended to the non-constant-sum version of the DCB game (see Section 5.2).

Unlike the generalized CB game, the (constant-sum) DCB game is a finite game; therefore, it has at least one (mixed) equilibrium (see Proposition 2.1.4). Moreover, the equilibrium can be found numerically in general cases through linear programming. Therefore, the main focus of the literature so far is aiming to design efficient and practical algorithms that compute an equilibrium of the DCB game, and the main challenge comes from the fact that the number of pure strategies of a player in this game is exponentially large in terms of the number of battlefields and the magnitude of players' budgets (numbers of strategies in several DCB instances are presented in Table 3.1). The main idea of state-of-the-art algorithms for computing equilibrium of the (constant-sum) DCB game is to find a better representation of players' strategy sets, then convert the problem to a linear programming. As seen in our literature review in Section 3.3.2, although these algorithms provide polynomial-time/size solutions, their running time grows fast (in an order of high-degree polynomials) in terms of the number of battlefields and the budgets. They are still too impractical for real-life situations. For example, it takes over 1 day for the algorithm of Behnezhad, Dehghani, et al. (2017) to solve instances with 45 battlefields and a budget of 75 in our simulations (on a computer with an Intel core i5-7500U 2.60GHz processor and 8GB of RAM—more results are reported in Section 5.3). Note that the size of practical applications of the DCB game can be scaled up to hundreds or thousands.[1]

Our main contribution in this chapter is to propose a special approximate equilibrium of the constant-sum DCB game that can be quickly constructed. This approximate equilibrium, called the DIU strategy, is based on the idea of the class of IU^{γ^*} strategies defined in the generalized CB game (see Section 4.2.1) with necessary adjustments to guarantee that the integer constraints are satisfied. We give the formal definition of the DIU strategy in Section 5.1. In Section 5.2, we show that DIU is efficiently constructed and that the approximation error in using DIU as an approximate equilibrium is negligible under specific conditions on the number of battlefields and players' budgets. Finally, in Section 5.3, we propose an efficient algorithm, based on dynamic programming, that computes a best response of a player against a given strategy of her opponent; based on this algorithm, we conduct and analyze several numerical experiments to illustrate the trade-off between the efficiency and the accuracy of using the DIU strategy in the constant-sum DCB game.

Throughout this chapter, in order to emphasize the integer constraints in the DCB model and to set an intuition about the setting, we change some notations and terminology making them slightly differ from the description of the (non-constant-sum) DCB game in Section 3.2.1 (and the generalized CB game): resources will be referred to as *troops*,[2] the budgets will be denoted by m and p (for players A and B respectively) instead of X^A, X^B ($m, p \in \mathbb{N}\backslash\{0\}$ and $m \leq p$) and integer allocations are denoted with a hat (e.g., $\hat{x}$). To avoid the confusion, we rewrite below a definition of the constant-sum DCB game using this new notation; and we will use this definition for the remainders of this chapter.

Definition. ***The constant-sum* DCB *game***, *denoted by* $\mathcal{DCB}_n^{m,p}$, *is the game that involves two players, A and B, who simultaneously allocate their troops to n battlefields* ($n \geq 3$). *A pure strategy of player A is a vector* $\hat{\boldsymbol{x}}^{\boldsymbol{A}} \in \mathbb{N}^n$, *with integer elements* $\hat{x}_i^A \geq 0$ *representing the allocation to battlefield* $i \in [n]$ *and satisfying the constraint* $\sum_{i=1}^n \hat{x}_i^A \leq m$. *Similarly, a pure strategy of player B is a vector* $\hat{\boldsymbol{x}}^{\boldsymbol{B}} \in \mathbb{N}^n$ *such that the constraint* $\sum_{i=1}^n \hat{x}_i^B \leq p$ *holds. Each*

[1]For instance, in airport security, the number of security targets (corresponding to battlefields) can be scaled up to hundreds and security forces (corresponding to budgets) might be up to tens of thousands (e.g., in 2011, the United States Transportation Security Administration (TSA) was tasked to protect 400 airports by allocating approximately 48000 TSA officers for security screening—see Pita et al. (2011) and TSA (2011)).

[2]This convention of terminology is inspired by the military setting in the description of the CB game by Gross (1950) and Gross and Wagner (1950).

battlefield i is commonly assessed by players with a fixed value $w_i > 0$. Players' payoffs are $\Pi^A(\hat{\boldsymbol{x}}^A, \hat{\boldsymbol{x}}^B) = \sum_{i=1}^n w_i \cdot \beta^A\left(\hat{x}_i^A, \hat{x}_i^B\right)$ and $\Pi^B(\hat{\boldsymbol{x}}^A, \hat{\boldsymbol{x}}^B) = \sum_{i=1}^n w_i \cdot \beta^B\left(\hat{x}_i^A, \hat{x}_i^B\right)$ when players play the pure strategies $\hat{\boldsymbol{x}}^A, \hat{\boldsymbol{x}}^B$ (here, the Blotto-rule functions β^A, β^B are defined in (3.1)).

In this chapter, we also reuse other notation in the generalized CB game for the $\mathcal{DCB}_n^{m,p}$ game, including $W = \sum_{i=1}^n w_i$—the total value of all battlefields; $\alpha \in [0,1]$—the tie-breaking parameter (implicitly presented in the definition of β^A, β^B) and we introduce the notation $\psi := p/m$—the ratio of players' budget. Note importantly that the results in this chapter are also obtained under Assumption (A0) presented in Chapter 4, i.e., all battlefields' values belong to a bounded range: $w_i \in [\underline{w}, \bar{w}], \forall i \in [n]$, with $0 < \underline{w} \leq \bar{w}$. In the above definition, to lighten the notation, we only include the subscript n (the number of battlefields) and the superscripts m, p (players' budgets) in the notation $\mathcal{DCB}_n^{m,p}$; however, this game also depends on α and $w_i, \forall i \in [n]$ (and $\underline{w}, \bar{w}$). Finally, given a game $\mathcal{DCB}_n^{m,p}$, we say that a game $\mathcal{CB}_n$ is the *continuous CB game corresponding to $\mathcal{DCB}_n^{m,p}$* when it has the same parameters as the $\mathcal{DCB}_n^{m,p}$ game but without the integer constraints.

5.1 The DIU Strategy

In this section, we propose a mixed strategy of the $\mathcal{DCB}_n^{m,p}$ game—called the *Discrete Independently Uniform* strategy (DIU strategy)—which will be proven to be an approximate equilibrium of the game. Intuitively, under the DIU strategy, players first draw *independently* numbers from some particular *uniform*-type distributions; then they rescale these numbers to guarantee the budget constraints; finally, they use a specific rounding process to ensure the *discrete/integer* requirements.

The first (naive) idea that we can attempt to follow is to consider the continuous game corresponding to $\mathcal{DCB}_n^{m,p}$, say $\mathcal{CB}_n$, then we construct the IU$^{\gamma^*}$ strategy of $\mathcal{CB}_n$ (based on the optimal univariate distributions in $\mathcal{CB}_n$) and simply round-up the allocations drawn from this IU$^{\gamma^*}$ strategy to the closest integers (to satisfy the integer constraints in $\mathcal{DCB}_n^{m,p}$). However, this procedure has a serious flaw: due to the rounding-up step, the budget constraints might be violated and it is non-trivial how to exploit the optimality of the marginals of the IU$^{\gamma^*}$ strategy (in $\mathcal{CB}_n$) to control the approximation error of the proposed mixed strategy in $\mathcal{DCB}_n^{m,p}$. Moreover, it might happen that a non-tie situation in the continuous game can lead to a tie situation in the discrete game; these situations can make a significant difference in the payoffs. Therefore, we need to adjust the IU$^{\gamma^*}$ strategy in a more elegant manner—somehow capturing the relation between $\mathcal{DCB}_n^{m,p}$ and its corresponding continuous game, this is our objective in designing the DIU strategy proposed below.

To formalize the DIU strategy definition, in any game $\mathcal{DCB}_n^{m,p}$, for any $i \in [n]$, we introduce the uniform-type distributions:

$$F_{\tilde{A}_i^*}(x) := \left(1 - \frac{1}{\psi}\right) + \frac{x}{2\frac{w_i}{W}\psi}\frac{1}{\psi}, \forall x \in \left[0, 2\frac{w_i}{W}\psi\right], \tag{5.1}$$

$$F_{\tilde{B}_i^*}(x) := \frac{x}{2\frac{w_i}{W}\psi}, \forall x \in \left[0, 2\frac{w_i}{W}\psi\right]. \tag{5.2}$$

These distributions depend on n (the number of battlefields), ψ (the ratio of players' budgets) and w_i/W (the normalized value of battlefield i). Moreover, we observe that $F_{\tilde{B}_i^*}$ is the (continuous) uniform distribution on $\left[0, 2w_i\psi/W\right]$ and $F_{\tilde{A}_i^*}$ is the distribution where we set a probability mass $\left(1 - 1/\psi\right)$ at 0 and uniformly distribute the remaining mass on $\left(0, 2w_i\psi/W\right]$. We denote by $\tilde{A}_i^*$ and $\tilde{B}_i^*$ the random variables that correspond to $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$. Note importantly that the distributions $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ are not the optimal univariate distributions of players to battlefield i in the continuous game that corresponds to $\mathcal{DCB}_n^{m,p}$. Instead, they are, at a high-level, the optimal univariate distributions of players in a "normalized" version of it.[3] We choose to work with $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ because in the DIU strategy (defined below), there is an additional step (that has not appeared in constructing the IU$^{\gamma^*}$ strategy for $\mathcal{CB}_n$) designed specifically to guarantee the integer constraints in $\mathcal{DCB}_n^{m,p}$. This step is based on a *rounding function*, defined as $r^m : \left[0, \frac{p}{m}\right] \rightarrow \left\{0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{p}{m}\right\}$, such that $r^m(x) = \frac{\hat{x}}{m}, \forall x$, where $\hat{x} \in \mathbb{N}$ is uniquely determined and satisfies $\frac{\hat{x}}{m} - \frac{1}{2m} \leq x < \frac{\hat{x}}{m} + \frac{1}{2m}$.

Definition 5.1.1 (The DIU strategy). *In the game* $\mathcal{DCB}_n^{m,p}$, $\mathrm{DIU_A}$ *(respectively,* $\mathrm{DIU_B}$*) is the mixed strategy where player A's allocation* $\hat{\boldsymbol{x}}^{\boldsymbol{A}}$ *(respectively, player B's allocation* $\hat{\boldsymbol{x}}^{\boldsymbol{B}}$*) is randomly generated from Algorithm 6.*

Algorithm 6: DIU strategy generation algorithm.

Input: $n, m, p \in \mathbb{N}$, and $w_i \in [\underline{w}, \bar{w}], \forall i \in [n]$
Output: $\hat{\boldsymbol{x}}^{\boldsymbol{A}}, \hat{\boldsymbol{x}}^{\boldsymbol{B}} \in \mathbb{N}^n$

1 $\psi := p/m \geq 1$
2 Draw $a_i \sim F_{\tilde{A}_i^*}, \forall i \in [n]$ independently
3 **if** $\sum_{j=1}^n a_j = 0$ **then** repeat line 2
4 Draw $b_i \sim F_{\tilde{B}_i^*}, \forall i \in [n]$ independently (and independently with line 2)
5 $s_0^A = s_0^B = 0$
6 **for** $i \in [n]$ **do**
7 $\quad s_i^A = \sum_{k=1}^i \frac{a_k}{\sum_{j=1}^n a_j}$; $s_i^B = \sum_{k=1}^i \frac{b_k}{\sum_{j=1}^n b_j} \frac{p}{m}$
8 $\quad \hat{x}_i^A := m\left[r^m\left(s_i^A\right) - r^m\left(s_{i-1}^A\right)\right]$
9 $\quad \hat{x}_i^B := m\left[r^m\left(s_i^B\right) - r^m\left(s_{i-1}^B\right)\right]$

Hereinafter, we use the term DIU to refer to the strategy profile $(\mathrm{DIU}_A, \mathrm{DIU}_B)$ and also to commonly address either $\mathrm{DIU_A}$ or $\mathrm{DIU_B}$ when it is unnecessary to emphasize a particular player.

[3]In fact, $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ are the optimal univariate distribution of players' allocations toward battlefield i in the (constant-sum) $\mathcal{CB}_n$ game where player A has the budget 1, player B has the budget ψ and battlefield i's value is w_i.

Algorithm 6 guarantees that the outputs are integers and satisfy the budget constraints (with equality, i.e., without any unallocated resource). More importantly, the DIU_A (resp., (DIU_B)) strategy is only implicitly defined via Algorithm 6, that is to say it is the *joint distribution* of all allocations $\{\hat{x}_i^A\}_i$ (resp., $\{\hat{x}_i^B\}_i$). Each pure strategy output from Algorithm 6 is only one realization of the DIU strategy. In other words, Algorithm 6 computes directly a realization of the strategy, rather than computing the mixed strategy of the players (i.e., the equilibrium distribution) as in Ahmadinejad et al. (2016) and Behnezhad, Dehghani, et al. (2017). In practice, this is what a player would need to generate his allocation. Besides, it is possible to generate this distribution with arbitrary precision simply by generating many realizations independently using Algorithm 6.

Algorithm 6 is easy to implement and runs very fast in expected time $O(n)$. Note that the *loop* in lines 2-3 is not guaranteed to end in a finite time. However, the probability that the loop runs over k times is $(1-1/\psi)^{kn}$ and converges to zero exponentially fast in k and n. To guarantee that the algorithm ends in finite time, it is possible to put a stopping criterion and assign an arbitrary allocation to player A if it is reached. For instance, we can set $a_i = 0, \forall i \in [n]$ as in the IU^{γ^*} strategy in the generalized CB game $\mathcal{CB}_n$. As the condition on line 3 will happen with increasingly low probability as n grows, it can be seen from the proof of Theorem 5.2.1 that the result will still hold. On the other hand, the summation $\sum_{j=1}^n b_j$ equals 0 only happens with probability zero, therefore we do not need an additional condition to guarantee that the output from line 4 satisfying $\sum_{j=1}^n b_j > 0$.

When applying the DIU strategy, player A's allocation to battlefield $i \in [n]$ follows the (marginal) distribution $F_{A_i^D}$ while player B 's allocation follows $F_{B_i^D}$ whose corresponding random variables are defined as:

$$A_i^D = m\left[r^m\left(\sum\nolimits_{k=1}^{i}\tilde{A}_k^n\right) - r^m\left(\sum\nolimits_{k=1}^{i-1}\tilde{A}_k^n\right)\right], \tag{5.3}$$

$$B_i^D = m\left[r^m\left(\sum\nolimits_{k=1}^{i}\tilde{B}_k^n\right) - r^m\left(\sum\nolimits_{k=1}^{i-1}\tilde{B}_k^n\right)\right], \tag{5.4}$$

where for any $k \in [n]$,

$$\tilde{A}_k^n := \frac{\tilde{A}_k^*}{\sum_{j=1}^n \tilde{A}_j^*} \text{ and } \tilde{B}_k^n := \frac{\tilde{B}_k^*}{\sum_{j=1}^n \tilde{B}_j^*}\frac{p}{m}, \tag{5.5}$$

and the random variables $\tilde{A}_k^*, \tilde{B}_k^*$ have distributions (5.1)-(5.2).

We end this section by briefly describing the intuition behind $F_{A_i^D}$, $F_{B_i^D}$ and the construction of the DIU strategy in definition 5.1.1. These distributions are r^m-rounded from terms expressed by distributions $F_{\tilde{A}_i^n}$ and $F_{\tilde{B}_i^n}$, which in turn, uniformly converge towards $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ when $n \to \infty$. The key idea is that, the requirement to have discrete allocations in the discrete game $\mathcal{DCB}_n^{m,p}$ is less and less significant when the granularity of the game increases (i.e. $\frac{m}{n}, \frac{p}{n} \to \infty$), which makes $\mathcal{DCB}_n^{m,p}$ similar to its corresponding continuous game. Thus, based on the optimality of $F_{\tilde{A}_i^*}$ against $F_{\tilde{B}_i^*}$ (and vice versa) in the generalized (continuous) variant, we expect to have near-optimality in playing $\mathrm{DIU_A}$ strategy against $\mathrm{DIU_B}$ strategy (and vice versa), with any arbitrary error $\bar{\varepsilon}W > 0$, given large parameters n, m, p.

5.2 Approximate Equilibria of the Constant-sum Discrete CB Game

We present the main result of this chapter in the following theorem (its proof is given in Section 5.4):

Theorem 5.2.1.

(i) The DIU *strategy is an* $\bar{\varepsilon}W$*-equilibrium of the game* $\mathcal{DCB}_n^{m,p}$ $(m \leq p)$*, where* W *is the total value of all battlefields and* $\bar{\varepsilon} \leq \max\{\tilde{O}(n^{-1/2}), O(n/m)\}$.[4]

(ii) Fix $\psi \geq 1$ *and* $\bar{\varepsilon} > 0$*; there exists* $N^* = O\left(\bar{\varepsilon}^{-2}\ln(\bar{\varepsilon}^{-1})\right)$ *such that for* $n \geq N^*$*, there exists* $M^* = O(n/\bar{\varepsilon})$ *such that for* $m \geq M^*$ *and* $p = m\psi \in \mathbb{N}$*, for any pure strategies* $\hat{\boldsymbol{x}}^{\boldsymbol{A}}$ *and* $\hat{\boldsymbol{x}}^{\boldsymbol{B}}$ *of player A and B,*

$$\Pi^A\left(\hat{\boldsymbol{x}}^{\boldsymbol{A}}, \mathrm{DIU_B}\right) \leq \Pi^A\left(\mathrm{DIU_A}, \mathrm{DIU_B}\right) + \bar{\varepsilon}W, \tag{5.6}$$

$$\Pi^B\left(\mathrm{DIU_A}, \hat{\boldsymbol{x}}^{\boldsymbol{B}}\right) \leq \Pi^B\left(\mathrm{DIU_A}, \mathrm{DIU_B}\right) + \bar{\varepsilon}W. \tag{5.7}$$

The two statements in Theorem 5.2.1 are two equivalent results. In Result (i), the upper bound on $\bar{\varepsilon}$ is important because it allows us to evaluate the approximation error in terms of the number of battlefields and amount of troops. At a high level, it confirms the intuition that if the number of battlefields and the budgets are large enough, then the DIU strategy yields a near-optimal payoff against the opponent's DIU strategy. Moreover, this result goes much beyond merely showing this convergence and it is interesting and non-trivial in a number of ways. We notice in particular that if the ratio m/n is small, then the approximation may not be good, however large n gets. Furthermore, Result (i)-Theorem 5.2.1 (also Result (ii)) involves a double limit, with two growing parameters (n and m), and it identifies a precise *scaling regime* (i.e., ratio between the two growing parameters) under which the convergence holds. Here, it shows that the DIU strategy converges towards an equilibrium as soon as m grows at least as fast as $n^{3/2}$. This implies that, if we first make m grow to infinity, and then make n grow to infinity, the result will hold. However, the reverse is not true: if n grows first, or simply if m grows too slowly compared to n, then the DIU does not converge towards an equilibrium. Intuitively, if the number of troops is low compared to the number of battlefields, then the average number of troops per battlefield at equilibrium becomes low and the DIU strategy based on a discretization of a uniform-type distribution is no longer close to optimal.

On the other hand, Result (ii) tells us exactly how the parameters m and n should be to reach a given level of approximation when we consider the class of $\mathcal{DCB}_n^{m,p}$ games

[4] We recall again that the $\tilde{O}$ notation is a variant of the big-O notation that "ignores" logarithmic factors.

where players' budgets have a fixed ratio (i.e., a fixed asymmetry in players' strength). Note that we limited the statement of our result here to emphasize the dependence on n and m but our proof also allows extracting the dependence of $\bar{\varepsilon}$ on α, $\underline{w}, \bar{w}$ and ψ. A more precise definition of the constants given in Result (ii)-Theorem 5.2.1 is $N^* := \mathcal{O}\left(\frac{2}{\bar{\varepsilon}^2} \ln\left(\frac{4}{\bar{\varepsilon}}\right)\left[2\left(\frac{\bar{w}}{\underline{w}}\right)^2 + \psi\right]^2 \left(\frac{\underline{w}}{\bar{w}}\right)^2\right)$ and $M^* := \mathcal{O}\left(\frac{4n\bar{w}}{\bar{\varepsilon}\underline{w}\psi}\right) \max\left\{1, \frac{1}{\psi-1}\right\}$. One then observes that the convergence is slower if $\bar{w}/\underline{w}$ is larger (i.e., the battlefields heterogeneity is higher) and if ψ is larger (i.e., the players asymmetry is higher). Note that we have written the above discussion with m. The exact same holds with p instead.

Now, we remark that $\mathcal{DCB}_n^{m,p}$ is a constant-sum game. Therefore, by using inequalities (5.6) and (5.7), we can straightforwardly prove that the DIU strategy is an approximately max-min strategy of the game. This is presented as the following corollary of Theorem 5.2.1.

Corollary 5.2.2. *$\forall \psi \geq 1, \forall \bar{\varepsilon} > 0$, $\exists N^* = \mathcal{O}(\bar{\varepsilon}^{-2} \ln(\bar{\varepsilon}^{-1}))$: $\forall n \geq N^*$, $\exists M^* = \mathcal{O}(n/\bar{\varepsilon})$: $\forall m \geq M^*$, $p = m\psi \in \mathbb{N}$, in the game $\mathcal{DCB}_n^{m,p}$, for any strategy $\hat{s}$ and $\hat{t}$ of player A and B,*

$$\min_{\hat{\theta}} \Pi^A(\hat{s}, \hat{\theta}) \leq \min_{\hat{\theta}} \Pi^A(\mathrm{DIU_A}, \hat{\theta}) + \bar{\varepsilon}W, \tag{5.8}$$
$$\min_{\hat{\sigma}} \Pi^B(\hat{\sigma}, \hat{t}) \leq \min_{\hat{\sigma}} \Pi^B(\hat{\sigma}, \mathrm{DIU_B}) + \bar{\varepsilon}W. \tag{5.9}$$

This corollary ensures that the DIU strategy gives the near-optimal payoff to any player $\phi \in \{A, B\}$ even in the worst-case (when the opponent $-\phi$ plays the strategy that minimizes ϕ's payoff). This emphasizes the fact that players can "safely" use the DIU strategy in practice.

Finally, we discuss the generalizability of the obtained results to the *non-constant-sum* DCB game—i.e., the DCB game where player $\phi \in \{A, B\}$ assesses battlefield i with a value w_i^ϕ and her payoff is $\Pi^\phi(\hat{\boldsymbol{x}}^A, \hat{\boldsymbol{x}}^B) = \sum_{i=1}^n w_i^\phi \cdot \beta^\phi\left(\hat{x}_i^A, \hat{x}_i^B\right)$ (here, it allows that $w_i^A \neq w_i^B$). We can show that the approximation scheme using the DIU strategy can easily be extended to construct the approximate equilibria of the non-constant-sum discrete CB game: we only need to replace the distributions $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ in Algorithm 6 by the corresponding optimal univariate distributions of the players in the "normalized" (continuous) generalized CB game where player A has the budget of 1, player B has the budget of ψ and players evaluate the battlefields' values by $w_i^A, w_i^B, \forall i \in [n]$. The approximation error of these approximate equilibria can be bounded in an order of m, n and p that is similar to that in Theorem 5.2.1.

5.3 Numerical Evaluation

In this section, we turn to the numerical computation of quantities related to the DIU strategy, in particular to evaluate the quality of the approximation it gives depending on the game's parameters.

5.3.1 A Best-Response Algorithm in the Discrete CB Game

First, computing the value of $\bar{\varepsilon}$ (or how close a given mixed strategy of player A is to equilibrium) requires finding player B's optimal allocation given that player A's allocation to battlefield $i = 1, 2 \ldots, n$ follows a given marginal distribution $\{G_i\}_{i=1,2,\ldots,n}$. This itself is a non-trivial problem since there is in principle an exponential number of possible allocations to investigate. We propose an efficient algorithm based on dynamic programming to solve this problem (see e.g., Bertsekas (2017) for a definition of dynamic programming). This is formally presented as the following proposition.

Proposition 5.3.1. *Algorithm 7 finds a best response strategy of player B and his optimal payoff against any set of player A's marginals with complexity* $O\left(p^2 \cdot n\right)$.

Algorithm 7: Dynamic programming algorithm searching for player B's best-response (tie-breaking rule $\alpha = 0$).

Input: $n, m, p \in \mathbb{N}$, $v \in [\underline{w}, \overline{w}]^n$ and marginals $\{G_i\}_{i\in[n]}$ of player A
Output: Payoff $\Pi(p, n)$ and the best-response $\left\{\hat{x}_1^B, \cdots, \hat{x}_n^B\right\}$ against $\{G_i\}_{i\in[n]}$

1 **for** $j = 0, 1, \ldots, p$ **do**
2     $\Pi(j, 0) = 0$
3   **for** $i = 1, 2, \ldots, n$ **do**
4         $H(j, i) = w_i G_i(j)$
5         $\Pi(j, i) = \max_{k=0,\ldots,j} \left\{\Pi(k, i-1) + H\left(j-k, i\right)\right\}$
6 $j = p$
7 **for** $i = n, n-1, \ldots, 1$ **do**
8     $\hat{x}_i^B = \arg\max_{k=0,1,\ldots,j} \left\{\Pi\left(j-k, i-1\right) + H(k, i)\right\}$
9     $j = j - \hat{x}_i^B$

Note that, although our primary motivation is to compute a best-response of a player to the DIU strategy, Algorithm 7 has a broader applicability since it works for any mixed strategy of the adversary.[5] We discuss here the main intuition behind Algorithm 7 and give a descriptive proof of Proposition 5.3.1. Note firstly that the algorithm is presented here with tie-breaking parameter $\alpha = 0$ for simplicity but could straightforwardly be adapted to any tie-breaking rule. In this algorithm, $H(j, i)$ denotes the expected payoff that player B gains from battlefield i by allocating j troops to it, which is computed via the equation in line 4. More specifically, since $\alpha = 0$, by allocating j troops to battlefield i, player B wins the value w_i if j is at least equal to player A's allocation. Since G_i is the marginal distribution of player A in this battlefield, then $G_i(j)$ is exactly the probability of this event, which implies the expected gain of player B. There are $\left(p+1\right) n$ terms $H(j, i)$ to be computed yielding the complexity of $O(p \cdot n)$ to do so.

[5] Not that this algorithm is used again in our numerical experiments in Chapter 9 computing the regret in the online learning setting of the DCB game.

On the other hand, we denote $\Pi(j,i)$ the optimal payoff of player B when he is allowed to spend j troops over the set $\{1,2,\ldots,i\}$ of battlefields; thus, $\Pi(p,n)$ is exactly the best-response payoff of player B. The computation of $\Pi(j,i)$ is done by working backwards with the recursive equation given in line 5. To spend j troops over i battlefields $\{1,2,\ldots,i\}$, player B has to choose $k \in \{0,1,\ldots,j\}$ which is the number of troops he would then allocate across the first $i-1$ battlefields (whose optimal payoff is denoted $\Pi(k,i-1)$), and put the remaining $(j-k)$ troops on i^{th}–battlefield (which induces the payoff $H(j-k,i)$). He then optimizes the payoff to find $\Pi(j,i)$ by selecting the number k which maximizes the summation between the payoffs gained from these two parts. There are $O(p \cdot n)$ terms $\Pi(j,i)$ needed to be computed, each is done by comparing between at most $(p+1)$ terms; thus it yields the complexity of $O(p^2 \cdot n)$ to do so. Finally, the algorithm finds a best response strategy yielding this optimal payoff with complexity $O(p \cdot n)$ as in lines 7-9. Therefore, we conclude the proof of Proposition 5.3.1.

Reversing the roles of A and B, we can construct a similar algorithm with complexity $O\left(m^2 \cdot n\right)$ to find the best response payoff of player A against any given set of player B's marginals.

5.3.2 Numerical Experiments

In practice, we first observe that a pure strategy instructing players to allocate their resources following the DIU strategy can be generated from Algorithm 6 in time $O(n)$, which is negligible even for extremely large values of the parameters.

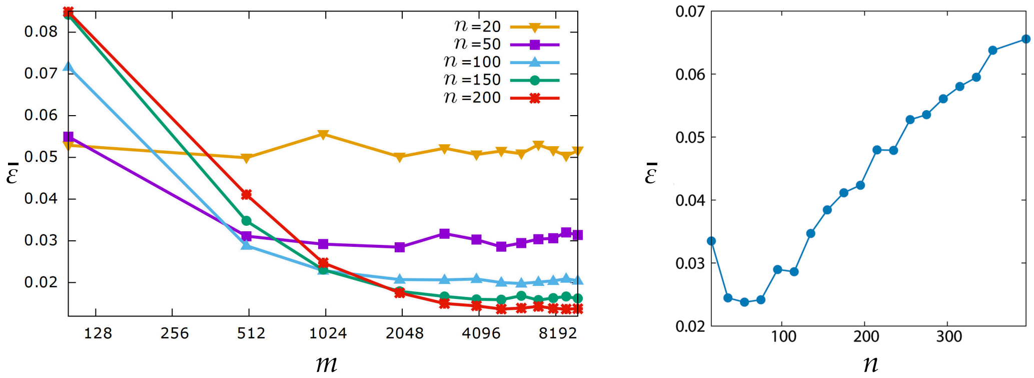
On the other hand, since the marginal allocations at battlefield i under the DIU strategy, $F_{A_i^D}$ and $F_{B_i^D}$, are not known in closed-form; we approximate them by the corresponding empirical CDFs denoted $\bar{F}_{A_i^D}$ and $\bar{F}_{B_i^D}$ computed by drawing "many" realizations of the DIU strategy from Algorithm 6. Indeed, it is known by the Glivenko-Cantelli theorem (see e.g., Vaart (1998)) that the empirical CDF converges uniformly towards the actual CDF, with a maximum difference in $O(K^{-1/2})$ where K is the number of realizations drawn. Then, to guarantee that the approximation of the DIU's CDF by its empirical CDF does not affect the computed value of $\bar{\varepsilon}$, we only need to take $K \geq O(n)$ (since $\bar{\varepsilon}$ is of the order $\tilde{O}(n^{-1/2})$ according to the previous section). Overall, generating a good approximation of the DIU's marginal distribution therefore takes time $O(n^2)$, still negligible even for large values. Finally, to compute $\bar{\varepsilon}$, for each player $\phi \in \{A,B\}$, we compare the expected payoff $\Pi^{\phi}(\mathrm{DIU_A}, \mathrm{DIU_B})$ to player ϕ's best-response payoff obtained from Algorithm 7 against the set of marginal distributions $\left\{\bar{F}_{(-\phi)_i^D}\right\}_i$ of player $-\phi$.

We construct several numerical experiments using R to illustrate the efficiency of using the DIU strategy as an approximate equilibrium of the discrete constant-sum Colonel Blotto games[6]. Our experiments run on a computer with an Intel core i5-7500U 2.60GHz processor and 8GB of RAM. In all the experiments, we keep $\alpha = 0$ and $\psi = p/m$ fixed, thus, the values of m and p always have the same growth rate (up to

[6]Our code for these experiments can be found at *https://github.com/dongquan11/Approx_discrete_Blotto*

the multiplicative constant ψ). For each instance (of $n, m, p, (w_1, w_2, \ldots, w_n)$) we run the simulations 3 times then take the average results.

Figure 5.1(a) shows the results. We first notice that when m (and p) increases, the error $\bar{\varepsilon}$ generally decreases in consistency with Theorem 5.2.1. Moreover, when m is relatively small, the error $\bar{\varepsilon}$ is higher with instances having higher number of battlefields n. This is predicted by Theorem 5.2.1, stating that when the ratios m/n and p/n are low (they decrease when n increases), the upper bound of $\bar{\varepsilon}$ is not good. For instances with higher values of m, these ratios are sufficiently large to ensure that $\bar{\varepsilon}$ decreases when either m or n (or both) increase. This interpretation is also consistent with the results shown in Figure 5.1(b). When the value of n increases, at the beginning where the ratio m/n is still sufficiently large, $\bar{\varepsilon}$ decreases. However, since we keep m (and p) fixed in this experiment, the ratio m/n gradually decreases, which makes the errors eventually get worse. Note that, for each experiment presented here, we independently generate a value for each battlefield uniformly distributed in $[\underline{w}, \bar{w}]$, with $\underline{w} = 1$ and $\bar{w} = 8$. This process explains the randomness observed in the plots.



(a) m (and p) increases, $\psi = \frac{p}{m} = 6/5$, log-scale x-axis.

(b) $m = 425, p = 510, \psi = 6/5$, n increases.

Figure 5.1: Approximation error $\bar{\varepsilon}$ of the DIU strategy stated in Theorem 5.2.1 as a function of the game parameters.

We finally compare our work with the algorithm proposed in Behnezhad, Dehghani, et al. (2017),[7] which finds an exact equilibrium of the game (we denote it by the *Algorithm EQ*). Table 5.1 shows the computation time of evaluating the error in using DIU strategy and elapsed time of Algorithm EQ for several instances. We observe that it takes remarkably less time to compute the DIU strategy payoffs and give an upper bound of the potential error by using Algorithm 7. Note that the computation times shown here include the time to compute the empirical CDF of the DIU strategy by drawing sufficiently many realizations (here we take $K = 10 \cdot n$ guaranteeing not to affect the evaluation on $\bar{\varepsilon}$) and the elapsed time of Algorithm 7. Moreover, the last column of Table 5.1 shows that the DIU strategy payoffs are very close to the exact

[7] We use the authors' implementation from *https://github.com/Soben713/ColonelBlotto*

equilibrium payoffs, even for instances with small values of the parameters m, n and p.

In conclusion, it is important to note that we do not claim that our algorithm can replace more efficiently the algorithm of Behnezhad, Dehghani, et al. (2017) (in fact, we are not computing the same thing). However, our results show that, for large values of n and p, the DIU strategy, which can be computed very efficiently, can be safely used by the players as it provides a good approximation to the equilibrium.

Table 5.1: Comparison between DIU error evaluation time and Algorithm EQ

Instances ($\psi = 6/5$)	DIU error's evaluation time			Algo. EQ elapsed time	$\frac{\lvert\text{DIU}-\text{EQ}\rvert}{W}$*
	eCDF generating	Algorithm 7	Total		
$n = 20, m = 50$	0.12s	0.36s	0.49	2540.2s	0.0066
$n = 35, m = 50$	0.34s	0.67s	1.01s	10238.7s	0.0054
$n = 50, m = 100$	0.83s	1.99s	2.83s	1.5 day	N/A
$n = 100, m = 5000$	106.46s	1396.33s	1502.79s	N/A	N/A
$n = 150, m = 8000$	380.14s	5153.11s	5533.25s	N/A	N/A
$n = 200, m = 10000$	895.36s	10991.66s	11887.02s	N/A	N/A

*The maximum difference between DIU payoffs and exact equilibrium payoffs (rescaled by the total payoff W).

5.4 Proof of Theorem 5.2.1

We end this chapter by presenting the main elements of the proof of Result(ii)-Theorem 5.2.1. We note that, although the main idea behind the DIU strategy is quite simple, the proof of this theorem is non-trivial and requires careful analysis to achieve the upper bound on $\bar{\varepsilon}$. We give here the proof of Inequality (5.6); the proof of (5.7) can be similarly done. More technical details are presented in Appendix B.

If $\bar{\varepsilon} > 1$, (5.6) and (5.7) trivially hold. In the following, we consider $\bar{\varepsilon} \leq 1$ and $\psi \geq 1$.

Proof. We start by rewriting (5.6). If player A plays a pure strategy $\hat{\boldsymbol{x}}^A$ against DIU_B, he strictly wins battlefield i with probability $F_{B_i^D}\left(\hat{x}_i^A - 1\right)$ and has a tie with probability $P(B_i^D = \hat{x}_i^A)$. Hence, according to our general tie-breaking rule,

$$\Pi^A\left(\mathbf{x}^\mathbf{A}, \text{DIU}_\text{B}\right) = \sum_{i=1}^{n} w_i F_{B_i^D}\left(\hat{x}_i^A - 1\right) + \sum_{i=1}^{n} \alpha w_i P\left(B_i^D = \hat{x}_i^A\right).$$

Similarly, player A's payoff when both players play the DIU strategy is

$$\Pi^A\left(\text{DIU}_\text{A}, \text{DIU}_\text{B}\right) = \sum_{i=1}^{n}\left[w_i \sum_{\hat{y}=0}^{m} F_{B_i^D}\left(\hat{y} - 1\right) P\left(A_i^D = \hat{y}\right)\right]$$

$$+\sum_{i=1}^{n}\left[\alpha w_i \sum_{\hat{y}=0}^{m} P\left(B_i^D=\hat{y}\right) P\left(A_i^D=\hat{y}\right)\right].$$

Therefore, to prove (5.6), it is sufficient to prove that, for all i,[8]

$$\sum_{i=1}^{n} w_i F_{B_i^D}\left(\hat{x}_i^A - 1\right) \le \sum_{i=1}^{n}\left[w_i \sum_{\hat{y}=0}^{m} F_{B_i^D}\left(\hat{y}-1\right) P\left(A_i^D=\hat{y}\right)\right] + \frac{\bar{\varepsilon}}{2} W, \tag{5.10}$$

$$P\left(B_i^D = \hat{x}_i^A\right) \le \sum_{\hat{y}=0}^{m} P\left(B_i^D=\hat{y}\right) P\left(A_i^D=\hat{y}\right) + \frac{\bar{\varepsilon}}{2\alpha}, \forall \alpha \neq 0. \tag{5.11}$$

We observe that (5.10) and (5.11) relate to the distributions $F_{A_i^D}$ and $F_{B_i^D}$, which are not expressed in closed form. However, we can approximate them with the distributions $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ defined in (5.1)–(5.2), as stated in the following lemma (its proof is given in Appendix B):

Lemma 5.4.1. *Fix $\psi \ge 1$, for any $\bar{\varepsilon}_1 \in (0,1]$, there exists $N^* := O\left(\bar{\varepsilon}_1^{-2} \ln\left(\bar{\varepsilon}_1^{-1}\right)\right)$, such that for any $n \ge N^*$, there exists $M_0 := O\left(n/\bar{\varepsilon}_1\right)$, such that for any $m \ge M_0$ and $i \in \{1,2,\dots,n\}$, we have*

$$\sup_{\hat{x}\in\mathbb{N}}\left|F_{A_i^D}(\hat{x}) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right)\right| < \bar{\varepsilon}_1 \text{ and } \sup_{\hat{x}\in\mathbb{N}}\left|F_{B_i^D}(\hat{x}) - F_{\tilde{B}_i^*}\left(\frac{\hat{x}}{m}\right)\right| < \bar{\varepsilon}_1.$$

Since A_i^D is a discrete random variables for any i, as a direct corollary of Lemma 5.4.1, for any $n \ge O\left(\bar{\varepsilon}_1^{-2} \ln\left(\bar{\varepsilon}_1^{-1}\right)\right)$ and $m \ge O\left(n/\bar{\varepsilon}_1\right)$, for $\hat{x} \in \mathbb{N}$, we have that

$$\left|\tilde{F}_{A_i^D}(\hat{x}) - \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right)\right| < \bar{\varepsilon}_1, \forall i \in \{1,2,\dots,n\}, \tag{5.12}$$

where we define

$$\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) := F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}-1}{m}\right) \text{ and } \tilde{F}_{A_i^D}(\hat{x}) := P\left(A_i^D = \hat{x}\right) = F_{A_i^D}(\hat{x}) - F_{A_i^D}(\hat{x}-1).$$

Step 1: We now prove (5.10) in 3 sub-steps.

Step 1.1: Upper bound of the left-hand-side of (5.10). Applying Lemma 5.4.1 with $\bar{\varepsilon}_1 = \bar{\varepsilon}/4$, for any $n \ge N^*$, $m \ge M_0$ and any pure strategy $\hat{\boldsymbol{x}}^A$ of player A, we have

$$\begin{aligned}
&\sum_{i=1}^{n} w_i F_{B_i^D}\left(\hat{x}_i^A - 1\right)\\
&\le \sum_{i=1}^{n} w_i \left[F_{\tilde{B}_i^*}\left(\frac{\hat{x}_i^A - 1}{m}\right) + \frac{\bar{\varepsilon}}{4}\right] \le \sum_{i=1}^{n} w_i \left(\frac{\hat{x}_i^A - 1}{m}\frac{W}{2w_i\psi} + \frac{\bar{\varepsilon}}{4}\right)\\
&= \sum_{i=1}^{n}\frac{W\hat{x}_i^A}{2\psi m} - \sum_{i=1}^{n}\frac{W}{2\psi m} + \frac{\bar{\varepsilon}}{4}W \le \frac{W}{2\psi} - \sum_{i=1}^{n}\frac{W}{2\psi m} + \frac{\bar{\varepsilon}}{4}W.
\end{aligned} \tag{5.13}$$

[8] Trivially, if $\alpha = 0$, then (5.10) implies directly (5.6); therefore, we focus on the cases where $\alpha \neq 0$.

Here, the second inequality comes from the definition of $F_{\tilde{B}_i^*}$ and the last inequality comes from the constraint $\sum_{i=1}^n \hat{x}_i^A \leq m$.

Step 1.2: Approximation of the right-hand-side of (5.10). We now need to show that the right-hand-side of (5.10) has a lower bound matching the upper bound given in (5.13). Based on Lemma 5.4.1 and Inequality (5.12), a naive approach would be to simply approximate $F_{B_i^D}(\hat{y}-1)\, P\left(A_i^D = \hat{y}\right)$ by $F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right) \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)$, for each $\hat{y} = 0, 1, \ldots, m$. However, summing all these approximation errors would lead to an error in $O(m\bar{\varepsilon})$, a large number when $m \to \infty$. Hence, we must do a finer approximation. To do so, we note that the probability of A_i^D being larger than $\left\lceil \frac{2w_i}{W} p \right\rceil$ can be bounded from above by a term independent of m. Specifically, let $\varepsilon' := \bar{\varepsilon}/8$; applying Lemma 5.4.1 with $\bar{\varepsilon}_1 = \varepsilon'/3$, for any $n \geq N^*$ and $m \geq \bar{M} := O(n/\bar{\varepsilon})$, we get that

$$P\left(A_i^D > \left\lceil 2\frac{w_i}{W}p \right\rceil\right) = 1 - F_{A_i^D}\left(\left\lceil 2\frac{w_i}{W}p \right\rceil\right) \leq 1 - F_{\tilde{A}_i^*}\left(\left\lceil 2\frac{w_i}{W}p \right\rceil \frac{1}{m}\right) + \frac{\varepsilon'}{3} = 1 - 1 + \frac{\varepsilon'}{3} = \frac{\varepsilon'}{3},$$

where the second-to-last equality comes from the fact that $\left\lceil 2\frac{w_i}{W}p \right\rceil \frac{1}{m} \geq 2\frac{w_i}{W}\psi$. Moreover,

$$\left| \sum_{\hat{y}=0}^{m} F_{B_i^D}(\hat{y}-1)\, P\left(A_i^D = \hat{y}\right) - \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} F_{B_i^D}(\hat{y}-1)\, P\left(A_i^D = \hat{y}\right) \right| \leq \frac{\varepsilon'}{3}. \tag{5.14}$$

Now, we can show that this approximate summation in (5.14) is very close to the term expressed with $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$. Indeed, we have

$$\begin{aligned} &\left| \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} F_{B_i^D}(\hat{y}-1)\, P\left(A_i^D = \hat{y}\right) - \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right) \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) \right| \\ \leq &\left| \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} F_{B_i^D}(\hat{y}-1) \left(\tilde{F}_{A_i^D}(\hat{y}) - \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right) \right| + \left| \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} \left(F_{B_i^D}(\hat{y}-1) - F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\right) \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) \right|. \end{aligned} \tag{5.15}$$

Rearranging the first term and applying corollary (5.12) with $\bar{\varepsilon}_1 := \varepsilon'/3$, there exists an $M_U := O(n/\bar{\varepsilon})$ such that for $m \geq M_U$, we have

$$\begin{aligned} &\left| \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} F_{B_i^D}(\hat{y}-1) \left(\tilde{F}_{A_i^D}(\hat{y}) - \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right) \right| \\ \leq &\left| \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} \tilde{F}_{B_i^D}(\hat{y}-1) \left(1 - F_{A_i^D}(\hat{y}) - 1 + F_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right) \right| \\ \leq &\sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} \tilde{F}_{B_i^D}(\hat{y}-1) \left| F_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) - F_{A_i^D}(\hat{y}) \right| \\ \leq &\frac{\varepsilon'}{3} \sum_{\hat{y}=0}^{\left\lceil 2\frac{w_i}{W}p \right\rceil} \tilde{F}_{B_i^D}(\hat{y}-1) \leq \frac{\varepsilon'}{3}. \end{aligned} \tag{5.16}$$

Here, the first inequality comes from the fact that[9] $F_{B_i^D}(\hat{y}-1) = \sum_{\hat{z}=0}^{\hat{y}-1} \tilde{F}_{B_i^D}(\hat{z})$; which implies

$$\begin{aligned}
&\sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} F_{B_i^D}(\hat{y}-1)\left(\tilde{F}_{A_i^D}(\hat{y}) - \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right) \\
&= \sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} \sum_{\hat{z}=0}^{\hat{y}-1} \tilde{F}_{B_i^D}(\hat{z})\left(\tilde{F}_{A_i^D}(\hat{y}) - \tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right) \\
&= \sum_{\hat{y}=1}^{\left\lceil \frac{2w_i}{W}p\right\rceil} \tilde{F}_{B_i^D}(\hat{y}-1)\left[\sum_{\hat{z}=\hat{y}}^{\left\lceil \frac{2w_i}{W}p\right\rceil} \tilde{F}_{A_i^D}(\hat{z}) - \sum_{\hat{z}=\hat{y}}^{\left\lceil \frac{2w_i}{W}p\right\rceil} \tilde{F}_{\tilde{A}_i^*}(\frac{\hat{z}}{m})\right] \\
&\leq \left|\sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} \tilde{F}_{B_i^D}(\hat{y}-1)\left(1 - F_{A_i^D}(\hat{y}) - 1 + F_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right)\right|.
\end{aligned}$$

For the second term of (5.15), we use again Lemma 5.4.1 with $\bar{\varepsilon}_1 := \varepsilon'/3$ to get $\bar{M}$ such that for $m \geq \bar{M}$,

$$\begin{aligned}
&\left|\sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil}\left(F_{B_i^D}(\hat{y}-1) - F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right| \leq \sum_{\hat{y}=1}^{\lceil 2\frac{w_i}{W}p\rceil} \frac{\varepsilon'}{3}\left|\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right| \\
&= \sum_{\hat{y}=1}^{\lceil 2\frac{w_i}{W}p\rceil} \frac{\varepsilon'}{3}\left|F_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) - F_{\tilde{A}_i^*}\left(\frac{\hat{y}-1}{m}\right)\right| \leq \sum_{\hat{y}=1}^{\lceil 2\frac{w_i}{W}p\rceil} \frac{\varepsilon'}{3m}\frac{W}{2w_i\psi^2} \\
&= \frac{\varepsilon'}{3m}\frac{W}{2w_i\psi^2}\cdot\left\lceil\frac{2w_i}{W}p\right\rceil \leq \frac{\varepsilon'}{3m}\frac{W}{2w_i\psi^2}\cdot\left(\frac{2w_i}{W}p+1\right) \\
&\leq \frac{\varepsilon'}{3\psi} + \frac{\varepsilon'(\psi-1)}{3\psi} = \frac{\varepsilon'}{3}, \forall m \geq M_V,
\end{aligned} \tag{5.17}$$

where we choose $M_V := \max\left\{\bar{M}, \frac{3n\bar{w}}{\psi 2\underline{w}(\psi-1)}\right\} = \mathcal{O}(n/\bar{\varepsilon})$.

Finally, by injecting (5.16) and (5.17) into (5.15) and combining with (5.14), we conclude that for any $n \geq N^*$ and $m \geq M_1 = \max\{M_U, M_V\} = \mathcal{O}(n/\bar{\varepsilon})$,

$$\left|\sum_{\hat{y}=0}^{m} F_{B_i^D}(\hat{y}-1)\, P\,(A_i^D=\hat{y}) - \sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right| \leq \varepsilon'. \tag{5.18}$$

Step 1.3: Lower bound of right-hand-side of (5.10). We can finally find a lower bound of the approximated sum in (5.18) (its proof is given in Appendix B):

[9] We recall the notation that $\tilde{F}_{B_i^D}(x) = F_{B_i^D}(x) - F_{B_i^D}(x-1), \forall x$

Lemma 5.4.2. *Fix $\psi \geq 1$. For any $\varepsilon' \in (0,1]$ and $n \geq N^*$, there exists an $M_2 := O(n/\varepsilon')$, such that for any $m \geq M_2$, we have*

$$\sum_{i=1}^{n}\left[w_i \sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right] \geq \frac{W}{2\psi} - \sum_{i=1}^{n}\frac{W}{2\psi m} - \varepsilon' W. \tag{5.19}$$

Conclusion of step 1. Combining (5.18) and (5.19) (with $\varepsilon' = \bar{\varepsilon}/8$), for any $n \geq N^*$ and $m \geq M_1^A := \max\{M_0, M_1, M_2\} = O\left(\frac{n}{\bar{\varepsilon}}\right)$, we have that

$$\sum_{i=1}^{n}\left[w_i \sum_{\hat{y}=0}^{m} F_{B_i^D}(\hat{y}-1)\, P\left(A_i^D = \hat{y}\right)\right] \geq \frac{W}{2\psi} - \sum\nolimits_{i=1}^{n}\frac{W}{2\psi m} - 2W\varepsilon'$$

$$\geq \sum_{i=1}^{n} w_i F_{B_i^D}\left(\hat{x}_i^A - 1\right) - \frac{\bar{\varepsilon}}{4}W - 2W\varepsilon', \qquad \text{(from (5.13))}$$

which implies exactly (5.10) (recall that $\varepsilon' = \bar{\varepsilon}/8$).

Step 2: Proof of Inequality (5.11). Choosing any $\varepsilon_2 = O(\bar{\varepsilon})$, as another corollary of Lemma 5.4.1, for any $n \geq N^* = O\left(\bar{\varepsilon}^{-2}\ln(\bar{\varepsilon}^{-1})\right)$ and $m \geq M' = O(n/\bar{\varepsilon})$), we have that $P\left(B_i^D = \hat{x}\right) = F_{B_i^D}(\hat{x}) - F_{B_i^D}(\hat{x}-1)$ is ε_2-approximated by $\tilde{F}_{\tilde{B}_i^*}\left(\frac{\hat{x}}{m}\right) := F_{\tilde{B}_i^*}\left(\frac{\hat{x}}{m}\right) - F_{\tilde{B}_i^*}\left(\frac{\hat{x}-1}{m}\right)$, $\forall \hat{x} \in \mathbb{N}, \forall i$. We now prove that (5.11) holds $\forall \alpha \neq 0$ by proving that $P(B_i^D = \hat{x}_i^A)$ gets arbitrary small when m and n increases.

- *Case 1*: if $\hat{x}_i^A \notin \left[1, \frac{2w_i}{W}p + 1\right]$, we have $F_{\tilde{B}_i^*}\left(\frac{\hat{x}_i^A}{m}\right) = F_{\tilde{B}_i^*}\left(\frac{\hat{x}_i^A - 1}{m}\right)$ (either both terms equal 0 or either they equal 1). By choosing $\varepsilon_2 = \bar{\varepsilon}/(2\alpha)$ then $P\left(B_i^D = \hat{x}_i^A\right)$ is ε_2-close to 0, which trivially leads to (5.11) when $m \geq M'$.

- *Case 2*: if $\hat{x}_i^A \in \left[1, \frac{2w_i}{W}p\right]$, we have $\tilde{F}_{\tilde{B}_i^*}\left(\frac{\hat{x}_i^A}{m}\right) = \frac{1}{m}\frac{W}{2w_i\psi}$.

On the other hand, if $\hat{x}_i^A = \left\lceil \frac{2w_i}{W}p \right\rceil$, since $\psi = \frac{p}{m}$, we also have

$$\tilde{F}_{\tilde{B}_i^*}\left(\frac{1}{m}\left\lceil\frac{2w_i}{W}p\right\rceil\right) = 1 - \frac{\left\lceil\frac{2w_i}{W}p\right\rceil - 1}{m}\frac{W}{2w_i\psi} \leq 1 - \frac{\frac{2w_i}{W}p}{m}\frac{W}{2w_i\psi} + \frac{1}{m}\frac{W}{2w_i\psi} = \frac{1}{m}\frac{W}{2w_i\psi}.$$

Therefore, by choosing $\varepsilon_2 = \bar{\varepsilon}/(4\alpha)$, for any $m \geq M_2^A := \max\left\{M', \frac{2\alpha n\bar{w}}{\bar{\varepsilon}\underline{w}\psi}\right\} = O(n/\bar{\varepsilon})$ and any x_i^A, we have

$$P\left(B_i^D = \hat{x}_i^A\right) \leq \tilde{F}_{\tilde{B}_i^*}\left(\frac{1}{m}\left\lceil\frac{2w_i}{W}p\right\rceil\right) + \frac{\bar{\varepsilon}}{4\alpha} \leq \frac{1}{m}\frac{W}{2w_i\psi} + \frac{\bar{\varepsilon}}{4\alpha} \leq \frac{\bar{\varepsilon}}{4\alpha} + \frac{\bar{\varepsilon}}{4\alpha} = \frac{\bar{\varepsilon}}{2\alpha}.$$

This directly implies (5.11).

Step 3: Conclusion of the proof. We have proved inequalities (5.10) and (5.11); thus, we conclude the proof of (5.6) by taking $M^* := \max\left\{M_1^A, M_2^A\right\} = \mathcal{O}(n/\bar{\varepsilon})$.

We can prove similarly (5.7) for player B and conclude that DIU strategy is indeed an $\bar{\varepsilon}W$-equilibrium of the game. □

. . . .

Summary: In this chapter, we studied the (constant-sum) discrete Colonel Blotto game. We proposed the DIU strategy, defined by a simple algorithm, and proved that it is an approximate equilibrium of the game. We also showed how large the number of troops and the number of battlefields of the game should be to ensure a certain level of approximation. We constructed a best-response dynamic programming algorithm and evaluated the approximation error of the DIU strategy via several numerical experiments. Our work extends the scope of applications of discrete Colonel Blotto games by trading off the accuracy with the computational efficiency, which is useful for analyzing games with large values of the parameters.

Chapter 6

Approximate Equilibria of Extensions of the Colonel Blotto Game

> *Some of the ideas and results presented in this chapter have previously appeared in our pre-print article Vu, Loiseau, and Silva (2019a). The numerical experiments presented in this chapter are conducted with our codes that are published at `https://github.com/dongquan11/GeneralizedRule_CBgame`.*

In this chapter, we consider two extensions of the Colonel Blotto game (CB game): the generalized Lottery Blotto game (LB game) and the generalized-rule Colonel Blotto game (GR-CB game). These games are used to capture situations where the CB game model is too restrictive: in the LB game, the winner in each battlefield is determined by a generic stochastic-rule (expressed as a contest success function); in the GR-CB game, players can pre-allocate resources before the start of the game and in a battlefield, one player' resources may be more effective than the other's. In Chapter 3, we presented their formal definitions, reviewed several related results and discussed motivations for studying them. To the best of our knowledge, our formulations of these games are the most general variants studied so far in the corresponding classes of games. Despite their high potential applicability, the literature still lacks detailed analyses on exact (and approximate) equilibria of the generalized LB game (with generic CSFs) and the GR-CB game. The key question still remains: how to play strategically in these games to obtain good guarantees on payoffs? To solve this question, we aim to extend the ideas of the IU strategies (Definition 4.2.1)—a class of approximate equilibria of the generalized CB game studied in Chapter 4—to these games. In each of these extensions, we encounter different sets of challenges and our contributions are to propose simply-constructed approximate equilibria with well-controlled errors of these games.

We organize this chapter into two sections, Section 6.1 and Section 6.2, respectively presenting results on our proposed approximate equilibria of the generalized Lottery Blotto game and the generalized-rule Colonel Blotto game.

6.1 Approximate Equilibria of the Generalized Lottery Blotto Game

To begin this section, we recall the basic idea that we use to build the formulation of a generalized LB game: starting from a generalized CB game with n battlefields, one replaces the Blotto functions (β^A, β^B) by a pair of (generic) contest success functions (CSFs) $\zeta = (\zeta_A, \zeta_B)$; we denote this game by $\mathcal{LB}_n(\zeta)$ (its formal definition is given in Definition 3.2.3). Recall that for a given generalized LB game $\mathcal{LB}_n(\zeta)$, we address the CB game used in formulating $\mathcal{LB}_n(\zeta)$ as its corresponding CB game (and vice versa). Importantly, in any generalized LB game, the players' strategy sets are the same as in its corresponding CB game. As a consequence, the IU^{γ^*} strategies (see Definition 4.2.1) of a generalized CB game are also feasible (mixed) strategies of the corresponding generalized LB game. In this section, we focus on the following questions: *In a generalized LB game, under which conditions do the* IU^{γ^*} *strategies yield good approximate equilibria? How to characterize the relation between the approximation errors derived from the* IU^{γ^*} *strategies and the parameters of some LB games with a particular form of CSFs?*

In Section 6.1.1, we present our solution for the first question mentioned above: we introduce novel notations to capture relations between a pair of arbitrary CSFs and the Blotto functions then show that the IU^{γ^*} strategies are approximate equilibria of the LB game with approximation errors that depend on these relations. In Section 6.1.2, we analyze the second question in the case of the LB games with the power-form CSF and the logit-form CSF that are two instances of the ratio-form CSFs—the most well-known class of CSFs used in the literature. In these instances of the LB game, under specific conditions on the number of battlefields and parameters of the CSFs, we show that the errors of the IU^{γ^*} strategies, relative to the magnitude of players' payoffs, are negligible.

For ease of reading, we recall several notations used in Chapter 4 for the generalized CB game that we adopt and reuse here for the generalized LB game. Given a game $\mathcal{LB}_n$,[1] for a player $\phi \in \{A, B\}$, X^A denote her budget, w_i^A is the value she assesses on battlefield $i \in [n]$, $W^\phi = \sum_{i\in[n]} w_i^\phi$; moreover, $W = \max\{W^A, W^B\}$ and α is the tie-breaking parameter. Corresponding to these parameters, we can rewrite Equation (4.5) and use the notation $\gamma^* \in \mathcal{S}_n^{(4.5)}$ to denote a positive solution of this equation. Moreover, we reuse the notations $F_{A_i^*}, F_{B_i^*}$ (defined formally in (4.12)) to denote the optimal univariate distributions of players in the CB game (corresponding to $\mathcal{LB}_n$) and the notations $F_{A_i^n}, F_{B_i^n}$ to denote the marginals of the strategies $\mathrm{IU}_A^{\gamma^*}$ and $\mathrm{IU}_B^{\gamma^*}$ (corresponding to the allocations toward battlefield i). Note finally that the results on the generalized LB games that we derive in this section are obtained under Assumption ($A0$) introduced in Chapter 4, that is $\exists \underline{w}, \bar{w} > 0 : \underline{w} \le w_i^\phi \le \bar{w}, \forall i \in [n], \forall \phi \in \{A, B\}$.

[1] Recall that we use the notation $\mathcal{LB}_n$ (without mentioning a specific pair of CSFs) to refer to a generalized LB game with n battlefields and a generic CSF.

6.1.1 Approximate Equilibria of the Generalized LB Game $\mathcal{LB}_n(\zeta)$ with Generic CSFs

We start by defining a parameter that expresses the dissimilarity between a given pair of CSFs $\zeta = (\zeta_A, \zeta_B)$ and the Blotto functions β_A, β_B (defined in (3.1)). Given any $\varepsilon > 0$, for any $x^* \in [0, 2X^B]$ and $y^* \in [0, 2X^B]$ (i.e., any number that can be sampled from $F_{A_i^*}, F_{B_i^*}, F_{A_i^n}$ or $F_{B_i^n}$),[2] we introduce the following sets:

$$\mathcal{X}_\zeta(y^*, \varepsilon) := \left\{x \in [0, 2X^B] : |\zeta_A(x, y^*) - \beta_A(x, y^*)| \geq \varepsilon\right\}, \tag{6.1}$$

$$\mathcal{Y}_\zeta(x^*, \varepsilon) := \left\{y \in [0, 2X^B] : |\zeta_B(x^*, y) - \beta_B(x^*, y)| \geq \varepsilon\right\}. \tag{6.2}$$

Definition 6.1.1. *For any pair of CSFs $\zeta = (\zeta_A, \zeta_B)$, $\varepsilon > 0$ and $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we define the following set*[3]

$$\Delta_{\gamma^*}(\zeta, \varepsilon) := \left\{\delta \in [0,1] : \left\{\begin{array}{l} \max\limits_{i \in [n]} \max\limits_{y^* \in [0, 2X^B]} \int_{\mathcal{X}_\zeta(y^*, \varepsilon)} \mathrm{d}F_{A_i^*}(x) \leq \delta, \\ \max\limits_{i \in [n]} \max\limits_{x^* \in [0, 2X^B]} \int_{\mathcal{Y}_\zeta(x^*, \varepsilon)} \mathrm{d}F_{B_i^*}(y) \leq \delta \end{array}\right.\right\}.$$

Intuitively, the set $\Delta_{\gamma^*}(\zeta, \varepsilon)$ contains all numbers $\delta \in [0,1]$ such that for any allocation y^* of player B toward an arbitrary battlefield i, if player A draws an allocation x from the distribution $F_{A_i^*}$, it only happens with probability at most δ that the value of the CSF ζ_A at (x, y^*) is significantly different (i.e., ε-away) from that of the Blotto function β_A; and we have a similar statement for the distribution $F_{B_i^*}$ of player B and any allocation x^* of player A. Note that the set $\Delta_{\gamma^*}(\zeta, \varepsilon)$ depends on $F_{A_i^*}$ and $F_{B_i^*}$, thus it depends on γ^*. We can trivially see that $\Delta_{\gamma^*}(\zeta, \varepsilon)$ is an interval with the form $[\delta_0, 1]$ since if $\delta_0 \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ then $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ for any $\delta \geq \delta_0$.

Based on the convergence of $F_{A_i^n}$ and $F_{B_i^n}$ toward $F_{A_i^*}$ and $F_{B_i^*}$ (see Lemma A.6 in Appendix A.2), we can prove the following lemma (a formal proof is given in Appendix C.1):

Lemma 6.1.2. *There exists $L_0 > 0$, such that for any $\varepsilon \in (0, 1]$, any $n \geq L_0 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$ and any game $\mathcal{LB}_n(\zeta)$, $\gamma^* \in \mathcal{S}_n^{(4.5)}$, $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and $i \in [n]$, we have:*

$$\max\left\{\sup_{y^* \in [0, 2X^B]} \int_{\mathcal{X}_\zeta(y^*, \varepsilon)} \mathrm{d}F_{A_i^n}(x), \sup_{x^* \in [0, 2X^B]} \int_{\mathcal{Y}_\zeta(x^*, \varepsilon)} \mathrm{d}F_{B_i^n}(y)\right\} \leq \delta + \varepsilon. \tag{6.3}$$

Intuitively, this lemma provides an upper-bound for the probability of the value of the CSFs ζ being ε-away from the Blotto functions when player A (resp. player B) plays such that her allocation to battlefields i follows $F_{A_i^n}$ (resp. $F_{B_i^n}$), i.e., when she plays the IU^{γ^*} strategy.

Based on the definition of $\Delta_{\gamma^*}(\zeta, \varepsilon)$ and Lemma 6.1.2, we can now show the following result on the IU^{γ^*} strategy in the generalized Lottery Blotto games.

[2] Recall that for any n and $i \in [n]$, the random variables A_i^*, B_i^* are upper-bounded by $2X^B$ (see Lemma A.1 in Appendix A) and by definition, the variables A_i^n, B_i^n are trivially upper-bounded by X^A, X^B (and thus by $2X^B$).

[3] Note that $F_{A_i^*}, F_{B_i^*}$ are continuous, bounded functions on $[0, 2X^B]$; therefore, they attain a maximum on this interval.

Theorem 6.1.3. ***(Approximate equilibria of the generalized Lottery Blotto game).***

(i) In any game $\mathcal{LB}_n(\zeta)$, there exists a positive number $\varepsilon \le \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, the following inequalities hold for any pure strategy $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ of players A and B:[4]

$$\Pi_\zeta^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*}) \le \Pi_\zeta^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (8\delta + 13\varepsilon)\, W^A, \tag{6.4}$$

$$\Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \boldsymbol{x}^B) \le \Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (8\delta + 13\varepsilon)\, W^B. \tag{6.5}$$

(ii) There exists $L^* > 0$, such that for any $\varepsilon \in (0,1]$ and in any game $\mathcal{LB}_n(\zeta)$ where $n \ge L^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, (6.4) and (6.5) hold for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and any pure strategy $\boldsymbol{x}^A, \boldsymbol{x}^B$ of players A and B.

The proof of this theorem is given in Appendix C.1. The main idea to prove these results is that we can approximate the players' payoffs in the game $\mathcal{LB}_n(\zeta)$ when they play the IU^{γ^*} strategies by that in the corresponding game $\mathcal{CB}_n$ (the difference between these payoffs is controlled by the parameter $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$); and then use the results from Chapter 4 for the game $\mathcal{CB}_n$ (involving the error ε) to prove (6.4) and (6.5). The coefficients (8 and 13) in front of the parameters δ and ε come from the application of several triangle inequalities to connect these approximate results. Note that if the CSFs ζ_A and ζ_B are Lipschitz continuous on $[0, 2X^B] \times [0, 2X^B]$, we can avoid the need to approximate several terms involved in the analysis of using the IU^{γ^*} strategy in the game $\mathcal{LB}_n(\zeta)$ via the corresponding terms in the game $\mathcal{CB}_n$; thus, we can improve the results in Theorem 6.1.3 to obtain an approximation error of $2\delta + 5\varepsilon$ instead of $8\delta + 13\varepsilon$ (see Remark C.3 in Appendix C.1 for more details). Here, to keep the generality, we do not include the continuity assumption of the CSFs in Theorem 6.1.3 (recall that our definition of a CSF allows for discontinuity).

Intuitively, Result (i) of Theorem 6.1.3 determines the order of the approximattion error while using IU^{γ^*} in any given game $\mathcal{LB}_n(\zeta)$. Straightforwardly, we can deduce that the IU^{γ^*} strategy is an approximate equilibrium of the game $\mathcal{LB}_n(\zeta)$, formally stated as follows:

Corollary 6.1.4. *In any game $\mathcal{LB}_n(\zeta)$, there exists a positive number $\varepsilon \le \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, the IU^{γ^*} strategy is an $(8\delta + 13\varepsilon)\, W$-equilibrium where $W := \max\{W^A, W^B\}$.*

We observe that the error bound in Theorem 6.1.3 (and in Corollary 6.1.4) is valid for any δ of the set $\Delta_{\gamma^*}(\zeta, \varepsilon)$. Naturally, it is the tightest for $\delta_0 = \min\{\delta : \delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)\}$; but this quantity is not always easy to compute; for instance, in the Lottery Blotto games with the power and logit form CSFs presented in Section 6.1.2. Still, we can obtain a good error's upper-bound for the IU^{γ^*} strategy in these games by showing that there exists an element of the corresponding set $\Delta_{\gamma^*}(\zeta, \varepsilon)$ that is negligibly small, given appropriate parameter's configurations of the game; these are all given in Section 6.1.2.

[4] Recall that Π_ζ^A and Π_ζ^B denote the payoffs functions of players A and B in the game $\mathcal{LB}_n(\zeta)$.

Note that, on the other hand, the generalized CB game $\mathcal{CB}_n$ can be considered as an instance of the game $\mathcal{LB}_n(\zeta)$ where the CSFs are $\zeta_A = \beta_A$ and $\zeta_B = \beta_B$; therefore, it also satisfies Theorem 6.1.3. In $\mathcal{CB}_n$, we trivially have $\mathcal{X}_\zeta(y^*, \varepsilon) = \mathcal{Y}_\zeta(x^*, \varepsilon) = \emptyset$ for any x^*, y^*; thus $\Delta_{\gamma^*}(\zeta, \varepsilon) = [0, 1]$ for any $\varepsilon > 0$ and $\min\{\delta : \delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)\} = 0$.[5] This is consistent with results obtained in Theorem 4.2.3 in Chapter 4.

In Theorem 6.1.3, Result (ii) is an equivalent statement of Result (i). It indicates the number of battlefields needed to guarantee a certain level of approximation error when using the IU^{γ^*} strategy in the game $\mathcal{LB}_n(\zeta)$. For instance, to obtain an approximate equilibrium of the game $\mathcal{LB}_n(\zeta)$ where the level of error is less than a certain number $\bar{\varepsilon}$, one needs $\varepsilon \leq \bar{\varepsilon}$ such that we can find a $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ satisfying $8\delta + 13\varepsilon \leq \bar{\varepsilon}$; from these parameters, by Result (ii), one can deduce the sufficient number of battlefields needed for an $\mathcal{LB}_n$ game to yield that desired level of error.

Finally, in the constant-sum variant of the Lottery Blotto game (i.e., when $w_i^A = w_i^B$, $\forall i \in [n]$), denoted by $\mathcal{LB}_n^c(\zeta)$, we can easily deduce from Theorem 6.1.3 that the IU^{γ^*} strategy is also an approximate max-min strategy:

Corollary 6.1.5. *In any game $\mathcal{LB}_n^c(\zeta)$, $\mathcal{S}_n^{(4.5)}$ is a singleton (more specifically, $\mathcal{S}_n^{(4.5)} = \{X^B/X^A\}$), and there exists $\varepsilon \leq \tilde{O}(n^{-1/2})$ such that for any $\gamma^* = X^B/X^A$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, the following inequalities hold for any strategy $\tilde{s}$ and $\tilde{t}$ of players A and B:*[6]

$$\min_t \Pi_\zeta^A(\tilde{s}, t) \leq \min_t \Pi_\zeta^A(\mathrm{IU}_A^{\gamma^*}, t) + (8\delta + 13\varepsilon)W,$$
$$\min_s \Pi_\zeta^B(s, \tilde{t}) \leq \min_s \Pi_\zeta^B(s, \mathrm{IU}_B^{\gamma^*}) + (8\delta + 13\varepsilon)W.$$

6.1.2 Approximate Equilibria of the Ratio-form LB Game

Besides the Lottery Blotto game with generic CSFs, we additionally consider LB games corresponding to the CSFs that belong to a special class called the *ratio-form* CSFs. These are the CSFs that are studied the most profoundly in the literature. We will use the games with these ratio-form CSFs to illustrate the results obtained in Section 6.1.1 for the generalized LB game.

Definition 6.1.6. *CSFs $\zeta_A, \zeta_B : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$ are called **ratio-form CSFs** if they have the form:*

$$\zeta_A(x, y) = \frac{\eta(x)}{\eta(x) + \psi(y)} \quad \text{and} \quad \zeta_B(x, y) = \frac{\psi(y)}{\eta(x) + \psi(y)},$$

where $\eta, \psi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ are non-negative functions such that ζ_A and ζ_B satisfy Definition 3.2.2.

Two classical ratio-form CSFs in the literature (see e.g., Corchón and Dahm (2010) and Hillman and Riley (1989)) are the power form where $\eta(z) = \psi(z) = z^R, \forall z \geq 0$ and the logit form where $\eta(z) = \psi(z) = e^{Rz}, \forall z \geq 0$, where $R > 0$ is a parameter chosen a priori. These functions yield the sharing 50-50 tie-breaking rule, i.e., if $x = y$, then

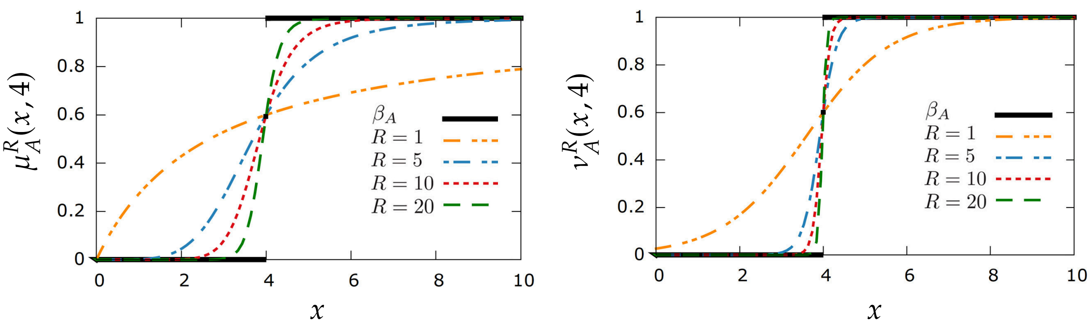
[5]Note also that for the case of the game $\mathcal{CB}_n$, the left-hand side in (6.3) equals zero for any n and $i \in [n]$.

[6]Recall that in the constant-sum variant, $W := \max\{W^A, W^B\} = W^A = W^B$.

$\zeta_A(x,y) = \zeta_B(x,y) = 1/2$. We define in Table 6.1 the generalized versions of these ratio-form CSFs using the parameter $\alpha \in (0,1)$ that leads to the general tie-breaking rule as in the Colonel Blotto game $\mathcal{CB}_n$.[7] Henceforth, we use the terms power and logit form to indicate the CSFs μ^R and ν^R with this generalization. It is trivial to verify that both pairs (μ^R_A, μ^R_B) and (ν^R_A, ν^R_B) satisfy the Conditions $(C1)$ and $(C2)$. An important remark is that both the power and logit form CSFs converge pointwise toward the Blotto functions β_A, β_B as R tends to infinity (we will revisit this remark with more details). This convergence can be observed in Figure 6.1 that illustrates several instances of the ratio-form CSFs in comparison with the Blotto functions. Note also that the LB games with the CSF μ^R has been introduced by Shubik and Weber (1981); however, no equilibrium (or approximate equilibrium) result has been given for this game. On the other hand, in our knowledge, the LB game $\mathcal{LB}_n(\mu^R)$ has not been studied before.

Table 6.1: Power and logit form CSFs with generalized tie-breaking rule ($\alpha \in (0,1)$).

Name	Notation	If $x^2+y^2>0$	If $x=y=0$
Power form	$\mu^R := (\mu^R_A, \mu^R_B)$	$\mu^R_A(x,y) = \frac{\alpha x^R}{\alpha x^R + (1-\alpha) y^R}$; $\mu^R_B(x,y) = \frac{(1-\alpha) y^R}{\alpha x^R + (1-\alpha) y^R}$	$\mu^R_A(x,y) = \alpha$ $\mu^R_B(x,y) = 1-\alpha$
Logit form	$\nu^R := (\nu^R_A, \nu^R_B)$	$\nu^R_A(x,y) = \frac{\alpha e^{xR}}{\alpha e^{xR} + (1-\alpha) e^{yR}}$ $\nu^R_B(x,y) = \frac{(1-\alpha) e^{yR}}{\alpha e^{xR} + (1-\alpha) e^{yR}}$	$\nu^R_A(x,y) = \alpha$ $\nu^R_B(x,y) = 1-\alpha$



(a) Power CSF, $X^A = 10, \alpha = 0.6$. (b) Logit CSF, $X^A = 10, \alpha = 0.6$.

Figure 6.1: Examples of power-form and logit-form CSFs in comparison with the Blotto functions.

We now consider LB games $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$; henceforth, we call them the ratio-form LB games. Note that for the CSFs μ^R and ν^R, we do not consider the degen-

[7] When $\alpha = 1/2$, the CSFs μ^R and ν^R match the classical power form and logit form CSFs. Note that we exclude the cases where $\alpha = 0$ or $\alpha = 1$ since these are the trivial cases: in the corresponding Lottery Blotto game, a player, say $p \in \{A, B\}$, always has the payoff W^ϕ while player $-p$'s payoff is always zero regardless how they allocate their resources.

erate cases where $\alpha = 0$ or $\alpha = 1$ in which trivial equilibria exist. The games $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$ are instances of the game $\mathcal{LB}_n(\zeta)$ studied in Section 6.1.1; therefore, by Theorem 6.1.3 (and Corollary 6.1.4), the IU$^{\gamma^*}$ strategy is also an approximate equilibrium of them. In this section, we focus on characterizing the approximation error of the IU$^{\gamma^*}$ strategy in these games according to n (the number of battlefields) and R (the corresponding parameter of the CSFs). We will show that this error quickly tends to zero as n and R increase under appropriate conditions. To do this, we first notice that although it is non-trivial to analyze the closed form of the sets $\Delta_{\gamma^*}(\mu^R, \varepsilon)$ and $\Delta_{\gamma^*}(\nu^R, \varepsilon)$ and find their minimum, we can find small elements of theses sets.

Lemma 6.1.7. *Fix $n \geq 2$, $R > 0$ and $\alpha \in (0, 1)$. Setting α as the tie-breaking rule for the games mentioned below, for any $\varepsilon < \min\{\alpha, 1-\alpha\}$, we have:*[8]

(i) *in any game $\mathcal{LB}_n(\mu^R)$, there exists $\delta_\mu = \min\{1, \mathcal{O}\left(n(\varepsilon^{\frac{-1}{R}} - 1)\right)\}$ such that $\delta_\mu \in \Delta_{\gamma^*}(\mu^R, \varepsilon)$ for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$;*

(ii) *in any game $\mathcal{LB}_n(\nu^R)$, there exists $\delta_\nu = \min\{1, \mathcal{O}\left(nR^{-1}\ln(\varepsilon^{-1})\right)\}$ such that $\delta_\nu \in \Delta_{\gamma^*}(\nu^R, \varepsilon)$ for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$.*

The proof of Lemma 6.1.7 is given in Appendix C.1. Note that for the sake of generality, the parameters δ_μ and δ_ν are indicated in this lemma in such a way that they do not depend on γ^*, but for each $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we can find smaller elements of the corresponding sets $\Delta_{\gamma^*}(\mu^R, \varepsilon)$ and $\Delta_{\gamma^*}(\nu^R, \varepsilon)$. More importantly, for a fixed n, the numbers δ_μ and δ_ν decrease as R increases; but δ_μ and δ_ν increase as ε decreases. While the lemma is valid for any parameter values, since 1 is a trivial element of $\Delta_{\gamma^*}(\mu^R, \varepsilon)$ and $\Delta_{\gamma^*}(\nu^R, \varepsilon)$, it is useful only if $\delta_\mu, \delta_\nu < 1$; this is guaranteed whenever $R \geq \mathcal{O}\left(n\ln(\varepsilon^{-1})\right)$. Note finally that the condition $\varepsilon < \min\{\alpha, 1-\alpha\}$ in the statement of Lemma 6.1.7 does not limit its use since our goal is to obtain asymptotic results on the IU$^{\gamma^*}$ strategy when ε tends to 0. Moreover, in the games $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$ where α is either very close to 0 or 1, one player has a very high advantage and always obtains large gains from all battlefields (where her allocation is strictly positive) while her opponent gains very little regardless of her allocations; therefore, there exist (many) trivial approximate equilibria with small errors.

Combining the results of Corollary 6.1.4 and Lemma 6.1.7, we can deduce directly that in any game $\mathcal{LB}_n(\mu^R)$ (resp. $\mathcal{LB}_n(\nu^R)$), there exists $\varepsilon \leq \tilde{\mathcal{O}}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the IU$^{\gamma^*}$ strategy is an $(8\varepsilon + 13\delta_\mu)W$-equilibrium (resp. $(8\varepsilon + 13\delta_\nu)W$-equilibrium). Next, we look for the asymptotic relation between these error terms and the parameters n, R. First, as n increases, the error level ε decreases; on the other hand, from Lemma 6.1.7, the number δ_μ (and δ_ν) decreases if R increases with a faster rate than $\tilde{\mathcal{O}}(n)$. However, there is a trade-off between ε and δ_μ (or δ_ν): as ε decreases, δ_μ (and δ_ν) increases and vice versa. To handle this trade-off between δ_μ and ε (resp. δ_ν and ε), we can first find a condition on n that generates a small error ε, and then find a condition on R (with respect to n) such that the error δ_μ (resp. δ_ν) is of the same

[8] The asymptotic notations are taken w.r.t. when $\varepsilon \to 0$.

order as ε. Formally, we state the result that the IU^{γ^*} strategy yields an approximate equilibrium of the games $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$ with any arbitrary small error in the next theorem.

Theorem 6.1.8. ***(Approximate equilibria of the ratio-form Lottery Blotto games)*** *For any $\bar{\varepsilon} > 0$ and $\alpha \in (0,1)$ such that $\bar{\varepsilon} < \min\{\alpha, 1-\alpha\}$, there exists $\tilde{L} > 0$ such that for any $n \geq \tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon}, 1/e\}}\right)$, $R \geq \mathcal{O}\left(\frac{n}{\bar{\varepsilon}} \ln\left(\frac{1}{\bar{\varepsilon}}\right)\right)$ and $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the IU^{γ^*} strategy is an $\bar{\varepsilon}W$-equilibrium of any game $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$ having α as the tie-breaking-rule parameter.*

The proof of this theorem is based on Theorem 6.1.3 and Lemma 6.1.7 (see Appendix C.1 for more details). Theorem 6.1.8 involves a double limit in R and n. Intuitively, if n and R increase but R increases with a slower rate, then ε decreases but the corresponding δ_μ and δ_ν do not decrease; thus, the total error is not guaranteed to decrease.

6.2 Approximate Equilibria of the Generalized-Rule Colonel Blotto Game

In this section, we turn our focus to the generalized-rule Colonel Blotto game (or GR-CB game in short). A definition of the $\mathcal{GR}\!-\!\mathcal{CB}_n$ game (i.e., a GR-CB game with n battlefields) was presented in Definition 3.2.4. We emphasize again that there is currently no work in the literature considering a similar model to ours (capturing situations with both pre-allocations of resources and asymmetric effectiveness). By Definition 3.2.4, the game $\mathcal{GR}\!-\!\mathcal{CB}_n$ is a non-constant-sum game. As in the case of the generalized CB games, one can obtain the constant-sum variant of the GR-CB game by simply adding the constraint requiring that players have the same evaluation on each battlefield's value. In this section, *we limit ourselves to study only the constant-sum variant* of the GR-CB game since it is simpler and more tractable. We will discuss the generalization of the obtained results into the non-constant-sum variant of the GR-CB game in the end of this section. To avoid confusion, we redefine here the constant-sum GR-CB game:

Definition 6.2.1. *In the* ***constant-sum generalized-rule Colonel Blotto game*** *with n battlefields, denoted by $\mathcal{GR}\!-\!\mathcal{CB}_n^C$, players A and B simultaneously allocate their resource (with budgets X^A and X^B). Each battlefield $i \in [n]$ has a value $w_i > 0$ commonly assessed by the players; moreover, it is embedded with two extra parameters $p_i \in \mathbb{R}$ and $q_i > 0$. Players' payoffs when they play the pure strategies $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ are $\Pi_{GR}^A(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum_{i=1}^n w_i \cdot \beta^A\left(x_i^A, q_i x_i^B - p_i\right)$ and $\Pi_{GR}^B(\boldsymbol{x}^A, \boldsymbol{x}^B) = \sum_{i=1}^n w_i \cdot \beta^B\left(x_i^A, q_i x_i^B - p_i\right)$; where β^A, β^B are the Blotto functions defined in* (3.1).

Within this section, when there is no ambiguity, we drop the term "constant-sum" in the name and address the $\mathcal{GR}\!-\!\mathcal{CB}_n^C$ game simply as the generalized-rule CB game. We recall the intuition that the parameter p_i indicates the difference between players' pre-allocations and q_i indicates the asymmetry in resources' effectiveness at battlefield i

(see Section 3.2.3 for more elaborated details). We reuse the notations $W := \sum_{i \in [n]} w_i$ and $\alpha \in [0,1]$ to denote the total of the battlefields' values and the tie-breaking parameter (implicitly defined in the functions β^A, β^B) of the $\mathcal{GR}-\mathcal{CB}_n^C$ game. Moreover, as in the previous chapters, some of the results in this section are obtained under Assumption (A0), that is for any $i \in [n]$, there exist $\underline{w}, \bar{w}$ such that $0 < \underline{w} \leq w_i \leq \bar{w}$.

We address two key questions in this section: In a $\mathcal{GR}-\mathcal{CB}_n^C$ game, can we find a set of optimal univariate distributions of the players? Can we use the ideas of the IU$^{\gamma^*}$-strategies (in the $\mathcal{CB}_n$ game) to obtain approximate equilibria of the game $\mathcal{GR}-\mathcal{CB}_n^C$ yielding a similar control on the approximation errors? To answer the first question, we study another model, called the all-pay auction with favoritism, and completely characterize the exact equilibria of this game (in Section 6.2.1). Based on these results, we construct an approximate equilibrium of the $\mathcal{GR}-\mathcal{CB}_n^C$ game in Section 6.2.4—this is our solution for the second question posed above.

6.2.1 The All-pay Auction with Favoritism

As discussed in Chapter 3, the all-pay auction (henceforth, APA) is a well-known problem that relates closely to the class of Blotto games. In APA, two players secretly decide their bids to compete for a common item; the higher bidder wins the item and gains its value; then, both players pay their bids. In the literature, it is common to use APA as a tool to find optimal univariate distributions[9] of the CB games (see e.g., Roberson (2006) and Schwartz et al. (2014)). In this section, we study a special extension of the APA model, called the all-pay auction with favoritism (or F-APA for short) where the rule determining the winner among the bidders is shifted by an affine transformation. We use the equilibria of this F-APA game to attempt to construct optimal univariate distributions of the players in the game $\mathcal{GR}-\mathcal{CB}_n^C$.[10]

A description of the F-APA game is as follows: It is an auction for a common item by two players, A and B. The item is evaluated by each player with a value (denoted u^A and u^B for players A and B ($u^A, u^B > 0$));[11] moreover, it is embedded with two additional parameters: $p \in \mathbb{R}$ and $q > 0$. Knowing these parameters, the players simultaneously submit their *bids*, denoted by x^A and x^B, where $x^A, x^B \geq 0$ (unlike in Blotto games, players can bid as large as they want in F-APA). The *winner-determination rule* of the game is defined as follows: if $x^A > qx^B - p$, player A wins the item and gains the corresponding value u^A; reversely, if $x^A < qx^B - p$, player B wins the item and gains the value u^B; in case of a tie, i.e., $x^A = qx^B - p$, each player earns a portion of the corresponding values: player A gains αu^A and player B gains $(1-\alpha)u^B$ (here, $\alpha \in [0,1]$ is a given parameter). Finally, *both players* pay the bids that they submitted (regardless of the winner). Formally, we have the following definition:

[9]See Definition 4.1.1 for a definition of the optimal univariate distributions in $\mathcal{CB}_n$.

[10]A definition of optimal univariate distributions in the $\mathcal{GR}-\mathcal{CB}_n^C$ game can be easily extended from the definition of optimal univariate distributions of the $\mathcal{CB}_n$ game (Definition 4.1.1) by replacing the terms in (4.3) and (4.4) by the corresponding payoffs of $\mathcal{GR}-\mathcal{CB}_n^C$.

[11]The case where either $u^A = 0$ or $u^B = 0$ is trivial (there exist trivial pure equilibria) and thus, is omitted in this work.

Definition 6.2.2. *The all-pay auction with favoritism (henceforth, the F-APA game) is the game with the above description and the players' payoffs when player A bids x^A and player B bids x^B are:*

$$\Pi^A\left(x^A, x^B\right) = u^A \cdot \beta^A\left(x^A, qx^B - p\right) - x^A, \tag{6.6}$$
$$\Pi^B\left(x^A, x^B\right) = u^B \cdot \beta^B\left(x^A, qx^B - p\right) - x^B. \tag{6.7}$$

Here, the functions β^A, β^B are defined in (3.1).

The main novelty setting the F-APA game apart from the classical (two-player first-price) APA is the parameters p and q. From the above definition, we remark that if $p > 0$, player A has an advantage (we call this the *additive favoritism*) in winning the item and reversely if $p < 0$, player B has this favoritism; likewise, when $0 < q < 1$, player A has a *multiplicative favoritism* to win the item and when $q > 1$, this type of favoritism is in favor of player B. Trivially, when $p = 0$ and $q = 1$, the game F-APA coincides with the classical APA and no player has favoritism nor advantage. Hereinafter, we commonly address p and q as the favoritism parameters. Note also that Definition 6.2.2 can be easily extended to the all-pay auctions with favoritism involving more than two players/bidders; however, in this work, we only analyze the two-player F-APA since it relates directly to the generalized-rule CB game $\mathcal{GR}\text{–}\mathcal{CB}_n^C$—that is our main focus.

Characterizing the equilibrium is one of the main focuses of the literature of APA and F-APA. For the (classical) APA, its equilibria have been completely characterized by Baye, Kovenock, and De Vries (1994) and Hillman and Riley (1989) (in games with any number of bidders); for the sake of reference, the results in the two-player APA is rewritten in our notation as follows:

Proposition 6.2.3 (extracted from Baye, Kovenock, and De Vries (1994) and Hillman and Riley (1989))**.** *In the two-player all-pay auction (i.e., an F-APA with $p = 0$, $q = 1$ and $\alpha = 1/2$), if $u^A > u^B$, there exists a unique mixed equilibrium where players A and B bid according to the following distributions:*

$$G_A(x) = \begin{cases} \frac{x}{u^B}, \forall x \in \left[0, u^B\right], \\ 1 \quad, \forall x > u^B, \end{cases} \quad \text{and} \quad G_B(x) = \begin{cases} \frac{u^A - u^B}{u^A} + \frac{x}{u^A}, \forall x \in \left(0, u^B\right] \\ 1 \qquad, \forall x > u^B. \end{cases}$$

In this equilibrium, player A's payoff is $\Pi^A_{APA} = u^A - u^B$, player B's payoff is $\Pi^B_{APA} = 0$.

Intuitively, $G_A(x)$ is the uniform distribution on $[0, u^B]$ and $G_B(x)$ is the distribution with a (strictly positive) probability mass at 0 and the remaining mass is distributed uniformly in $(0, u^B]$. In the case where $u^B > u^A$, players exchange their roles and a similar statement to Proposition 6.2.3 can be easily deduced by symmetry.

A note on terminology: the name "all-pay auction with favoritism" is adopted from Fu and Wu (2019); it is also referred to as the APA with head-starts and handicaps by Kirkegaard (2012) and as the APA with incumbency advantages by Konrad (2002). A review on the literature of all-pay auctions with favoritism was presented in Section 3.3.4. In particular, Konrad (2002) studies the case where players assess the item with the same value and where the tie-breaking rule is sharing the value equally among the bidders; in this setting, the equilibria is characterized only for the cases where both kinds of favoritism is in favor of one player. For the sake of reference, we rewrite this result in our notation as follows:

Proposition 6.2.4 (extracted from Konrad (2002)). *In the F-APA where* $u^A = u^B = u$, $p > 0$, $0 < q < 1$ *(i.e., player A has both kind of favoritism) and* $\alpha = 1/2$,

- (i) *If* $qu - p \leq 0$, *there exists a unique pure equilibrium where players' bids are* $x^A = x^B = 0$ *and their equilibrium payoffs are* $\Pi^A_{\text{F-APA}} = u$ *and* $\Pi^B_{\text{F-APA}} = 0$.
- (ii) *If* $0 < qu - p$, *there exists no pure equilibrium; the unique mixed equilibrium is where players A and B draw their bids from the following distributions:*

$$G_{\bar{A}}(x) = \begin{cases} \frac{p}{qu} + \frac{x}{qu}, \forall x \in \left[0, qu - p\right], \\ 1 \qquad\qquad , \forall x > qu - p, \end{cases} \quad \text{and} \quad G_{\bar{B}}(x) = \begin{cases} 1 - q + \frac{p}{u}, \forall x \in \left[0, \frac{p}{q}\right) \\ 1 - q + \frac{q \cdot x}{u}, \forall x \in \left[\frac{p}{q}, u\right], \\ 1 \qquad\qquad , \forall x > u. \end{cases}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = u(1-q) + p$ *and* $\Pi^B_{\text{F-APA}} = 0$.

Intuitively, $G_{\bar{A}}(x)$ is the distribution placing a positive mass at 0 and distributing the remaining mass uniformly on $(0, qu - p]$ and $G_{\bar{B}}(x)$ is the distribution placing a mass at 0 and distributing the remaining mass uniformly on $\left(p/q, u\right]$. It is easy to deduce similar results for the case where $p < 0$ and $q > 1$, i.e., player B has both kind of favoritism (it is not stated explicitly in Konrad (2002)).

The formulation of the F-APA game that we introduce in Definition 6.2.2 is more general than the definitions of the (two-player) models found in the works mentioned above; notably, we allow players to have different evaluations on the item's values and we include a general tie-breaking rule (with a generic parameter α). This extension, especially with asymmetry item' values, is essential for the analysis of the $\mathcal{GR}-\mathcal{CB}^C_n$ game. To the best of our knowledge, except from our work, only Siegel (2009) studies an F-APA game with asymmetric values, a general tie-breaking rule, and where a player has one kind of favoritism and her opponent has the other. However, Siegel (2009) does not explicitly construct the equilibrium strategies of the bidders in these cases.

6.2.2 Exact Equilibria of the All-pay Auction with Favoritism

In this section, we analyze in detail and give explicit solutions for the exact equilibria of the F-APA game with any parameters configuration (that are important for our study of the approximate equilibria of the $\mathcal{GR}-\mathcal{CB}^C_n$ game presented in Section 6.2.3 and Section 6.2.4). To do this, we need to consider the F-APA in several cases of its parameters configuration.

The F-APA game with $p \geq 0$ and any $q > 0$.

Theorem 6.2.5. *In the F-APA game where* $p \geq 0$, *we have the following results:*

(i) *If* $qu^B - p \leq 0$, *there exists a unique pure equilibrium where players' bids are* $x^A = x^B = 0$ *and their equilibrium payoffs are* $\Pi^A_{\text{F-APA}} = u^A$ *and* $\Pi^B_{\text{F-APA}} = 0$ *respectively.*

(ii) *If* $0 < qu^B - p \leq u^A$, *there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution* $G_{A_2^+}$ *(resp.* $G_{B_2^+}$*) defined as follows:*

$$G_{A_2^+}(x) = \begin{cases} \frac{p}{qu^B} + \frac{x}{qu^B}, \forall x \in \left[0, qu^B - p\right], \\ 1 \qquad\qquad , \forall x > qu^B - p, \end{cases} \tag{6.8}$$

$$\textit{and } G_{B_2^+}(x) = \begin{cases} 1 - \frac{qu^B}{u^A} + \frac{p}{u^A}, \forall x \in \left[0, \frac{p}{q}\right), \\ 1 - \frac{qu^B}{u^A} + \frac{q \cdot x}{u^A}, \forall x \in \left[\frac{p}{q}, u^B\right], \\ 1 \qquad\qquad\qquad , \forall x > u^B. \end{cases} \tag{6.9}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = u^A - qu^B + p$ *and* $\Pi^B_{\text{F-APA}} = 0$.

(iii) *If* $qu^B - p > u^A$, *there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution* $G_{A_3^+}$ *(resp.* $G_{B_3^+}$*) defined as follows:*

$$G_{A_3^+}(x) = \begin{cases} 1 - \frac{u^A}{qu^B} + \frac{x}{qu^B}, \forall x \in \left[0, u^A\right], \\ 1 \qquad\qquad\qquad , \forall x > u^A, \end{cases} \tag{6.10}$$

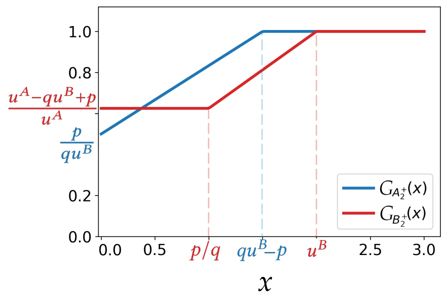
$$\textit{and } G_{B_3^+}(x) = \begin{cases} 0 \qquad\qquad , \forall x \in \left[0, \frac{p}{q}\right), \\ -\frac{p}{u^A} + \frac{q \cdot x}{u^A}, \forall x \in \left[\frac{p}{q}, \frac{u^A + p}{q}\right], \\ 1 \qquad\qquad , \forall x > \frac{u^A + p}{q}. \end{cases} \tag{6.11}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = 0$ *and* $\Pi^B_{\text{F-APA}} = u^B - (u^A + p)/q$.

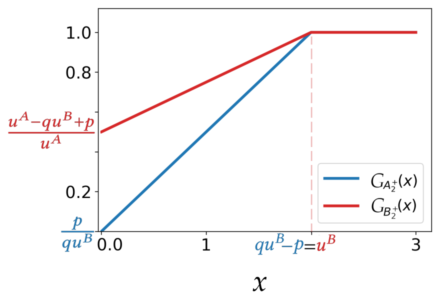
A detailed proof of Theorem 6.2.5 in given in Appendix D. In the notations used for the distributions involved in this theorem, the superscript $^+$ refers to the condition $p \geq 0$ that is being considered (to distinguish with the case with $p < 0$ presented later) and the subscript index—either 2 or 3—simply indicates that these distributions are the equilibrium in the Result (ii) or (iii). Note that we can easily verify that all the functions $G_{A_2^+}, G_{B_2^+}, G_{A_3^+}$ and $G_{B_3^+}$ in Results (ii) and (iii) of Theorem 6.2.5 satisfy the conditions of being distributions (under the corresponding conditions of the game's parameters in each case); these functions are also continuous on $[0, \infty)$.

Now, we present here an intuition for the equilibria presented in Theorem 6.2.5. First, we observe that in all equilibria, player A does not bid more than $\min\{u^A, qu^B - p\}$ and player B does not bid more than $\min\{u^B, (u^A + p)/q\}$. This can be explained by the observation that *no player has an incentive to bid more than the value* she assesses on the item: if player A (resp. player B) bids strictly more than u^A (resp. u^B), she surely receives a negative payoff (even if she wins the item); on the other hand, by bidding

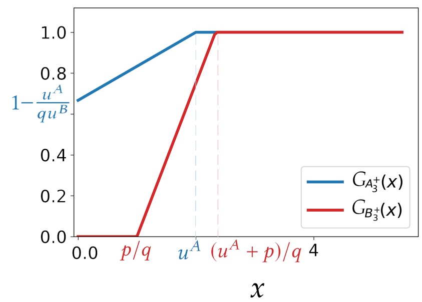
0, she can always guarantee to have a non-negative payoff.[12] Knowing this, player A (resp. player B) does not need to bid more than $qu^B - p$ (resp. $(u^A + p)/q$) in order to win the item. The condition in Result (i) of this theorem implies that player A has an advantage that is larger than the value assessed on the item by player B, i.e., player A can win with any positive bid; naturally, it is optimal for both players to bid zero. The distributions $G_{A_2^+}$, $G_{B_2^+}$, $G_{A_3^+}$ and $G_{B_3^+}$ in Results (ii) and (iii) of Theorem 6.2.5 all relate to the uniform distributions. By drawing the bid from $G_{A_2^+}$, player A puts a non-negative probability mass at zero, and then uniformly distributes the remaining mass on the range $(0, qu^B - p]$; on the other hand, from $G_{B_2^+}$, player B puts a probability mass at zero, then uniformly distributes the remaining mass on $[p/q, u^B]$ (she chooses the bid in the range $(0, p/q)$ with zero probability). Following $G_{A_3^+}$, player A puts a probability mass at zero and uniformly distributes the remaining mass on $(0, u^A]$; while player B, by following $G_{B_3^+}$, distributes uniformly on $[p/q, (u^A + p)/q]$. Figure 6.2 illustrates the plots of the distributions $G_{A_2^+}, G_{B_2^+}, G_{A_3^+}, G_{B_3^+}$ corresponding to some instances of F-APA with $p \geq 0$ (we include the instances with $p = 0$, $q = 1$, i.e., the classical APA for the sake of comparison).



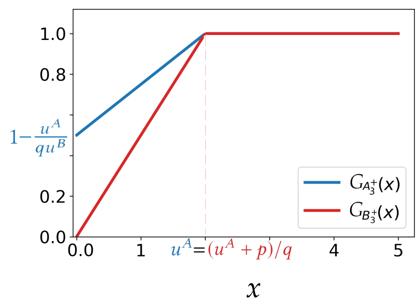
(a) F-APA instance with $u^A = 4$, $u^B = 2$, $p = 1.5$, $q = 1.5$ (i.e., $0 \leq qu^B - p < u^A$).



(b) F-APA instance with $u^A = 4$, $u^B = 2$, $p = 0$, $q = 1$ (i.e., $0 \leq qu^B - p < u^A$).



(c) F-APA instance with $u^A = 2$, $u^B = 4$, $p = 1.5$, $q = 1.5$ (i.e., $qu^B - p > u^A$).



(d) F-APA instance with $u^A = 2$, $u^B = 4$, $p = 0$, $q = 1$ (i.e., $qu^B - p > u^A$).

Figure 6.2: The mixed equilibrium of an instance of the F-APA game with $p \geq 0$.

[12] That is $\Pi^A_{\text{F-APA}}\left(x^A, x^B\right) < 0, \forall x^A > u^A, \forall x^B$ while $\Pi^A_{\text{F-APA}}\left(0, x^B\right) \geq 0$ and similarly, $\Pi^B_{\text{F-APA}}\left(x^A, x^B\right) < 0, \forall x^A, \forall x^B > u^B$ while $\Pi^B_{\text{F-APA}}\left(x^A, 0\right) \geq 0$

Moreover, in Result (ii), as long as the condition $0 < qu^B - p \leq u^A$ is satisfied, the larger p becomes and/or the smaller q gets, the larger the equilibrium payoff that player A obtains. This is in coherence with the intuition that when player A has an advantage, she can gain more. However, if p is too large (and/or q is too small) such that the condition in Result (i)-Theorem 6.2.5 is satisfied, player B will give up totally and player A gains a fixed payoff u^A even if p keeps increasing (and/or q keeps decreasing). Similar intuition can be addressed for player B and Result (iii).

It is clear that in the classical all-pay auction (i.e., the F-APA game with $p = 0, q = 1, \alpha = 1/2$), only the condition in Result (ii) of Theorem 6.2.5 is satisfied; and the equilibrium in this case coincides with the classical result stated in Proposition 6.2.3. Likewise, Result (i) and (ii) of Theorem 6.2.5 are also in coherence with Proposition 6.2.4 for the game F-APA where the item has the same value to the players.[13]

The F-APA game with $p < 0$ and any $q > 0$

Now, we consider the F-APA game where $p < 0$. The results in this case will be based on results of the case $p \geq 0$ in the previous section. We first define $p' = -p/q$ and $q' = 1/q$. If $p < 0$, we have $p' > 0$. Moreover, for any x^A, x^B, we have:

$$\beta^{\phi}\left(x^A, qx^B - p\right) = \beta^{\phi}\left(q'x^A - p', x^B\right), \forall \phi \in \{A, B\}.$$

Therefore, the F-APA game with $p < 0$, $q > 0$ is equivalent to an F-APA game with $p' > 0$, $q' > 0$ in which the roles of players are exchanged.[14] Applying Theorem 6.2.5 to this *new* game, we can deduce the following theorem:

Theorem 6.2.6. *In the F-APA game where $p < 0$, we have the following results:*

(i) If $(u^A + p)/q \leq 0$,[15] there exists a unique pure equilibrium where players' bids are $x^A = x^B = 0$ and their equilibrium payoffs are $\Pi^A_{F\text{-}APA} = 0$ and $\Pi^B_{F\text{-}APA} = u^B$ respectively.

(ii) If $0 < (u^A + p)/q \leq u^B$,[16] there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution $G_{A_2^-}$ (resp. $G_{B_2^-}$) defined as follows:

$$G_{A_2^-}(x) = \begin{cases} 1 - \frac{u^A}{qu^B} - \frac{p}{qu^B}, \forall x \in \left[0, -p\right), \\ 1 - \frac{u^A}{qu^B} + \frac{x}{qu^B}, \forall x \in \left[-p, u^A\right], \\ 1 \qquad\qquad\quad, \forall x > u^A, \end{cases} \tag{6.12}$$

$$\text{and } G_{B_2^-}(x) = \begin{cases} -\frac{p}{u^A} + \frac{qx}{u^A}, \forall x \in \left[0, \frac{u^A+p}{q}\right] \\ 1 \qquad\qquad, \forall x > \frac{u^A+p}{q}. \end{cases} \tag{6.13}$$

[13] Note that Proposition 6.2.4, extracted from Konrad (2002), does not concern the conditions involved in Result (iii) of Theorem 6.2.5.

[14] The tie-breaking rule does not change, i.e., in cases where $q'x^A - p' = x^B$, player A gains αu^A and player B gains $(1-\alpha)u^B$.

[15] That is $q'u^B - p' \leq 0$.

[16] That is $0 \leq q'u^A - p' \leq u^B$.

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = 0$ *and* $\Pi^B_{\text{F-APA}} = u^B - (u^A + p)/q$.

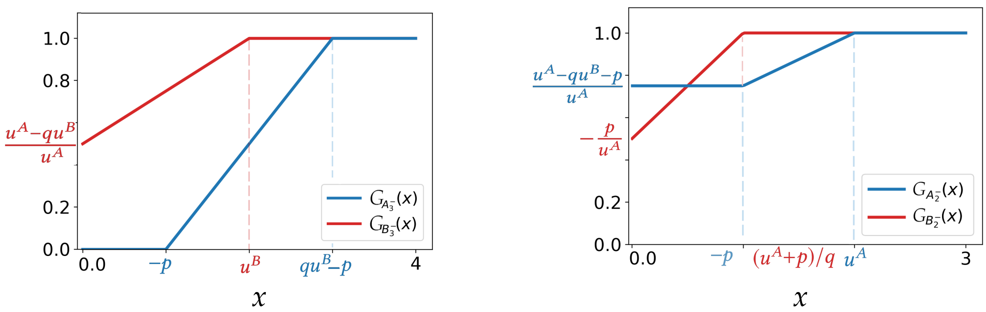
(iii) *If* $(u^A + p)/q > u^B$,[17] *there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution* $G_{A_3^-}$ *(resp.* $G_{B_3^-}$*) defined as follows:*

$$G_{A_3^-}(x) = \begin{cases} 0 & , \forall x \in [0, -p), \\ \frac{p}{qu^B} + \frac{x}{qu^B} & , \forall x \in [-p, qu^B - p], \\ 1 & , \forall x > qu^B - p, \end{cases} \tag{6.14}$$

$$\text{and } G_{B_3^-}(x) = \begin{cases} 1 - \frac{qu^B}{u^A} + \frac{q \cdot x}{u^A} & , \forall x \in [0, u^B], \\ 1 & , \forall x > u^B. \end{cases} \tag{6.15}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = u^A - qu^B + p$ *and* $\Pi^B_{\text{F-APA}} = 0$.

Similarly to the case where $p \geq 0$, we can verify that in Theorem 6.2.6, all the functions $G_{A_2^-}, G_{B_2^-}, G_{A_3^-}$ and $G_{B_3^-}$ satisfy the conditions of a distribution and they are continuous on $[0, \infty)$. The interpretation of these functions are very similar to the analysis of $G_{A_2^+}, G_{B_2^+}, G_{A_3^+}$ and $G_{B_3^+}$; and their illustrations in some instances of F-APA are given in Figure 6.3.



(a) F-APA instance with $u^A = 4$, $u^B = 2$, $p = -1$, $q = 1$ (i.e., $(u^A + p)/q > u^B$).

(b) F-APA instance with $u^A = 2$, $u^B = 4$, $p = -1$, $q = 1$ (i.e., $0 \leq (u^A + p)/q < u^B$).

Figure 6.3: The mixed equilibrium of an instance of the F-APA game with $p < 0$.

6.2.3 Optimal Univariate Distributions of the $\mathcal{GR}\text{−}\mathcal{CB}^C_n$ Game

We now turn our focus back to the generalized-rule CB game $\mathcal{GR}\text{−}\mathcal{CB}^C_n$. First, recall that in our analysis of the generalized CB game $\mathcal{CB}_n$ in Chapter 4, finding optimal univariate distributions (see Definition 4.1.1) is an essential step in constructing our proposed approximate equilibria (the IU$^{\gamma^*}$ strategies). Following this remark, in this section, to study the equilibrium (and the approximate equilibria) of the game $\mathcal{GR}\text{−}\mathcal{CB}^C_n$, we look for a set of optimal univariate distributions of $\mathcal{GR}\text{−}\mathcal{CB}^C_n$; i.e., the distributions that satisfy two conditions: by allocating according to them, players

[17] That is $q'u^A - p' > u^B$.

guarantee the budget constraints *in expectation*; it is the best response (among strategies guaranteeing budget constraints in expectation) against the allocation of the opponent in each battlefield. Based on the equilibria of the F-APA game found in the previous section, we can construct such a set of distributions. To do this, we first introduce new notation and definitions. Given a $\mathcal{GR}\text{−}\mathcal{CB}_n^C$ game as defined in Definition 6.2.1, for each pair of numbers $\left(\kappa^A, \kappa^B\right) \in \mathbb{R}^2_{>0}$, let us define the following sets:

$$
\begin{aligned}
I_1^+(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i \geq 0, q_i w_i \kappa^B - p_i \leq 0, \right\} \\
I_2^+(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i \geq 0, 0 < q_i w_i \kappa^B - p_i \leq w_i \kappa^A \right\} \\
I_3^+(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i \geq 0, q_i w_i \kappa^B - p_i > w_i \kappa^A \right\} \\
I_1^-(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i < 0, w_i \kappa^A \leq -p_i \right\} \\
I_2^-(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i < 0, -p_i < w_i \kappa^A \leq q_i w_i \kappa^B - p_i \right\} \\
I_3^-(\kappa^A, \kappa^B) &:= \left\{i \in [n] : p_i < 0, w_i \kappa^A > q_i w_i \kappa^B - p_i \right\}
\end{aligned}
$$

It is trivial to see that $\left[\bigcup_{j=1}^{3} I_j^+(\kappa^A, \kappa^B)\right] \bigcup \left[\bigcup_{j=1}^{3} I_j^-(\kappa^A, \kappa^B)\right] = [n]$ for any $(\kappa^A, \kappa^B) \in \mathbb{R}^2_{>0}$, i.e., these sets are disjoint and they contain the indices of all the battlefields in the game $\mathcal{GR}\text{−}\mathcal{CB}_n^C$. Hereinafter, we refer to these 6 sets by the *indices sets*. Note that the superscript $^+$ (resp. $^-$) in the notation of these sets implies that they include battlefield i with the corresponding parameter $p_i \geq 0$ (resp. $p_i < 0$). The subscript $j \in \{1, 2, 3\}$ in the notations I_j^+ (resp. notations I_j^-) indicates that the conditions in I_j^+ (resp. I_j^-) corresponds to the conditions in the Results (i), (ii) or (iii) of Theorem 6.2.5 (resp. Theorem 6.2.6) where we replace $p = p_i, q = q_i$, $u^A = w_i \cdot \kappa^A$ and $u^B = w_i \cdot \kappa^B$. Note also that depending on the parameters configurations of $\mathcal{GR}\text{−}\mathcal{CB}_n^C$ and values of κ^A, κ^B, these sets can be empty. Henceforth, in places where it is not necessary to state the particular parameter κ^A, κ^B, we lighten the notation by writing I_j^+ and I_j^- instead of $I_j^+(\kappa^A, \kappa^B)$ and $I_j^-(\kappa^A, \kappa^B)$.

Now, based on the sets introduced above, we have the following definition.

Definition 6.2.7. *Given* $(\kappa^A, \kappa^B) \in \mathbb{R}^2_{>0}$ *and a game* $\mathcal{GR}\text{−}\mathcal{CB}_n^C$*, for each* $i \in [n]$*, we define* $G_{A_i^{\kappa^A,\kappa^B}}$ *and* $G_{B_i^{\kappa^A,\kappa^B}}$ *to be the pair of distributions that forms the equilibrium of the all-pay auction with favoritism F-APA where* $p = p_i$, $q = q_i$, $u^A = w_i \cdot \kappa^A$ *and* $u^B = w_i \cdot \kappa^B$*. The explicit formulas of these distributions are given in Table 6.2, defined for each configuration of* w_i, p_i, q_i, κ^A *and* κ^B *(i.e., when battlefield* i *belongs to one of the indices set* I_j^+ *or* I_j^-*,* $j \in \{1, 2, 3\}$*).*

From this definition and the fact that the profile $\left(G_{A_i^{\kappa^A,\kappa^B}}, G_{B_i^{\kappa^A,\kappa^B}}\right)$ is the equilibrium of the corresponding F-APA game, we can easily prove the following: in a game $\mathcal{GR}\text{−}\mathcal{CB}_n^C$, player A (resp. player B) has no pure strategy (and no feasible mixed strategy) that provides her a better payoff than when she draws her allocation in each battlefield $i \in [n]$ from the distribution $G_{A_i^{\kappa^A,\kappa^B}}$ (resp. $G_{B_i^{\kappa^A,\kappa^B}}$), given that her opponent's allocation to this battlefield follows $G_{B_i^{\kappa^A,\kappa^B}}$ (resp. $G_{A_i^{\kappa^A,\kappa^B}}$). This implies that for any $(\kappa^A, \kappa^B) \in \mathbb{R}^2_{>0}$, the set $\left\{G_{A_i^{\kappa^A,\kappa^B}}, G_{B_i^{\kappa^A,\kappa^B}}\right\}_{i \in [n]}$ satisfies the second condition in the definition of optimal univariate distributions of the game $\mathcal{GR}-\mathcal{CB}_n^C$. To check if $\left\{G_{A_i^{\kappa^A,\kappa^B}}, G_{B_i^{\kappa^A,\kappa^B}}\right\}_{i \in [n]}$ are optimal univariate distributions or not, we need to check the first condition, that is to answer the following question: when will it happen that by drawing allocations from $\left\{G_{A_i^{\kappa^A,\kappa^B}}\right\}_{i \in [n]}$ (resp. $\left\{G_{A_i^{\kappa^A,\kappa^B}}\right\}_{i \in [n]}$), player A (resp. player B) can guarantee that her budget constraint holds in expectation? In other words, we need to solve the following system of equation (with variables κ^A, κ^B):

$$\begin{cases} \sum_{i \in [n]} \mathbb{E}_{x \sim G_{A_i^{\kappa^A,\kappa^B}}}[x] = X^A, \\ \sum_{i \in [n]} \mathbb{E}_{x \sim G_{B_i^{\kappa^A,\kappa^B}}}[x] = X^B, \end{cases} \tag{6.16}$$

Applying the definition of $G_{A_i^{\kappa^A,\kappa^B}}$ and $G_{B_i^{\kappa^A,\kappa^B}}$, we can rewrite the system (6.16) as the following system of equations:

$$\begin{cases} f^A(\kappa^A, \kappa^B) = \kappa^A, \\ f^B(\kappa^A, \kappa^B) = \kappa^B, \end{cases} \tag{6.17}$$

where $f^A, f^B : \mathbb{R}^2 \to \mathbb{R}$ are defined as follows (for each given instance of the $\mathcal{GR}-\mathcal{CB}_n^C$ game):

$$f^A\left(\kappa^A, \kappa^B\right) = \frac{1}{X^B}\left[\sum_{i \in I_2^+(\kappa^A,\kappa^B)} \frac{\left(q_i w_i \kappa^B\right)^2 - \left(p_i\right)^2}{2 q_i w_i} + \sum_{i \in I_3^+(\kappa^A,\kappa^B)} \frac{\left(w_i \kappa^A + p_i\right)^2 - \left(p_i\right)^2}{2 q_i w_i}\right] + \frac{1}{X^B}\left[\sum_{i \in I_2^-(\kappa^A,\kappa^B)} \frac{\left(w_i \kappa^A + p_i\right)^2}{2 q_i w_i} + \sum_{i \in I_3^-(\kappa^A,\kappa^B)} \frac{\left(q_i w_i \kappa^B\right)^2}{2 q_i w_i}\right],$$

$$\text{and } f^B\left(\kappa^A, \kappa^B\right) = \frac{1}{X^A}\left[\sum_{i \in I_2^+(\kappa^A,\kappa^B)} \frac{\left(q_i w_i \kappa^B - p_i\right)^2}{2 q_i w_i} + \sum_{i \in I_3^+(\kappa^A,\kappa^B)} \frac{\left(w_i \kappa^A\right)^2}{2 q_i w_i}\right] + \frac{1}{X^A}\left[\sum_{i \in I_2^-(\kappa^A,\kappa^B)} \frac{\left(w_i \kappa^A\right)^2 - p_i^2}{2 q_i w_i} + \sum_{i \in I_3^-(\kappa^A,\kappa^B)} \frac{\left(q_i w_i \kappa^B - p_i\right)^2 - p_i^2}{2 q_i w_i}\right].$$

We can have a trivial proposition:

Proposition 6.2.8. *Given a game $\mathcal{GR}-\mathcal{CB}_n^C$, assume that System (6.17) has a positive solution $(\kappa^A_*, \kappa^B_*) \in \mathbb{R}^2_{>0}$; then the distributions $\left\{G_{A_i^{\kappa^A_*,\kappa^B_*}}, G_{B_i^{\kappa^A_*,\kappa^B_*}}\right\}_{i \in [n]}$ that correspond to (κ^A_*, κ^B_*) are optimal univariate distributions of the game.*

Table 6.2: Uniform-type distributions of the $\mathcal{GR}-\mathcal{CB}_n^C$ game. Here, F-APA* is the all-pay auction with favoritism where $p = p_i, q = q_i, u^A = w_i \cdot \kappa^A$ and $u^B = w_i \cdot \kappa^B$.

Conditions	Definition	Notes
$i \in I_1^+(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) := 1, \forall x \geq 0,$ $G_{B_i^{\kappa^A,\kappa^B}}(x) := 1, \forall x \geq 0.$	
$i \in I_2^+(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) := \begin{cases} \frac{p_i}{q_i w_i \kappa^B} + \frac{x}{q_i w_i \kappa^B}, \forall x \in \left[0, q_i w_i \kappa^B - p_i\right], \\ 1 \quad , \forall x > q_i w_i \kappa^B - p_i, \end{cases}$ $G_{B_i^{\kappa^A,\kappa^B}}(x) := \begin{cases} 1 - \frac{q_i \kappa^B}{\kappa^A} + \frac{p_i}{w_i \kappa^A}, \forall x \in \left[0, \frac{p_i}{q_i}\right), \\ 1 - \frac{q_i \kappa^B}{\kappa^A} + \frac{q_i \cdot x}{w_i \kappa^A}, \forall x \in \left[\frac{p_i}{q_i}, w_i \kappa^B\right], \\ 1 \quad , \forall x > w_i \kappa^B. \end{cases}$	corrsp. to $G_{A_2^+}$ and $G_{B_2^+}$ of F-APA*.
$i \in I_3^+(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) = \begin{cases} 1 - \frac{\kappa^A}{q_i \kappa^B} + \frac{x}{q_i w_i \kappa^B}, \forall x \in \left[0, w_i \kappa^A\right], \\ 1 \quad , \forall x > w_i \kappa^A, \end{cases}$ $G_{B_i^{\kappa^A,\kappa^B}}(x) = \begin{cases} 0 \quad , \forall x \in \left[0, \frac{p_i}{q_i}\right), \\ -\frac{p_i}{w_i \kappa^A} + \frac{q_i \cdot x}{w_i \kappa^A}, \forall x \in \left[\frac{p_i}{q_i}, \frac{w_i \kappa^A + p_i}{q_i}\right], \\ 1 \quad , \forall x > \frac{w_i \kappa^A + p_i}{q_i}. \end{cases}$	corrsp. to $G_{A_3^+}$ and $G_{B_3^+}$ of F-APA*.
$i \in I_1^-(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) := 1, \forall x \geq 0,$ $G_{B_i^{\kappa^A,\kappa^B}}(x) := 1, \forall x \geq 0.$	
$i \in I_2^-(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) := \begin{cases} 1 - \frac{\kappa^A}{q_i \kappa^B} - \frac{p_i}{q_i w_i \kappa^B}, \forall x \in \left[0, -p_i\right), \\ 1 - \frac{\kappa^A}{q_i \kappa^B} + \frac{x}{q_i w_i \kappa^B}, \forall x \in \left[-p_i, w_i \kappa^A\right], \\ 1 \quad , \forall x > w_i \kappa^A, \end{cases}$ $G_{B_i^{\kappa^A,\kappa^B}}(x) : \begin{cases} -\frac{p_i}{w_i \kappa^A} + \frac{q_i \cdot x}{w_i \kappa^A}, \forall x \in \left[0, \frac{w_i \kappa^A + p_i}{q_i}\right], \\ 1 \quad , \forall x > \frac{w_i \kappa^A + p_i}{q_i}. \end{cases}$	corrsp. to $G_{A_2^-}$ and $G_{B_2^-}$ of F-APA*.
$i \in I_3^-(\kappa^A, \kappa^B)$	$G_{A_i^{\kappa^A,\kappa^B}}(x) = \begin{cases} 0 \quad , \forall x \in \left[0, -p\right), \\ \frac{p}{q u^B} + \frac{x}{q u^B}, \forall x \in \left[-p, q u^B - p\right], \\ 1 \quad , \forall x > q u^B - p, \end{cases}$ $G_{B_i^{\kappa^A,\kappa^B}}(x) = \begin{cases} 1 - \frac{q u^B}{u^A} + \frac{q \cdot x}{u^A}, \forall x \in \left[0, u^B\right], \\ 1 \quad , \forall x > u^B. \end{cases}$	corrsp. to $G_{A_3^-}$ and $G_{B_3^-}$ of F-APA*.

In principle, if we can construct an n-variate joint distribution of these optimal univariate distributions such that any realization satisfies the budget constraint, we have found an exact equilibrium of $\mathcal{GR}-\mathcal{CB}_n^C$. However, the construction of this joint distribution is as challenging as in the case of the $\mathcal{CB}_n$ game and it still remains unknown (even its existence is still an open question). As a different perspective, we can apply a similar idea to that of the IU$^{\gamma^*}$ strategy in the $\mathcal{CB}_n$ game (see Chapter 4) to construct an approximate equilibrium of the $\mathcal{GR}-\mathcal{CB}_n^C$ game with a well-controlled approximation error by letting player A (resp. player B) draw independently n numbers from $\left\{G_{A_i^{\kappa_*^A,\kappa_*^B}}\right\}_{i\in[n]}$ (resp. $\left\{G_{B_i^{\kappa_*^A,\kappa_*^B}}\right\}_{i\in[n]}$) then rescale them to guarantee the budget constraint. However, all these results are obtained only under the assumption that System (6.17) has a positive solution and that we can compute it (efficiently).

Therefore, the question of interest now becomes *checking whether System* (6.17) *has positive solutions* $(\kappa^A, \kappa^B) \in \mathbb{R}^2_{>0}$ *and compute these solutions if they exist.*[18] Note that in System (6.17), when we change the values of the variables (κ^A, κ^B), the indices sets I_j^+ and I_j^- ($j \in \{1, 2, 3\}$) also change and the conditions determining whether a battlefield i belongs to a summation involved in computing $f^A(\kappa^A, \kappa^B)$ and $f^B(\kappa^A, \kappa^B)$ also change. Therefore, *System* (6.17) *is* not *simply a system of quadratic equations*. Only when we *fix* a configuration of partitioning the numbers $\{1, 2, \ldots, n\}$ into the sets I_1^+, I_2^+, I_3^+ and I_1^-, I_2^-, I_3^-, then (6.17) becomes a system of quadratic equations. Therefore, in principle, a simple way to solve (6.17) is to consider all the possible configuration of I_i^+ and I_i^-, then solve the particular system corresponding to each of these cases. However, this approach is extremely inefficient due to the fact that in the worst cases, the number of ways to partition $\{1, 2, \ldots, n\}$ into 6 indices sets is exponential in terms of n. The following toy example (Example 6.2.9) illustrates the inefficiency of this approach.

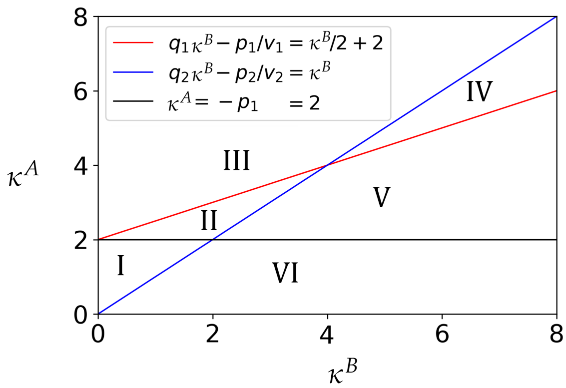


Figure 6.4: Conditions to partition the battlefields into the indices sets of the $\mathcal{GR}\text{−}\mathcal{CB}_n^C$ game in Example 6.2.9.

Example 6.2.9. *Consider the game* $\mathcal{GR}\text{−}\mathcal{CB}_n^C$ *with* $n = 2$; $X^A = X^B = 2$; $w_1 = w_2 = 1$; $p_1 = -2, p_2 = 0$ *and* $q_1 = 1/2, q_2 = 1$. *Even in this simple game, there are 6 possible configurations of* I_j^+ *and* I_j^- ($j \in \{1, 2, 3\}$). *In Figure 6.4, we illustrate these cases by the 6 regions in the first quadrant of the* κ^A-κ^B *plane separated by the axes and the polynomial involved in the conditions that determines the indices sets. We need to consider each of these cases. For example, when* $0 \le \kappa^A, \kappa^B \le 2$ *and* $\kappa^B \le \kappa^A$ *(i.e., when the point* (κ^A, κ^B) *lies within Region I in Figure 6.4), we have* $I_1^- = \{1\}$, $I_2^+ = \{2\}$ *and all other indices sets are empty. In this*

[18] Note that $(0, 0)$ is a trivial solutions of System (6.17); however, we need a (strictly) positive solution (κ^A, κ^B) in order to apply the results in the equilibrium analysis of the F-APA game (see Theorem 6.2.5 and Theorem 6.2.6).

case, (6.17) *becomes*

$$\begin{cases} (\kappa^B)^2/2 = 2\kappa^A, \\ (\kappa^B)^2/2 = 2\kappa^B. \end{cases}$$

This system has a unique solution in $\mathbb{R}^2_{>0}$*, which is* $\kappa^A = \kappa^B = 4$*. However, this solution does not satisfy the condition* $0 \le \kappa^A, \kappa^B \le 2$*; therefore,* (6.17) *has no positive solution in Region I.*

Similarly, we can check for all the remaining cases. We see that there is no positive solution of (6.17) *in the cases corresponding to Regions II, III, IV and VI of Figure 6.4. Only when* $2 < \kappa^A < \min\{\kappa^B, \kappa^B/2 + 2\}$ *(i.e., the point* (κ^A, κ^B) *lies within Region V in Figure 6.4), we have* $I_2^- = \{1\}$ *and* $I_3^+ = \{2\}$ *and thus, System* (6.17) *becomes:*

$$\begin{cases} (\kappa^A)^2/2 + (\kappa^A - 2)^2 = 2\kappa^A, \\ (\kappa^A)^2/2 + (\kappa^A)^2 - 2^2 = 2\kappa^B. \end{cases}$$

This system has a unique solution in $\mathbb{R}^2_{>0}$ *that is* $\kappa^A = 2 + \sqrt{4/3}$ *and* $\kappa^B = 2 + \sqrt{12}$*. This solution satisfies the conditions in this case; therefore, it is a positive solution of System* (6.17)*.*

6.2.4 Heuristic Algorithms Finding an Approximate Solution of System (6.17)

We aim to find a more efficient method to solve System (6.17) or at least to efficiently find a good approximation of one of its positive solution in order to construct an approximate equilibrium of the game $\mathcal{GR}\text{−}\mathcal{CB}_n^C$. First, we address the question: does there exist a positive solution of System (6.17)? Recall that in Chapter 4, to obtain the optimal univariate distributions of the generalized CB game $\mathcal{CB}_n$, we also need to resolve Equation (4.5) (with the variable γ). To prove the existence of a positive solution of this equation, Kovenock and Roberson (2015) convert it to the form $f(\gamma) = \gamma$ and use the intermediate value theorem to prove that the function f has a fixed point. Generalizing this approach, for each game instance $\mathcal{GR}\text{−}\mathcal{CB}_n^C$, we define the function:

$$\begin{aligned} F\colon \mathbb{R}^2 &\to \mathbb{R}^2 \\ (\kappa^A, \kappa^B) &\mapsto F(\kappa^A, \kappa^B) = \left(f^A(\kappa^A, \kappa^B) - \kappa^A, f^B(\kappa^A, \kappa^B) - \kappa^B\right). \end{aligned}$$

We also denote $F^A(\kappa^A, \kappa^B) := f^A(\kappa^A, \kappa^B) - \kappa^A$ and $F^B(\kappa^A, \kappa^B) := f^B(\kappa^A, \kappa^B) - \kappa^B$. Then, System (6.17) can be rewritten under the form $F(\kappa^A, \kappa^B) = (0, 0)$. Therefore, to prove that System (6.17) has a positive solution, we only need to prove that F has a positive zero.[19] A possible generalization of the intermediate value theorem to the cases involved functions with 2-dimensional inputs and outputs is the *Poincaré-Miranda* theorem (see e.g., Kulpa (1997)). Unfortunately, we have not succeeded in proving that the sufficient condition of this theorem is satisfied by the function F corresponding to any instance of the game $\mathcal{GR}\text{−}\mathcal{CB}_n^C$; therefore, whether one can use the Poincaré-Miranda theorem to prove the existence of positive solution of System (6.17) is an open question.[20]

[19] A zero of a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ is any point (x, y) such that $F(x, y) = (0, 0)$.

[20] Note that, even in the case of the generalized CB game $\mathcal{CB}_n$ where it requires to find a fixed-point of a 1-dimensional inputs and outputs function, which is considerably easier, the Brouwer's fixed point theorem (and the Poincaré-Miranda theorem) is *not* applicable (see Kovenock and Roberson (2015) for more details).

Another approach is to use the *main theorem of connectedness* in topology, stated briefly as follows: let X and Y be topological spaces and let $g : X \to Y$ be a continuous function; if X is connected then the image $g(X)$ is also connected (see e.g., Viro et al. (2008) for a definition of connectedness). It is easy to check that the function F in any $\mathcal{GR}-\mathcal{CB}_n^C$ game is continuous (with respect to the standard topology of $\mathbb{R}^2$); therefore, if we can find a subset of $\mathbb{R}^2$, say D, in the first quadrant of the κ^A-κ^B plane[21] such that it is a connected space (e.g., D can be a convex subset of $\mathbb{R}^2$) and that its image $F(D)$ ($\subset \mathbb{R}^2$) contains the point $(0,0)$, we can deduce that there exists a zero of F in D. However, to prove (theoretically) the existence of such a set D for any instance of the game $\mathcal{GR}-\mathcal{CB}_n^C$ turns out to be non-trivial and we leave it here as an open question.

As a numerical solution for this question, we propose a heuristic procedure (Algorithm 8) to find such a set D in $\mathbb{R}^2$ for the function F corresponding to any instance of the game $\mathcal{GR}-\mathcal{CB}_n^C$ (if it exists). Moreover, this heuristic procedure also helps us quickly compute an approximate (positive) solution of System (6.17) with an arbitrary small approximation error $\tilde{\delta} > 0$; that is, given a game $\mathcal{GR}-\mathcal{CB}_n^C$, we can find a point $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ such that

$$\begin{cases} \tilde{\kappa}^A > \tilde{\delta}, \tilde{\kappa}^B > \tilde{\delta}, \\ \left|F^A(\kappa^A, \kappa^B)\right| = \left|f^A(\tilde{\kappa}^A, \tilde{\kappa}^B) - \kappa^A\right| \le \tilde{\delta}, \\ \left|F^B(\kappa^A, \kappa^B)\right| = \left|f^B(\tilde{\kappa}^A, \tilde{\kappa}^B) - \kappa^B\right| \le \tilde{\delta}. \end{cases} \tag{6.18}$$

Intuitively, any $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ satisfying (6.18) is bounded away from $(0,0)$ and $F(\tilde{\kappa}^A, \tilde{\kappa}^B)$ is $\tilde{\delta}$-closed to $(0,0)$. Hereinafter, we call such $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ as a $\tilde{\delta}$-approximate solution of System (6.17).

Intuitively, Algorithm 8 is a dichotomy procedure (i.e., a bisection method) where, starting from a rectangle D in the first quadrant of the κ^A-κ^B plane, we verify to see whether the point $(0,0)$ is contained in its image via the function F (i.e., whether $(0,0) \in F(D) \subset \mathbb{R}^2$). This can be done by computing the winding number of $F(D)$ around $(0,0)$.[22] If this winding number is non-zero, $F(D)$ contain $(0,0)$. Now, the rectangle D (including its boundary and interior) is a convex subset of $\mathbb{R}^2$, thus it is a connected space (with the standard subspace topology); due to the main theorem of connectedness, we know that there exists a point inside D yielding to be a zero of F (i.e., the image via F of this point is $(0,0)$). We divide the rectangle D into smaller rectangles and repeat the above procedure for each one of them. The algorithm terminates when it finds a rectangle, say D^*, such that $F(D^*)$ is small enough (with a diameter smaller than $\tilde{\delta}$) and $F(D^*)$ has a non-zero winding number around $(0,0)$. There exists a zero of F inside D^* and the value of F at the center of D^* is $\tilde{\delta}$-closed to $(0,0)$. Therefore, this center satisfies (6.18) and it is a $\tilde{\delta}$-approximate solution of (6.17).

This algorithm is heuristic since we do not have a proof guaranteeing the existence of a positive solution of System (6.17), that is a positive zero of F. Therefore, it might

[21] i.e., if $(\kappa^A, \kappa^B) \in D$ then $\kappa^A > 0$ and $\kappa^B > 0$

[22] Briefly put, the winding number of a closed curve in the 2-D plane around a given point is the integer representing the total number of times that curve travels counterclockwise around the point.

Algorithm 8: Heuristic algorithm finding a $\tilde{\delta}$-approximate solution of (6.17)

Input: $\mathcal{GR}-\mathcal{CB}_n^C$ game, $\tilde{\delta} > 0$, $M \gg \tilde{\delta}$.
Output: $(\kappa^A, \kappa^B) \in \mathbb{R}^2_{>0}$ satisfying (6.18).

1 Let D be the rectangle with four vertices $(\tilde{\delta}, \tilde{\delta}), (\tilde{\delta}, M), (M, M), (M, \tilde{\delta})$
2 Set ω_D to be the winding number of $F(D)$ around $(0, 0)$
3 **if** $\omega_D = 0$ **then**
4   M:= 2M and $\tilde{\delta} := \tilde{\delta}/2$, then repeat line 1
5 **else if** $\omega_D \neq 0$ **then**
6   Divide D into two rectangles, D_1 and D_2, with equal areas
7   Set ω_{D_1} to be the winding number of $F(D_1)$ around $(0, 0)$
8   **if** $\omega_{D_1} \neq 0$ **then**
9   **if** *diameter of* $F(D_1)$ *is less than* $\tilde{\delta}$ **then**
10   Stop and return output to be the center of D_1
11   **else**
12   Set $D := D_1$ and repeat line 6
13   **else**
14   **if** *diameter of* $F(D_2)$ *is less than* $\tilde{\delta}$ **then**
15   Stop and return output to be the center of D_2
16   **else**
17   Set $D := D_2$ and repeat line 6

happen that for certain instances of $\mathcal{GR}-\mathcal{CB}_n^C$, there exists no rectangle D such that[23] $\omega_D \neq 0$ and thus, Algorithm 8 will loop infinitely at line 4. However, we conjecture that there indeed exists a positive zero of F and that by enlarging the search space, i.e., increasing the size of the rectangle D (as in line 4), we will eventually find a rectangle large enough to contain this zero. Our numerical experiments (see below) support this conjecture: we *do not* find any instance of $\mathcal{GR}-\mathcal{CB}_n^C$ such that when using it as inputs, Algorithm 8 does not terminate. Note also that if the winding number of $F(D)$ around $(0, 0)$ is non-zero, among the smaller rectangles constituting D, there must exist at least one rectangle whose F-image also has an non-zero winding number around $(0, 0)$ (this is due to the additive property of the winding number).[24] Therefore, as long as we find a rectangle whose image has a non-zero winding number, Algorithm 8 surely terminates.[25] Note that F may have multiple (positive) zeros, but Algorithm 8 only computes an approximation of one zero among them and that in order to obtain a $\tilde{\delta}$-approximate solution with smaller and smaller $\tilde{\delta}$, it takes Algorithm 8 longer time to terminate (because of the loop in lines 6-17).

Finally, to run Algorithm 8 more efficiently, we can choose to additionally use some

[23]Recall that we denote by ω_D the winding number of $F(D)$ around $(0, 0)$.
[24]Thus, at line 13 of Algorithm 8, we know that $\omega_{D_2} \neq 0$
[25]Due to the continuity of F, when we consider a sequence of smaller and smaller rectangles, their images via F will eventually becomes small enough that Algorithm 8 terminates either by line 9 or 15.

techniques as follows. To check whether the winding number of a curve around a given point is non-zero (involved in lines 3, 5 and 8 of Algorithm 8), we can simply approximate it by the winding number of the curve's piece-wise approximation, which is a polygon. The winding number of a polygon around $(0,0)$ can be computed in $\mathcal{O}(V)$ where V is its number of vertices. The larger V is, the more precise the approximation becomes. In our numerical experiments, we heuristically set V to be 500.[26] On the other hand, the conditions at lines 9 and 14 of Algorithm 8 can be easily checked by the following observation: given a rectangle D having four vertices denoted by $(V_A^j, V_B^j) \in \mathbb{R}^2_{>0}$ $(j \in \{1,2,3,4\})$, $F(D)$ is contained inside the circle having the center $(0,0)$ and the radius $r := \max_{j \in \{1,2,3,4\}}\{|F^A(V_A^j, V_B^j)|, |F^B(V_A^j, V_B^j)|\}$.[27]

To illustrate the usefulness of Algorithm 8, we consider the following toy examples.[28]

Example 6.2.10. *We recall the simple game instance $\mathcal{GR}-\mathcal{CB}_n^C$ $(n=2)$ considered in Example 6.2.9 in which we know that the corresponding System (6.17) has one positive (exact) solution $(\kappa_*^A, \kappa_*^B) := (2+\sqrt{4/3}, 2+\sqrt{12}) \approx (3.1547005, 5.4641016)$. With the input $\tilde{\delta} = 10^{-6}$, $M = 10$, Algorithm 8 running in this game outputs the solution $(\tilde{\kappa}^A, \tilde{\kappa}^B) = (3.1547010, 5.4641018)$. We observe that not only $F(\tilde{\kappa}^A, \tilde{\kappa}^B)$ is $\tilde{\delta}$-closed to $(0,0)$ (guaranteed by Algorithm 8) but the point $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ is also $\tilde{\delta}$-closed to the exact solution (κ_*^A, κ_*^B). The computation time here is merely ~ 2.78 seconds.*

Example 6.2.11. *Consider a game $\mathcal{GR}-\mathcal{CB}_n^C$ with $n = 4$, $X^A = 4, X^B = 4$; the battlefield's values are $w_1 = w_3 = 1; w_2 = w_4 = 2$, the favoritism parameters are $p_1 = p_2 = 1$; $p_3 = p_4 = -1$ and $q_i = 1, \forall i$. We illustrate in Figure 6.5 the values of the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to this game. Figure 6.5(a) represents the output plane where each point is mapped with a color; e.g., if a point has the color blue, we know that both coordinates of this point are positive. Figure 6.5(b) presents the input plane. Function F maps each point in this input plane with a point in the output plane. Then, in Figure 6.5(b), we colorize each point in the input plane with the corresponding color of its output (colors are chosen according to Figure 6.5(a)). For example, in Figure 6.5(b), for $\kappa^A \geq 2$ and $\kappa^B \leq 1$ (corresponding to the region of green points), by looking up the output plane and find the position of green points, we know that $F(\kappa^A, \kappa^B)$ has a negative x-coordinate and a positive y-coordinate. Finally, Figure 6.5(c) illustrates the curve that is the image of a rectangle, called D, with the vertices $(1,1)$, $(1,4)$, $(4,4)$, $(4,1)$.[29]*

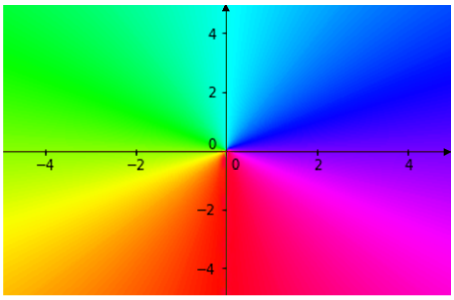
We observe that in Figure 6.5(b), when one draws a curve around the point $(2,2)$ in the input plane, this curve passes through all colors, which indicates that its image via F goes around the

[26]It would be interesting to theoretically characterize the relation between the choice of V and the accuracy of the solution of Algorithm 8. However, in this thesis, we do not focus on this question and leave it to future work.

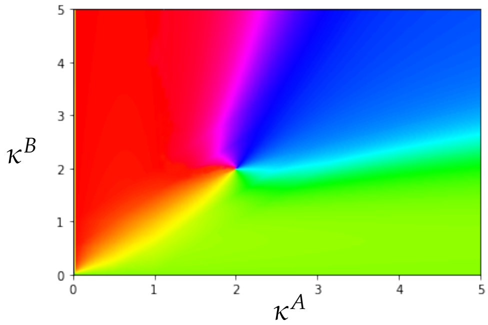
[27]This result comes directly from the definitions of f^A and f^B. An illustration of this result for an instance of the $\mathcal{GR}-\mathcal{CB}_n^C$ game is given in Figure 6.5(c).

[28]We publish the codes used in these examples at `https://github.com/dongquan11/GeneralizedRule_CBgame`.

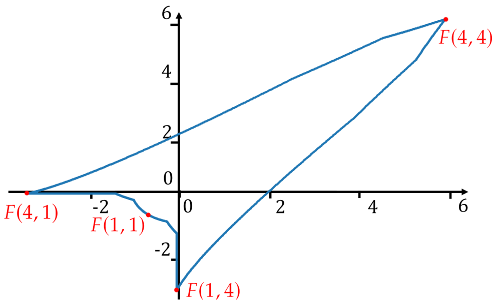
[29]For the sake of illustration, we choose D with this set-up instead of a rectangle with a vertex (e.g., $(\tilde{\delta}, \tilde{\delta})$) that is very close to $(0,0)$ because although the F-image of the latter also contains $(0,0)$, it is hard to illustrate visibly this fact.



(a) Output plane that is a heatmap to be used as reference



(b) Input plane with the colors of the values of function F



(c) The image via function F of the rectangle D

Figure 6.5: Illustration of the function F on an instance of $\mathcal{GR}\text{−}\mathcal{CB}_n^C$ (Example 6.2.11)

origin $(0,0)$ *of the output plane. This is confirmed by Figure 6.5(c) showing that the rectangle D (contains* $(2,2)$*) having the image contains* $(0,0)$*. Therefore, we "guess" that* $(2,2)$ *is a zero of F. On the other hand, with inputs* $\tilde{\delta} = 10^{-6}$ *and* $M = 4$*, Algorithm 8 (initializing with a rectangle containing D) outputs the point* $(\tilde{\kappa}^A, \tilde{\kappa}^B) = (1.9999995, 1.9999995)$ *and we can confirm that* $F(\tilde{\kappa}^A, \tilde{\kappa}^B) = (-6.80 \times 10^{-7}, -6.80510 \times 10^{-7})$ *(i.e., the error is smaller than* $\tilde{\delta}$*). Finally, by listing out all possible combination of indices sets (note again that this is inefficient), we indeed find that* $(2,2)$ *is an exact zero of F (moreover, it is unique).*

Finally, to illustrate the trade-off between the approximation error from the output of Algorithm 8 and its running time, we conduct the following experiment (running with a machine with an Intel Xeon CPU 2.20GHz and 12Gb RAM).[30] For each $n \in \{5, 10, 20, 50, 100\}$, we randomly generate 10 instances[31] of $\mathcal{GR}\text{−}\mathcal{CB}_n^C$. We then run Algorithm 8 on each game instance with the input $\tilde{\delta} \in \{10^{-1}, 10^{-2}, \ldots, 10^{-6}\}$ and $M := 10 \cdot \min\{X^A, X^B\}$; we then measure the time it takes to output the $\tilde{\delta}$-approximate

[30] We publish the codes used in this experiment at https://github.com/dongquan11/GeneralizedRule_CBgame.

[31] We choose $X^A, X^B \in \{1, 2, \ldots, 100\}$ randomly at uniform ($X^A \leq X^B$); then, for each $i \in [n]$, we randomly generate a battlefield value $w_i \sim \mathcal{U}(0, X^A]$ and with equal probability, we choose either $p_i > 0$ or $p_i =< 0$ or $p_i = 0$; then draw p_i from $\mathcal{U}(0, X^A)$ or $\mathcal{U}(-X^A, 0)$ or set it equal 0 respectively; then, with equal probability, we choose either $q_i > 1$ or $q_i \in (0,1)$ or $q_i = 1$; then draw q_i from $\mathcal{U}(1, X^A)$ or $\mathcal{U}(1/X^A, 1)$ or set it equal 1 respectively.

solution of the corresponding System (6.17). Figure 6.6 shows the average running time of Algorithm 8 taken from the 10 instances for each n and $\tilde{\delta}$. We can see that, even with large instance ($n = 100$), it does not take too much time for Algorithm 8 to compute a good approximate solution of (6.17).

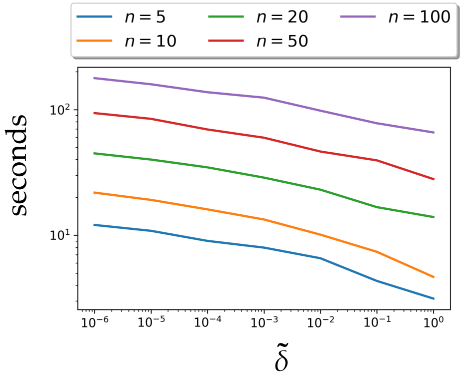


Figure 6.6: The trade-off between the running time of Algorithm 8 and $\tilde{\delta}$. Here, both the axes are drawn with log-scale.

6.2.5 Partial Results on Approximate Equilibria of the $\mathcal{GR}-\mathcal{CB}_n^C$ Game

In this section, let us denote by $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ a $\tilde{\delta}$-approximate solution of System (6.17) that is computed by Algorithm 8. For any game instance $\mathcal{GR}-\mathcal{CB}_n^C$ and any $\tilde{\delta} > 0$, we assume that this $\tilde{\delta}$-approximate solution exists and we consider as if they are constants. Now, observe that when player A (resp. B) draws allocations from $\left\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$ (resp. $\left\{G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$), in expectation, her budget constraint may be violated. However, because $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ satisfies (6.18), the margin of violation here is just $\tilde{\delta}$—a small number. Therefore, although $\left\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$ (corresponding to $\tilde{\kappa}^A, \tilde{\kappa}^B$) are not optimal univariate distributions of the game $\mathcal{GR}-\mathcal{CB}_n^C$, by using $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, \forall i \in [n]$ with a modified version of the IU^{γ^*} strategy (see Definition 4.2.1), we can still construct an approximate equilibrium of the game $\mathcal{GR}-\mathcal{CB}_n^C$ with a well-controlled approximation error. We call this the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy and present its definition as follows:

Definition 6.2.12. *Given a game $\mathcal{GR}-\mathcal{CB}_n^C$, let $(\tilde{\kappa}^A, \tilde{\kappa}^B)$ be a $\tilde{\delta}$-approximate solution of the corresponding System (6.17), for any player $\phi \in \{A, B\}$, the $\mathrm{IU}_\phi^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ is the **mixed** strategy of player ϕ where her allocation $\boldsymbol{x}^\phi$ is randomly generated from Algorithm 9.*

Hereinafter, we use $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ to commonly refer to either the $\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ or $\mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy and also the strategy profile $\left(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}, \mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right)$. The intuition about the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ in the $\mathcal{GR}-\mathcal{CB}_n^C$ game is very similar to that of the IU^{γ^*} strategy in the game $\mathcal{CB}_n$. The

Algorithm 9: $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy-generation algorithm.

Input: $n \in \mathbb{N}$, $w_i \in [\underline{w}, \bar{w}], \forall i \in [n]$, budgets X^A, X^B, $(\tilde{\kappa}^A, \tilde{\kappa}^B) \in \mathbb{R}^2_{>0}$ satisfying (6.18)

Output: $\boldsymbol{x}^A, \boldsymbol{x}^B \in \mathbb{R}^n_{\geq 0}$

Draw $a_i \sim G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, b_i \sim G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, \forall i \in [n]$ independently
if $\sum_{j\in[n]} a_j = 0$ **then**
  $x_i^A := 0, \forall i \in [n]$
else
  $x_i^A := \frac{a_i}{\sum_{j\in[n]} a_j} X^A, \forall i \in [n]$
if $\sum_{j\in[n]} b_j = 0$ **then**
  $x_i^B := 0, \forall i \in [n]$
else
  $x_i^B := \frac{b_i}{\sum_{j\in[n]} b_j} X^B, \forall i \in [n]$

main difference is that in the $\mathrm{IU}_{\phi}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy, player ϕ draws n independent numbers from $\left\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$ (instead of from $F_{A_i^*}$ as in the IU^{γ^*} strategy) before rescaling them to guarantee the budget constraint. We emphasize that the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ is only implicitly defined by Algorithm 9 and each output of this algorithm is a realization of the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy. Finally, we have the following result:

Theorem 6.2.13. *In any game $\mathcal{GR}-\mathcal{CB}_n^C$, under Assumption (A0), there exists a positive number $\varepsilon = \tilde{\mathcal{O}}\left(n^{-1/2}\right)$ such that if there exists $(\tilde{\kappa}^A, \tilde{\kappa}^B) \in \mathbb{R}^2_{>0}$ satisfying (6.18) (i.e., a $\tilde{\delta}$-approximate solution of (6.17)), the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy is an $(\varepsilon + \tilde{\delta})W$-equilibrium.*

The proof of Theorem 6.2.13 is presented in Appendix D. This proof is based on the proof of Theorem 4.2.3 (showing that the IU^{γ^*} strategies are approximate equilibria of the $\mathcal{CB}_n$ game) and the main difference is that proving Theorem 6.2.13 requires to justify that the error $\tilde{\delta}$ in approximating a solution of System (6.17) only contributes an additional term $\tilde{\delta}W$ into the approximation error in the players' payoffs where they use the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy. From Theorem 6.2.13, we make the following observation: as the number of battlefields n increases, ε decreases (the relation of these terms are presented as $\varepsilon = \tilde{\mathcal{O}}(n^{-1/2})$), thus, the level of approximation error $\tilde{\delta} + \varepsilon$ also decreases. Therefore, assuming that System (6.17) has a positive solution, we can run Algorithm 8 with the input $\tilde{\delta} = \varepsilon$ to obtain $(\tilde{\kappa}^A, \tilde{\kappa}^B)$; in this case, the approximation error in using the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy in $\mathcal{GR}-\mathcal{CB}_n^C$ is $2\varepsilon W$. This error, relative to the magnitude of the players' payoffs (represented via W—the players' total payoff), is small when the number of battlefields is large.

Now, to see how the favoritism (parameters p_i and q_i) changes the outcome of the players in the Colonel Blotto games with favoritism, we can compute the approximate equilibrium's payoff. First, from the proof of Theorem 6.2.13, we note that when player

ϕ plays her $\mathrm{IU}_{\phi}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy, her allocation toward battlefield i follows the marginal distribution that uniformly converge toward the distribution $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$. Therefore, we have a lower-bound on the payoff that player A obtains when both players follow the $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy in $\mathcal{GR}-\mathcal{CB}_n^C$ as follows:

$$\begin{aligned}
&\Pi_{\mathrm{GR}}^A\left(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B},\mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right)\\
&\geq \sum_{i=1}^{n}\int_0^{\infty} w_i G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\left(\frac{x+p_i}{q_i}\right)\mathrm{d}G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}(x)-(\varepsilon+\tilde{\delta})W\\
&=\sum_{i\in I_1^+(\tilde{\kappa}^A,\tilde{\kappa}^B)} w_i\tilde{\kappa}^A+\sum_{i\in I_2^+(\tilde{\kappa}^A,\tilde{\kappa}^B)\cup I_3^-(\tilde{\kappa}^A,\tilde{\kappa}^B)}\left[w_i(\tilde{\kappa}^A-q_i\tilde{\kappa}^B)+p_i\right]-(\varepsilon+\tilde{\delta})W. \qquad (6.19)
\end{aligned}$$

Similarly, a lower-bound of player B's payoff when they play $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ is:

$$\begin{aligned}
&\Pi_{\mathrm{GR}}^B\left(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B},\mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right)\\
&\geq \sum_{i=1}^{n}\int_0^{\infty} w_i G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\left(q_i x-p_i\right)\mathrm{d}G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}(x)-(\varepsilon+\tilde{\delta})W\\
&=\sum_{i\in I_3^+(\tilde{\kappa}^A,\tilde{\kappa}^B)\cup I_2^-(\tilde{\kappa}^A,\tilde{\kappa}^B)}\left[w_i\left(\tilde{\kappa}^B-\frac{\tilde{\kappa}^A}{q_i}\right)+\frac{p_i}{q_i}\right]+\sum_{i\in I_1^-(\tilde{\kappa}^A,\tilde{\kappa}^B)}\left[w_i\tilde{\kappa}^B\right]-(\varepsilon+\tilde{\delta})W. \qquad (6.20)
\end{aligned}$$

At a high-level, Inequality (6.19) supports the intuition that when p_i increases and/or q_i decreases, that is player A has more additive favoritism and/or multiplicative in the battlefields, she can guarantee a better payoff. Reversely, Inequality (6.20) indicates that player B can guarantee a better payoff when she has more favoritism (i.e., p_i decreases and/or q_i increases). However, these remarks remains only as intuitive statement since in principle, when p_i and q_i changes, the values of the solution $(\tilde{\kappa}^A,\tilde{\kappa}^B)$ found by Algorithm 8 also changes and we cannot explicitly keep track of these changes. More careful numerical experiments are needed to analyze the relation between the players' payoffs in playing $\mathrm{IU}^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy and the favoritism parameters p_i, q_i. We leave this for future works.

Finally, we discuss the generalizability of the results obtained in this section to the non-constant-sum generalized-rule CB game $\mathcal{GR}-\mathcal{CB}_n$. Unlike the constant-sum game $\mathcal{GR}-\mathcal{CB}_n^C$ considered above, in $\mathcal{GR}-\mathcal{CB}_n$, each battlefield i may have different values to each player (denoted by w_i^A and w_i^B). This adds complexity into the problem at hand, but in a high-level perspective, the main challenges encountered in characterizing the equilibria and approximate equilibria of the game $\mathcal{GR}-\mathcal{CB}_n$ is not different from what we encounter in the case of the game $\mathcal{GR}-\mathcal{CB}_n^C$. A trivial generalization is as follows: extend from Definition 6.2.7, we can obtain a set of distributions in the $\mathcal{GR}-\mathcal{CB}_n$ game such that following them is the best response of each player at each battlefield in $\mathcal{GR}-\mathcal{CB}_n$; specifically, for any $(\kappa^A,\kappa^B)\in\mathbb{R}_{>0}^2$, the pair of distributions corresponding to battlefield i matches the equilibrium of the F-APA with $p := p_i$,

$q := q_i$, $u^A := w_i^A \kappa^A$ and $w^B := w_i^B \kappa^B$. However, the conditions determining that a battlefield i of $\mathcal{GR}-\mathcal{CB}_n$ corresponds to which case in Theorem 6.2.5 or Theorem 6.2.6 (characterizing the equilibria of F-APA) is more complicated since we need to check all configurations of w_i^A and w_i^B. We leave more detailed analyses for future work.

. . . .

Summary: In this chapter, we studied the generalized Lottery Blotto game (LB game)—an extension of the CB game model—and showed that the IU$^{\gamma^*}$ strategies are also the approximate equilibria of this game. We characterized the approximation error of the IU$^{\gamma^*}$ strategies in two special instances—the LB games with the power-form and logit-form CSFs—and showed that these errors are negligible relative to the players' payoffs under a condition of the number of battlefields and the parameters of the involved CSFs.

Moreover, we presented several initial results in the constant-sum generalized-rule Colonel Blotto game ($\mathcal{GR}-\mathcal{CB}_n^C$) (to the best of our knowledge, our analysis is the first study conducted for this game). First, we characterized the exact equilibria (in all parameters' configurations) of the all-pay auction with favoritism and used these results as tools to study the $\mathcal{GR}-\mathcal{CB}_n^C$ game. We encountered challenges in computing the exact optimal univariate distributions of $\mathcal{GR}-\mathcal{CB}_n^C$; as an alternative, we proposed an efficient heuristic algorithm that provides a set of distributions approximating optimal univariate distributions of $\mathcal{GR}-\mathcal{CB}_n^C$. We used these distributions to construct approximate equilibria of the game $\mathcal{GR}-\mathcal{CB}_n^C$ (under an assumption on the existence of solutions of an equation).

Part II

ONLINE LEARNING IN RESOURCE ALLOCATION GAMES WITH COMBINATORIAL STRUCTURES

Chapter 7

Online Resource Allocation Games as Online Shortest Path Problems (OSPs)—Formulation and Related Works

Some of the ideas and results presented in this chapter have previously appeared in our following publications: Vu, Loiseau, and Silva (2019b) and Vu, Loiseau, Silva, and Tran-Thanh (2020).

In this second part of the manuscript (that includes Chapters 7, 8 and 9), we consider resource allocation games under an *online learning setting* in which a player plays and learns on-the-fly a sequence of games (without having complete information of the instantaneous game when making decisions). We aim to investigate a subclass of this model that covers a wide range of applications: the set of *resource allocation games with combinatorial structures*, i.e., cases where there exists a one-to-one mapping between the strategy set (at each stage) of a player and a subset of $\{s \in \mathbb{N}^n : \sum_{i\in[n]} s_i \leq X\}$ (for fixed $n, X \in \mathbb{N}$). Several game models introduced in the previous chapters satisfy such condition, including the discrete CB game and other discrete Blotto games (i.e., where players' allocations are constrained to be integers) as well as the multi-looking hide-allocation game and the hide-and-seek game (with discrete search).[1]

Among several possible approaches, we focus on the *regret-minimization analysis* of online resource allocation games with combinatorial structures. Our high-level perspective is to convert them into the online combinatorial optimization (OComb) framework (see Section 2.2.1 for a definition of OComb) and the leading question is how to exploit games' structures in order to improve existing learning policies in OComb when applying them to these games. As a case study, we investigate especially *the online discrete CB game* in which at each stage, a learner is given a budget k to play a discrete CB game against an adversary across n battlefields (see Section 7.1 for a formal definition). We consider this game under several feedback settings (received

[1] See Section 3.2.1 for a definition of the discrete CB game and see Section 2.1.3 for discussions on the multi-looking hide-allocation game and the hide-and-seek game.

by the learner at the end of each stage) that model different sets of applications. The online discrete CB game, among other online resource allocation games (with combinatorial structures), motivates us to study an important instance of OComb: the online shortest path problem (OSP)—see Section 2.2.3 for a definition. There is still room for improvement for state-of-the-art algorithms in OSP. We aim to first design new algorithms running in any generic instance of OSP (under several feedback settings) that improves the running time and regret guarantees in comparison with existing algorithms. Then, we apply these findings into several online resource allocation games, including the online discrete CB game and the online hide-and-seek game,[2] and show corresponding improvements.

The *outline of this part* is as follows: This chapter (Chapter 7) serves as an introduction. In Section 7.1, we first give definitions of several online resource allocation games; particularly, the online discrete CB game model is presented in Section 7.1.1, the online version of other Blotto games and the online hide-and-seek game are discussed in Section 7.1.3. An important result in this chapter is that we *show the connection between online resource allocation games (with combinatorial structures) and the framework of the online shortest path problem* (OSP). In particular, the conversion of the online discrete CB game into an OSP is presented in Section 7.1.2 and similar conversions for other games are presented in Section 7.1.3. Next, as a preparation for our studies in the following chapters, we also introduce the model of OSP with side-observations (henceforth, SOOSP)—an instance of OSP that has not been explicitly formulated in the literature. Finally, to pinpoint the challenges encountered in studying these games, we review the literature of OComb, OSP and other related problems in Section 7.3. In Chapter 8, motivated by the online discrete CB game with semi-bandit feedback, we study the SOOSP model and design a novel regret-minimization algorithm that provides improvements in performance guarantees and the running time. Then, we turn our interest back to online resource allocation games (being cast as SOOSP) and illustrate the benefits of this algorithm in playing the online semi-bandit CB game and the online hide-and-seek game. We also discussed the application of the proposed algorithm to the online discrete CB game with full-information feedback (see Chapter 8). Finally, in Chapter 9, to provide an improving learning policy for playing the online discrete CB game with bandit feedback, we study the OSP with bandit feedback model and provide some methods to improve existing algorithms in this model.

A note on the terminology: Some results in this chapter are extracted from our publications: Vu, Loiseau, and Silva (2019b) and Vu, Loiseau, Silva, and Tran-Thanh (2020). In these publications, the online shortest path problem is actually referred to as the path planning problem (the acronyms PPP and SOPPP are used instead of OSP and SOOSP). In the literature of OComb and MAB, these terms are often used interchangeably. However, it appears that the term "path planning" has also been used by other communities. In this thesis, lest the readers have unnecessary confusions, we make this change of the terminology and use the term online shortest path.

[2]The online hide-and-seek game is defined in Section 7.1.3.

7.1 Online Resource Allocation Games

In the game-theory literature, a variety of resource allocation games have been studied intensively in the one-shot complete-information form; in Part I of this thesis, we analyze a subclass—the Blotto games—also in this perspective. Although it has been showed to be applicable in many situations (e.g., see Chapter 1 for a list of applications of the CB game), the one-shot complete-information game model has its limits and fails to capture other type of applications in practice. In certain situations, it is more natural to consider a model in which players need to play a sequence of games and their objective is to maximize the cumulative payoffs. One of the common approaches used to analyze these situations is to model them as *online learning problems*—a well-established framework with many applicable results (see e.g., Cesa-Bianchi and Lugosi (2006) for more details about online learning and the relations with games). We call an *online resource allocation game* the model broadly defined as follows: a resource allocation game, between a learner and an adversary, is played repeatedly in T stages ($T \in \mathbb{N} \backslash \{0\}$); when making decisions at each stage, the learner may not be informed about certain information (e.g., in the CB game, the parameters such as the values of battlefields or the opponent's budget could be unknown); after the strategies are chosen and played, the learner receives (possibly restricted) feedback about the payoff and/or the outcomes of the game at that stage. In this setting, players are often required to sequentially learn the game on-the-fly and adjust the trade-off between exploiting known information and exploring to gain new information. Note also that besides the online learning setting, there exist other models used to study sequential plays in games such as sequential games, repeated games, stochastic games or Markov decision processes. We do not follow these approaches and only focus on the online learning setting described above.

As previously discussed, we limit our study in this part of the thesis to the subclass of online resource allocation games with combinatorial structures; one of them—the online discrete Colonel Blotto game—especially attracts our attention. As discussed in Chapter 1, this is one of the resource allocation games with the simplest rule and description, it also possesses typical characteristics of the whole class. We study the online discrete CB game intensively as an example of how one can exploit the structure of the game to improve existing learning policies. In this section, we first define formally this game in Section 7.1.1; then we develop an important result: any online discrete CB game can be converted into an online shortest path problem (this is done in Section 7.1.2); finally, we introduce the formulations of some other online resource allocation games in Section 7.1.3 and show similar conversions.

7.1.1 The Online CB game

We first consider two motivational examples as follows:

> In radio resource allocation (in a cognitive radio network), a solution obtaining the balance between efficiency and fairness is to provide the users with fictional

budgets (the same budget at each stage) and let them bid across n spectrum carriers simultaneously to compete for obtaining as many bandwidth portions as possible, the highest bidder to each carrier wins the corresponding bandwidth (see e.g., Chien et al. (2019)).[3] If this system only allows the bids to be integers (that allows a simple transmission with fewer units of information), from the point of view of each user, she is a learner who plays a sequence of discrete CB games on n battlefields (i.e., n spectrum carriers).

In advertising, we consider an online marketing campaign who distributes the broadcasting time (rounded up to integers) in n advertisement slots; once per day, its performance is reviewed and based on this information, a better strategy can be learned for the following days. If it is assumed that among two products promoted by two competing marketing campaign on the same n slots, the one with a longer broadcasting-time is more likely to be bought, then each marketing campaign can be considered as a learner who needs to play a sequence of discrete CB game on n battlefields (n advertisement slots).

Now, we present a formal definition of the online discrete CB game. Note that in this definition, certain elements are similar to the definition of the one-shot discrete CB game (Definition 3.2.1) presented in Chapter 3; for these elements, we reuse the corresponding notations.

Definition 7.1.1. ***The online discrete Colonel Blotto game*** *is a game between a learner and an adversary over $n \geq 1$ battlefields within a time horizon $T > 0$. At stage $t \in [T]$, each battlefield $i \in [n]$ has a value $\boldsymbol{b}_t(i) > 0$ (unknown to the learner) such that $\sum_{i=1}^{n} \boldsymbol{b}_t(i) = 1$. At stage t, the learner needs to distribute k troops ($k \geq 1$ is fixed) towards the battlefields while the adversary simultaneously allocates hers; that is, the learner chooses a vector $\boldsymbol{z}_t$ in the strategy set $S_{k,n} := \{\boldsymbol{z} \in \mathbb{N}^n : \sum_{i=1}^{n} \boldsymbol{z}(i) = k\}$. At stage t and at battlefield $i \in [n]$, if the adversary's allocation is strictly larger than the learner's allocation $\boldsymbol{z}_t(i)$, the learner loses this battlefield and she suffers the loss $\boldsymbol{b}_t(i)$; if they have tie allocations, she suffers the loss $\boldsymbol{b}_t(i)/2$; otherwise, she suffers no loss. The learner's loss at each stage is the aggregate of the losses from all the battlefields. At the end of stage t, the learner observes some (possibly restricted) feedback about her instantaneous payoff, the parameters and/or the outcomes of the game played in that stage. The objective of the learner is to minimize her expected regret.*

In the remainder of this thesis, in places where there is no ambiguity, we drop the term "discrete" and address the game defined in Definition 7.1.1 simply as *the online CB game*. Hereinafter, we also refer to the elements of $S_{k,n}$ in the online CB game as the *pure strategies* of the learner. On the other hans, in Definition 7.1.1, we note that the values of battlefields[4] are allowed to change over time and they are unknown to the learner when she makes decisions at each stage. This setting is generic and it allows a large scope

[3] Note that in this model, the network users play the roles of players; this is unlike the radio management system from Hajimirsaadeghi and Mandayam (2017) that we discussed in Chapter 1 (see also Figure 1.1) where they are modeled as the battlefields.

[4] The battlefields may have different values to the learner and the adversary.

of applications of the online CB game. For instance, in the motivational example of radio resource allocation described above, the actual data rate of the bandwidth (which corresponds to the values of battlefields) are often subject to unpredictably random noises; therefore, they may change over time. On the other hand, the term "adversary" here can indicates either a single opponent or a set of opponents of the learner (i.e., in each stage, we can consider an N-player CB game where $N \geq 2$); moreover, we do not impose any constraint on how the adversary generates her strategies, that is, the adversary involved in the online CB game might be a non-oblivious adversary. Note also that the adversary's budget can be changed over time (as long as it is not larger than $n \times k$) and they are also unknown to the learner.[5] Finally, another difference from the one-shot discrete CB game is that we include, in Definition 7.1.1, the fixed tie breaking rule sharing equally the value to the learner and the adversary. This is simply to avoid unnecessary complications in the model; the results on the online CB game that we present in the next chapters can easily be extended to the game with a general tie-breaking rule.

Importantly, in Definition 7.1.1, the online CB game is defined with a generic description of the feedback that the learner observes at the end of each stage. By changing this feedback, we obtain different instances of the game. In this thesis, there are three feedback settings that we study for the online CB game:

> The *full-information setting*: at the end of stage t, the learner observes the values of all battlefields (i.e., $b_t(i), i \in [n]$) and the allocations that the adversary chose in that stage.
>
> The *semi-bandit setting*: at the end of stage t, the learner only knows whether she wins or loses (or has a tie) and the loss she suffers from each battlefield (but *not* the adversary's allocations).
>
> The *bandit setting*: at the end of stage t, the learner only observes the aggregate loss she suffers from all the battlefields and nothing else.

Hereinafter, we refer to the game defined in Definition 7.1.1 where the learner receives the feedback according to the full-information setting (resp. the semi-bandit/ bandit settings) shortly as the *online full-information CB game* (resp. the *online semi-bandit CB game* and the *online bandit CB game*). The feedback settings described above are modeled from common types of information available to the learners in the motivational applications of the online CB game model. For example, the radio resource allocation problem described at the beginning of this section can be modeled by an online CB game with semi-bandit feedback if we allow each user to observe her own data rate (the gain/loss) achieved via each carrier (corresponding to battlefields' values) but not other users' bids. This is a realistic assumption since it is easy to measure the data rate while it requires transmissions with more information units to keep track of all

[5] Technically, the model of the online CB game may allow the adversary to have any arbitrary finite budget at each stage; however, the case when it is larger than $n \times k$ is trivial: the learner always loses at all battlefields and suffers the same loss regardless of what she does; that is, the regret is always zero.

Table 7.1: The feedback settings in the online CB game.

	Adversary's allocations	Battlefields' values	Battlefields' outcomes	Aggregate loss from battlefields
Full-information	✓	✓	✓	✓
Semi-bandit		✓	✓	✓
Bandit				✓

users' bids (that might cause delays in the system). If it is needed to further reduce the amount of information transmitted in the system for the purpose of this bidding protocol, we might let each user measures only the total data rate that she transmits rather than that in each carrier; in this case, we may model this system as an online CB game with bandit feedback. Another example of the bandit setting is resource allocation problems in advertising (described above) under an additional assumption that the marketing campaign (the learner) observes the total revenue of the day, i.e., the total gains (also, the total loss), without knowing exactly which advertisement slot (among which the ads is broadcasted) has encouraged the customer to buy the product. The available information that the learner observes in each feedback setting of the online CB game is summarized in Table 7.1.

One might observe that the terminology we use to address the feedback settings in the online CB game correspond to the three standard settings of the online combinatorial optimization (OComb) framework (see Definition 2.2.2 for a definition). This is due to the following simple observation: *the online CB game can be formulated as a standard* OComb *problem*. In particular, in the game where the learner, at each stage, needs to allocate k troops across n battlefields, each pure strategy of the learner can be mapped to a 0-1 vector of the form $\big(\boldsymbol{p}(1,0),\dots,\boldsymbol{p}(1,k),\dots,\boldsymbol{p}(n,0),\dots,\boldsymbol{p}(n,k)\big) \in \{0,1\}^{k\times n}$ where $\boldsymbol{p}(i,j) = 1$ if and only if in that pure strategy, j troops are allocated on the i-th battlefield. One can easily verify that the feedback settings described above in the online CB game match exactly with the definition of those settings in the corresponding OComb.

This connection between the online CB game and the OComb framework also indicates that the learner in the online CB game can apply no-regret algorithms from the OComb literature to obtain the corresponding guarantees on the expected regret. However, there are two drawbacks to this approach: (i) the standard algorithms in OComb provide no overall guarantee that they run in polynomial time in terms of the dimension of the action vectors; this implies that in the worst-case scenarios of the online CB game, their running time is exponential in terms of n (the number of battlefields) and k (the budget)—too *inefficient to be implemented in practice*; (ii) these algorithms—designed for generic setting of OComb—do not exploit the structure of the CB game, thus there is still room for improvement in the regret guarantees provided by these algorithms. We will emphasize these remarks again in Section 7.3 where we review the literature of OComb and other related problems. Before doing that, in the next section, we first introduce another presentation, somewhat more refined, of the online CB game: we view this game as an online shortest path problem.

7.1.2 The Online CB Game as an Online Shortest Path Problem (OSP)

The online shortest path problem (OSP) is a special instance of OCOMB in which each element of the action set of the learner corresponds to a path on a directed acyclic graph (DAG)—see Section 2.2.3 for a definition of OSP. Importantly, there exist variants of regret-minimization algorithms for OCOMB, especially in the class of EXP3-type algorithms,[6] that run more efficiently in OSP than in OCOMB thanks to possible exploitation of the graphical structure (more details are given in Section 7.3 where we give a literature review on OSP). At a high-level, the OSP model is more tractable than the OCOMB framework; therefore, we desire to see whether any arbitrary instance of the online CB game can be cast into an OSP. Our answer to this question is positive.

Given a CB game with the parameters k and n as presented in Definition 7.1.1, we create a DAG, denoted by $G := G_{k,n}$, defined as follows:

Definition 7.1.2 (CB Graph)**.** *The graph $G_{k,n}$ is a DAG that contains:*

(i) *$V := 2 + (k+1)(n-1)$ vertices arranged into $n+1$ layers. Layer 0 and Layer n, each contains only one vertex, respectively labeled $s := (0,0)$–the source vertex and $d := (n,k)$–the destination vertex. Each Layer $i \in [n-1]$ contains $k+1$ vertices whose labels are ordered from left to right by $(i,0),(i,1),\dots,(i,k)$.*

(ii) *There are directed edges from vertex $(0,0)$ to every vertex in Layer 1 and edges from every vertex in Layer $n-1$ to vertex (n,k). For $i \in \{1,2,\dots,n-2\}$, if $k \geq j_2 \geq j_1 \geq 0$, there exists an edge connecting vertex (i,j_1) in Layer i to vertex $(i+1,j_2)$ in Layer $(i+1)$.*

An example illustrating the graph $G_{k,n}$ of an instance of the CB game is given in Figure 7.1. From Definition 7.1.2, let us respectively denote by E and P the number of edges and paths (going from vertex s to vertex d) in the graph $G_{k,n}$; we have:

$$E = (k+1)\left[4 + (n-2)(k! + 2)\right]/2 = \Omega(nk^2), \qquad P = \binom{n+k-1}{n-1} = \Omega(2^{\min\{n-1,k\}}).$$

The edge connecting vertex (i,j_1) to vertex $(i+1,j_2)$ for any $i \in \{0,1,\dots,n-1\}$ represents allocating $(j_2 - j_1)$ troops to battlefield $i+1$. The length of every path from s to d is n. More importantly, each path from s to d in the graph G represents a pure strategy in $S_{k,n}$. This is formally stated in Proposition 7.1.3.

Proposition 7.1.3. *Given k and n, there is a one-to-one mapping between the strategy set $S_{k,n}$ of the learner in the online CB game (with k troops and n battlefields) and the set of all paths from vertex s to vertex d of the graph $G_{k,n}$.*

The proof of this proposition is trivial and can be intuitively seen in Figure 7.1. We note that a similar graph is studied by Behnezhad, Dehghani, et al. (2017) (for the one-shot complete information discrete CB game); however, it is used for a completely different

[6]See our discussion on this class of algorithms in Section 2.2.2.

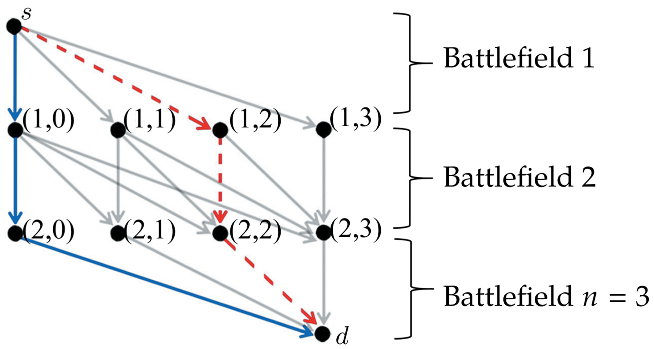


Figure 7.1: The graph $G_{3,3}$ corresponding to the CB game with $k=n=3$. For example, the bold-blue path represents the strategy $(0,0,3)$ while the dash-red path represents the strategy $(2,0,1)$. Note that the different colors and the dash lines here are just for the purpose of illustrating these examples.

purpose and it also contains more edges and paths than $G_{k,n}$ (that are not useful for our purpose in this chapter).

From Proposition 7.1.3, we conclude that *each instance of the online CB game model is equivalent to an* OSP where at each stage the learner chooses a path in $G_{k,n}$ and the loss on each edge is generated from the allocations of the learner (corresponding to that edge) and the adversary (in the corresponding battlefield) according to the rules of the game. Moreover, we can see that the definitions of the three feedback settings in the online CB game (described in the previous section) match exactly with the definitions of OSP with full-information (i.e. observing the losses of all edges), semi-bandit (i.e., only observing the losses of the edges belonging to the chosen path) and bandit feedback (i.e., only observing the aggregate of the edges' losses belonging to the chosen path).[7]

Importantly, similar conversions can be conducted to cast other online resource allocation games with combinatorial structures into OSP by constructing DAGs corresponding to the strategy's set of the learner in these games (an example is given in Section 7.1.3). For this reason, the high-level perspective of our studies in Chapters 8 and 8 is as follows: we consider OSP on generic graphs under different feedback settings and aim to improve existing algorithms in these cases; after that, we analyze the application of our findings into the particular OSP corresponding to the online CB game (with the graph $G_{k,n}$) and show how the learner benefits from that. In particular, Chapter 9 investigates the algorithms in OSP (and the online CB game) with bandit feedback. On the other hand, the case of online semi-bandit CB games motivates us to introduce and study another feedback setting of OSP, called the OSP with side-observations (SOOSP), that has not been explicitly analyzed in the literature.[8] We give the formulation of SOOSP in Section 7.2; then discuss the conversions of several online

[7]Formal definitions of these feedback settings in OSP are presented in Section 2.2.3.

[8]Indeed, the structure and rule of the online CB game allows the learner to deduce extra information from the semi-bandit feedback (see present this in Section 8.1).

resource allocation games into SOOSP and design new no-regret algorithms for it in Chapter 8.

7.1.3 Other Online Resource Allocation Games and the Relation to OSPs

In this section, we discuss several online resource allocation games with combinatorial structure other than the online CB game. We start with an interesting observation as follows: in making the conversions of the online CB game into the OComb framework (see Section 7.1.1) and into the OSP framework (see Section 7.1.2), only the strategy set of the learner is involved and the Blotto-rule determining the winner in each battlefield plays no particular role. Therefore, these conversions can be trivially generalized to other Blotto games where the players have the same strategy sets as in the CB game. For example, one can easily extend Definition 7.1.1 to define the online version of the discrete Lottery Blotto game[9] and also cast it into an OSP by the same conversion. Now, we introduce another game that has different motivation and formulation from that of the CB game.

The Online Hide-and-Seek Game

A one-shot version of the hide-and-seek game has been introduced in Section 2.1. Briefly put, in a hide-and-seek game (with discrete search), a seeker chooses n among k locations ($n \leq k$) to search for a hider, who chooses the probability of hiding in each location; the seeker's payoff is the summation of the probability that the hider hides in the chosen locations and the hider's payoff is the probability that she successfully escapes the seeker's pursuit. Several variants of this game have been used to model surveillance situations (see e.g., Bhattacharya et al. (2014)), anti-jamming problems (see e.g., Navda et al. (2007) and Wang and M. Liu (2016)) and vehicles control problems (see e.g.,Vidal et al. (2002)).

As in the case of the CB game, there exist applications of the hide-and-seek game that requires players to play repeatedly the game and learn on-the-fly to improve the payoffs (or reduce the losses). A motivational example is the following spectrum sensing problem in the opportunistic spectrum access context (see e.g., Yucek and Arslan (2009)):

> The network users are classified into two groups: primary users who are prioritized to use the spectrum and secondary users who can only use the remaining bandwidth after the usage of the primary users (if available). The secondary users' usage of the spectrum must not cause interference to the connection of the primary users. To do this, at each stage, each secondary user—as a learner—needs to choose to send the sensing signal to at most n among k channels (due to energy constraints, she cannot sense all channels, thus $n \leq k$) with the objective of sensing the channels with the (total) availability as high as possible. Note

[9] See Definition 3.2.3 for a definition of a one-shot Lottery Blotto game.

that the channels' availability depend on primary users' decisions that might be non-stochastic.

Formally, *the online Hide-and-Seek game (the online HS game)* is a repeated game (within the time horizon $T > 0$) between a hider and a seeker. Here, we consider that the learner plays the role of the seeker and the hider is the adversary. There are k locations, indexed from 1 to k. At stage t, the learner sequentially chooses n locations ($1 \leq n \leq k$), called an *n-search*, to seek for the hider, that is, she chooses $z_t \in [k]^n$ (if $z_t(i) = j$, we say that location j is her i-th move). The hider maliciously assigns losses on all k locations.[10] The learner's loss at stage t is the sum of the losses from her chosen locations in the n-search at stage t, that is $\sum_{i\in[n],j\in[k]} \mathbb{I}_{\{z_t(i)=j\}} \boldsymbol{b}_t(j)$. At the end of stage t, the learner observes some feedback about her losses from the locations. The learner's objective is to minimize the expected regret over T.

In the application of spectrum sensing mentioned above, the secondary user, as the learner, is a seeker; k spectrum channels correspond to k locations, each is embedded with a loss that can be interpreted as the unreliability of the channel (these losses can be normalized to represent the hiding probability of an artificial hider).

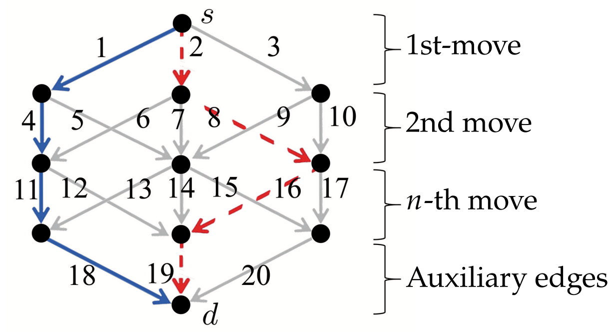


Figure 7.2: The graph $G_{3,3,1}$ corresponding to the HS game with $k = n = 3$ and $k_0 = 1$. E.g., the blue-bold path represents the $(1, 1, 1)$ search and the red-dashed path represents the $(2, 3, 2)$ search. Note that the different colors and the dash lines here are just for the purpose of illustrating these examples.

As in the case of the CB game, tackling the online HS game as a standard OComb is computationally involved. As such, we follow the OSP formulation instead. First, we note that there are a variety of instances of the online HS game that we can consider since the n-search of the learner often needs to satisfy some constraints in practice. In this thesis, as an example, we use the following constraint: for a fixed $k_0 \in [0, k-1]$, $|z_t(i) - z_t(i+1)| \leq k_0, \forall i \in [n-1]$ (called the *k_0-coherence constraint*), i.e., the seeker cannot search too far away from her previously chosen location.[11] To cast an HS game

[10] For example, these losses can be the wasted time supervising a mismatch location or the probability that the hider does not hide there.

[11] Our results can be applied to HS games with other constraints, for example, $z_t(i) \leq z_t(i+1), \forall i \in [n]$, i.e., she can only search forward; or, $\sum_{i\in[n]} \mathbb{I}_{\{z_t(i)=k^*\}} \leq k_0$, i.e., she cannot search a location $k^* \in [k]$ more than k_0 times.

where the learner can do n-searches in k locations under a k_0-coherent constraint to an OSP, we create a DAG $G := G_{k,n,k_0}$ whose paths set has a one-to-one correspondence to the set containing all feasible searches of the learner. Figure 7.2 illustrates the corresponding graph of an instance of the HS game and we give a formal definition of G_{k,n,k_0} in Appendix E.6. This variant of the online HS game is equivalent to the OSP where the learner chooses a path in G_{k,n,k_0} and edges' losses are generated by the adversary (i.e., the hider's probability of hiding in the corresponding location) at each stage. Note also that to ensure all paths end at d, there are n auxiliary edges in G_{k,n,k_0} that are always embedded with 0 losses. In G_{k,n,k_0}, there are $E = \mathcal{O}(k^2 n)$ edges and $P = \Omega(k_0^{n-1})$ paths. .

Finally, note that we do not claim that any online resource allocation game (with combinatorial structures) can be cast to OComb or OSP. In fact, the conversions used for the cases of the online CB game and the online HS game depend on the fact that there is linearity in the players' payoff functions in these games. Therefore, it might exists resource allocation game that is impossible to be cast into OComb nor OSP.

7.2 The Online Shortest Path Problem with Side-Observations (SOOSP)

In our review on the OSP in Section 2.2.3, we have introduced three standard feedback settings in OSP: full-information, semi-bandit and bandit. From the connection that we established between online resource allocation games and OSP in the previous sections, we observe that there exist situations where the feedback in these games are not fully captured when converting them into OSP with the standard settings. This is often due to the fact that structures of the games may allow extra information to be deduced from the feedback (see Chapter 8 for an example—the case of the online semi-bandit CB game). Therefore, in this section, we present the definition of OSP under another feedback settings, called OSP with side-observations (denoted by SOOSP).[12] To the best of our knowledge, SOOSP has not been explicitly defined in the literature; although it can be considered as a special case of the online combinatorial optimization with side-observations framework.[13]

As in the generic formulation of OSP (introduced in Section 2.2.3), SOOSP is also defined on a directed acyclic graph (DAG), denoted by G. We also reuse several notations with respect to G including its vertices set $\mathcal{V}$, its edges set $\mathcal{E}$ (and $V = |\mathcal{V}|, E = |\mathcal{E}|$), the source vertex s, the destination vertex d, the set $\mathcal{P}$ containing all *paths* starting from s and ending at d (and $P := |\mathcal{P}|$); moreover, let n be the length of the longest paths in $\mathcal{P}$ and $\boldsymbol{\ell}_t(e)$ be the loss generated by the adversary on edge e at stage t. The formal definition of SOOSP, based on the OSP framework, is as follows:

[12] In Chapter 8, we show how the SOOSP model captures the feedback and side-observations in the online CB game and the online HS game.

[13] See the definition of SOComb in Section 2.2.1.

Definition 7.2.1. *Given a time horizon $T \in \mathbb{N}\backslash\{0\}$ and a* DAG *G, an* ***online shortest path problem with side-observations*** *(*SOOSP*) on G is an* OSP *(on G) such that the learner's feedback at the end of stage $t \in [T]$ after choosing the path $\tilde{\boldsymbol{p}}_t \in \mathcal{P}$ is presented as follows:*[14] *First, she receives semi-bandit feedback, i.e., she observes the edges' losses $\boldsymbol{\ell}_t(e)$, for any e belonging to the chosen path $\tilde{\boldsymbol{p}}_t$. Additionally, each edge $e \in \tilde{\boldsymbol{p}}_t$ may reveal the losses on several other edges. To represent these side-observations at time t, we consider a graph, denoted G^O_t, containing E vertices: each vertex v_e of G^O_t corresponds to an edge $e \in \mathcal{E}$ of the graph G. There exists a directed edge from a vertex v_e to a vertex $v_{e'}$ in G^O_t if, by observing the edge loss $\boldsymbol{\ell}_t(e)$, the learner can also deduce the edge loss $\boldsymbol{\ell}_t(e')$; we also denote this by $e \longrightarrow e'$ and say that the edge e reveals the edge e'. The objective of the learner is to minimize the cumulative expected regret.*

Hereinafter, we use the term *observation graphs* to refer to G^O_t. In general, these observation graphs can depend on the decisions of both the learner and the adversary. On the other hand, all vertices in G^O_t always have self-loops, i.e., every edge in the graph G reveals itself when get chosen (this comes from semi-bandit feedback). Naturally, the feedback setting defined in SOOSP interpolates between the semi-bandit and full-information of OSP. Indeed, in the case where none among G^O_t ($t \in [T]$) contains any other edge than the self-loops, no side-observation is allowed and the problem is reduced to the classical semi-bandit setting. If all $G^O_t (t \in [T])$ are complete graphs, SOOSP corresponds to the full-information OSPs. On the other hand, there are two situations regarding the observation graphs: in the *informed setting*, the learner observes G^t_O before she makes the decisions; in the *uninformed* setting, the learner observes G^O_t only after making the decisions at time t. In this thesis, we work with the uninformed setting of SOOSP since it is more general and some cases of interest in studying online resource allocation games belong to this setting.

7.3 Literature Review on Regret-Minimization Analysis in Bandit Problems and Beyond

In the previous sections, we have shown that online resource allocation games (that are our primary objects of study in this part of the thesis) have connections to a variety of online learning frameworks; our aim is to improve existing regret-minimization algorithms in these frameworks and apply our findings to have more effective and efficient learning policies in these games. In order to gain an understanding of the existing research and challenges we encounter in studying these problems, in this section, we give a brief literature review on regret-minimization analyses in OLO, OComb, OSP and MAB. Moreover, we also review the problem of learning equilibria in games—another connection between online learning and game theory. This section involves a quite intensive use of acronyms; for ease of reading, we summarize them in Table 7.2. Throughout this section, we use the notation T for the time horizon, S for the

[14]Recall that we use the letter $\boldsymbol{p}$ to denote an arbitrary path in $\mathcal{P}$ and use the *tilde notation* with a subscript (i.e., $\tilde{\boldsymbol{p}}_t$) to emphasize that this is *the path chosen by the learner at stage t*.

action set of the learner, D to be the dimension of the action and $n = \max_{p \in S}\{\|p\|_1\}$. The asymptotic notations $O, \tilde{O}$ and Ω are used with respect to $T, n, D \to \infty$. We first start with a review on the most basic model—the well-known MAB problem.

Table 7.2: Acronyms of several online learning models.

Acronyms	Models
MAB	Multi-armed bandit (with finitely many arms)
OLO	Online learning optimization
OCOMB	Online combinatorial optimization
SOCOMB	Online combinatorial optimization with side-observations
OSP	Online shortest path problem
OSPBAND	Online shortest path problem with bandit feedback
SOOSP	Online shortest path problem with side-observations

7.3.1 The MAB with Finitely Many Arms

The multi-armed bandit problem (MAB) is one of the most basic models in sequential learning (also sequential prediction) with limited feedback. The name "bandit" is inspired by situations where a player, in a casino, faces a number of "one-armed bandit" machines (a slang word to call casino's slot machines) and she must repeatedly choose one among these machines to pull, then observes and gains a reward (or suffers a loss). There are two basic sub-classes of MAB: *stochastic bandits* and *adversarial bandits*.[15] At a high-level, the adversarial bandit model is more general and contains the class of stochastic bandits; however, in the literature, solving these classes involves different sets of techniques. A definition of adversarial bandits has been introduced in Section 2.2.2.

In a stochastic bandit with D arms, at each stage, the loss on each arm i is independently drawn from an unknown probability distribution, called μ_i; without knowing this, a learner chooses one among D arms then observes and suffers the corresponding loss. This model has a long-standing history; it can be traced back to Thompson (1933) with applications in clinic trials. Modern analyses of stochastic bandits are considered to begin with the work of Lai and Robbins (1985) in which the upper confidence bounds (UCB) technique is first introduced.[16] In a stochastic bandit, the performance of a learning policy is often measured by the (pseudo) regret, defined as $R_T = \sum_{i \in [D]} \Delta_i \mathbb{E} C_i(T)$; where $C_i(s)$ is the number of times arm i is selected within the first s stages and we

[15]Markovian bandits are sometimes considered to be another basic subclass of MAB; however, since the sets of techniques and results involved in Markovian bandits are very different from the ones related to our results in this thesis, we do not review it in detail here. We refer the interested readers to see e.g., Gittins et al. (2011) and Mahajan and Teneketzis (2008) for surveys on Markovian bandits.

[16]Briefly put, UCB is based on the principle of optimism in the face of uncertainty and the use of concentration inequalities.

define $\Delta_i = m_i - m^*$ where m_i is the mean of μ_i and $m^* := \min_{j \in [D]}\{m_j\}$.[17] In the case where the losses are bounded in $[0, 1]$, the α-UCB algorithm—proposed by Auer, Cesa-Bianchi, and Fischer (2002), is essentially optimal (i.e., it provides a pseudo regret's upper-bound matching in order with a known lower-bound). It guarantees that $R_T \leq \sum_{i:\Delta_i>0} \frac{2\alpha}{\Delta_i} \log T + \frac{\alpha}{\alpha-2}$; i.e, the regret is logarithmic in terms of T ($\alpha > 2$). The constants involved in this bound are improved by Bubeck, Cesa-Bianchi, et al. (2012) and Garivier and Cappé (2011). Based on the UCB technique, several algorithms are proposed with more refined results (i.e., the gap between the upper and lower bounds is further reduced) including the UCB-V algorithm by Audibert, Munos, et al. (2009) and the KL-UCB algorithm by Garivier and Cappé (2011) and Maillard et al. (2011). On the other hand, in the worst-case analysis (i.e., with the worst distributions of arms' losses), the regret of α-UCB is sub-optimal (it is in the order of $\sqrt{DT \log T}$). To improve this, Audibert and Bubeck (2010) modify UCB to design an algorithm, called MOSS, that obtains the bound $\sqrt{DT}$. For the cases where $\mu_i, i \in [D]$ all belong to a particular class of distributions, refined results can be obtained; e.g, Kaufmann et al. (2012) re-discover that Thompson sampling is optimal for Bernoulli bandits, Kaufmann et al. (2018), Korda et al. (2013), and Menard and Garivier (2017) prove the asymptotic optimality for a variety of families of exponential distributions. Note also that the algorithms mentioned above work with an assumption that the time horizon T is known in advance; a well-known technique to obtain an anytime algorithm is the "Doubling Trick" that is often integrated into the other algorithms (see Besson and Kaufmann (2018) for an overview on this technique). High-probability guarantees of UCB-type algorithms are also studied by Audibert, Munos, et al. (2009) and Bubeck, Cesa-Bianchi, et al. (2012). Finally, results on lower-bounds of minimax regret in special classes of losses distributions can be found in several works; e.g., Auer, Cesa-Bianchi, Freund, et al. (1995) give results in the cases of Bernoulli bandits, Gerchinovitz and Lattimore (2016) provide results in Gaussian bandits with unit variance.

Note that in this thesis, our objects of study are online resource allocation games in which, losses of each arm (i.e., of each pure strategy) may not be generated from i.i.d. distributions over time; therefore, they cannot be modeled by stochastic bandits.

Unlike stochastic bandits, in adversarial bandits, no stochastic assumption is made on how losses are generated (besides the requirement that the losses are bounded); hence, they are also called the non-stochastic bandits. Adversarial bandits are also often formulated as a sequence of games, between a learner and an adversary, that involves the basic trade-off between exploration and exploitation in making decisions. Note that in early works, adversarial bandits are developed and studied independently from the stochastic bandit model; pioneer works that present adversarial bandits by game-theoretic formulations are Banos et al. (1968) and Hannan (1957). One of the

[17] In the case of stochastic bandits, this definition of pseudo regret is equivalent to the definition of the expected regret given in (2.5), i.e., the difference between the expected cumulative loss and the expected loss from playing repeatedly the best arm.

most common performance measure in adversarial bandits is the expected regret (see Section 2.2.1 for a definition). An upper-bound of order $O(\sqrt{TD\ln(D)})$ of the expected regret in adversarial bandits is guaranteed by the Exp3 algorithm, proposed by Auer, Cesa-Bianchi, Freund, et al. (2002). Exp3 is considered as a standard algorithm in adversarial bandits and we have introduced this algorithm in detail in Section 2.2.2. Note that Exp3 is based on a strategy called exponential weight algorithm that has a longer history (see Littlestone and Warmuth (1989)). In the case with oblivious adversaries,[18] this bound is improved to $O(\sqrt{TD})$ by the class of INF algorithms proposed by Audibert and Bubeck (2009) (which can also be viewed as an Mirror Descent algorithm, see e.g., Audibert, Bubeck, and Lugosi (2011)). On the other hand, algorithms designed for adversarial bandits do not work well in stochastic bandits and vice versa. For example, it is proved that UCB can have a regret that is linear in T when being applied to adversarial bandits and that Exp3 does not guarantee a regret as good as that of UCB in several cases of stochastic bandits. To solve this problem, an algorithm that can guarantee the "best of both worlds" regret bounds is proposed by Bubeck and Slivkins (2012). Several works also consider a more general notion of regret with respect to best-switching strategy (instead of the best fixed action in hindsight). For example, Audibert and Bubeck (2010) and Auer (2002) show that Exp3-type algorithms obtain the bound of order $\sqrt{TD\tilde{t}\log(TD/\tilde{t})}$ if one compares the cumulative loss with that of the best strategy switching at most $\tilde{t} \leq T$ times. Concerning high-probability bounds, a modified version of Exp3, called Exp3.P, is also proposed by Auer, Cesa-Bianchi, Freund, et al. (2002), it provides a regret guarantee matching that of Exp3 (see also Abernethy and Rakhlin (2009)). This bound is further improved by Neu (2015) such that $R_t \leq O\left(\sqrt{(\log D + \log(1/\delta))TD}\right)$ happens with a probability at least $1-\delta, \delta \in [0,1]$. Finally, as for the lower-bound in adversarial bandits, minimax lower-bound in order of $\Omega(\sqrt{TD})$ is given by Auer, Cesa-Bianchi, Freund, et al. (2002).

Online resource allocation games with combinatorial structures (where the learner's payoff at each stage can be deduced from feedback) can be modeled directly as adversarial bandits such that each pure strategy corresponds to an arm. However, by doing this, the feedback and information about the games are not fully captured in several situations. Moreover, it is often the case that the number of strategies of the learner is exponential in terms of games' parameters (due to combinatorial structures), e.g., in the online CB game and the online HS game. Therefore, the best regret guarantee and the best running time (that are linear in terms of the number of arms) provided by algorithms in adversarial bandits are still exponentially large (in terms of games' parameters) and are not useful in practice. For these reasons, we do not use adversarial bandits to model these games in this thesis.

The literature of bandit problems goes much further than these two basic subclasses. Several worth-mentioning classes and extensions are: continuous bandits—where the learner has infinitely many (stochastic) "arms" (see e.g., Bubeck, Munos,

[18] In these cases, the expected regret is also equivalent to the definition of the pseudo-regret.

et al. (2011) and Kleinberg (2005)), Lipschitz bandits—where the expectation of the loss is determined by a Lipschitz function (see e.g., Abernethy, Hazan, et al. (2008)), sparse bandits—where most of the arms have means of rewards equal 0 (see e.g., Gerchinovitz (2013)), contextual bandits—where the learner has access to additional information at the beginning of each round (see e.g., Kakade et al. (2008) and Rakhlin and Sridharan (2016)) and bandits with knapsacks (see e.g., S. Agrawal and Devanur (2014), Badanidiyuru et al. (2013), and Tran-Thanh et al. (2012)). Note that MAB can also be considered as a specific case of a broader class, called partial monitoring (see e.g., Lugosi et al. (2008) and Perchet (2011)). In the next section, we review some generalizations of MAB, the OComb and OLO models, that are the main framework of our results in this part of the thesis.

7.3.2 Regret-Minimization in OLO, OComb and OSP

The **online linear optimization** (OLO) model is an important generalization of adversarial bandits with the main difference is that the set of arms is replaced with an action set $S \subset \mathbb{R}^D$ and the loss that the learner suffers at each stage is a linear function of her chosen action and an adversarially chosen loss vector. A definition of OLO and its connection to MAB[19] have been given in Section 2.2.1. In OLO *with full-information feedback*, the Hedge algorithm, proposed by Freund and Schapire (1997), has been proved to be optimal: it provides a regret upper-bound in $\mathcal{O}(\sqrt{T \log |S|})$ matching a lower-bound; this setting is also considered quite intensively by Koolen et al. (2010). On the other hand, OLO *with bandit feedback* provides more challenges. Dani et al. (2008) proposes an algorithm, called Geometric Hedge, guaranteeing the expected regret to be at most $\mathcal{O}(D^{3/2}\sqrt{T})$—this is the first algorithm having a regret bound in OLO with bandit feedback that is sub-linear in T. Geometric Hedge is based on an Exp2 strategy mixing with an exploration distribution that is uniform over a barycentric spanner of S. This bound is improved (for cases with an oblivious adversary) to the order of $\mathcal{O}(\sqrt{TD \log(|S|)})$ by the Exp2 with John's exploration algorithm, proposed by Bubeck, Cesa-Bianchi, and Kakade (2012). Note that the computation of a barycentric and a John's ellipsoid of S remains non-trivial and inefficient in general.[20] Another line of works in OLO with bandit feedback is the family of Online Mirror Descent (OMD) algorithms; that is based on the Mirror Descent (MD) method in (offline) convex optimization (introduced by Nemirovski and Yudin (1983)). The connection between MD and online learning is first mentioned in Cesa-Bianchi and Lugosi (2006) and the first MD algorithm in OLO with bandit feedback is from Abernethy, Hazan, et al. (2008). An improved variant for the case where the actions set S is a compact and convex set, called the OSMD for the Euclidean ball, is proposed by Bubeck, Cesa-Bianchi, and Kakade (2012); it obtains the expected regret upper-bound in the order of $\mathcal{O}(\sqrt{DT \log T})$. Mirror descent is also applied to prove high probability bounds of OLO with bandit feedback by Abernethy and Rakhlin (2009). Note that OSMD can also be considered to be equivalent to the

[19]That is each OLO can be considered as an MAB where each action vector in S of the OLO corresponds to an arm of the MAB; and reversely, each MAB can be modeled as an OComb that is an instance of OLO.

[20]The computation of a John's ellipsoid of a set is an NP-hard problem.

class of Follow-the-Regularized-Leader (FTRL) algorithms (see Rakhlin, Abernethy, et al. (2009)); that, in turns, contains the class of Follow-the-Perturbed-Leader (FTPL) algorithms introduced by Kalai and Vempala (2005) (by adding a perturbation of the losses as an implicit form of regularization). There are two main issues in applying MD-type (and FTRL-type) algorithms to OLO: (i) their performances depend (heavily) on the chosen potential function; to our knowledge, currently, there is no universal choice for the potential function and given an arbitrary OLO instance, it is non-trivial which potential should be chosen or how to find it; (ii) they require an efficient oracle solving a convex optimization problem at each stage; this is problem-dependent and in general, this step might be computationally expensive.

It still remains an open question to design an algorithm that provides optimal (in order) guarantees on the expected regret while runs efficiently in a general instance of OLO *with bandit feedback. In this thesis, we focus on a subclass of* OLO, *called* OComb, *that allows more precise analyses when modeling online resources allocation games with combinatorial structures.*

An important case of OLO is **online combinatorial optimization** (OComb) in which the action set of the leaner is $S \subset \{0,1\}^D$. A definition of OComb, with the three standard feedback settings (full-information, semi-bandit and bandit), was presented in Section 2.2.1. Because of the combinatorial structure of the learner's action set, $|S| = \Omega(\exp(D))$. In the setting of OComb, it is desired to have algorithms not only with good regret guarantees but also with implementability (we call an algorithm to be efficient if it runs in polynomial time in terms of D).

The *full-information feedback setting of* OComb is first formalized by Kalai and Vempala (2005); it is studied quite intensively by Koolen et al. (2010) who propose several extended variants of the Hedge algorithm guaranteeing the regret in the order of $\mathcal{O}(\sqrt{Tn\log(D/n)})$ where $n = \max_{p\in S}\{\|p\|_1\}$. Helmbold and Warmuth (2009) proposes an OSMD-type algorithm in this setting with a similar regret bound. Several other works consider OComb under full-information with specific action sets of the learner are Takimoto and Warmuth (2003), Warmuth, Koolen, et al. (2011), and Warmuth and Kuzmin (2008).

The model of OComb *with bandit feedback* (also called combinatorial bandits) is formulated by Cesa-Bianchi and Lugosi (2012); an algorithm called ComBand is proposed by these authors, it guarantees an expected regret of at most $2\sqrt{Tn^2\log|S|/\lambda^* + TD\log|S|}$ where λ^* is the smallest eigenvalue of the co-occurrence of the exploration distribution (see Chapter 9 for more details on these terms). Importantly, it is shown by Cesa-Bianchi and Lugosi (2012) that in many instances[21] of OComb with bandit feedback (with specific action sets of the learner), by choosing the uniform distribution on S as the exploration distribution (having $\lambda^* = \Omega(n^2/D)$), ComBand guarantees an expected regret of at most $\mathcal{O}(\sqrt{TD\log|S|})$. On the other hand, in general, ComBand runs inefficiently in $\mathcal{O}(|S|T) = \mathcal{O}(\exp(D)T)$ time. A modified version of ComBand is

[21] Note importantly that this *excludes* the case of the online shortest path problem (OSP).

proposed by Combes et al. (2015), called the CombEXP algorithm. By mixing with ideas of the OSMD algorithm, CombEXP improves the complexity of ComBand in several cases[22] while maintaining the regret guarantees. Note that CombEXP still uses the uniform distribution as an exploration distribution that is sub-optimal in several instances of OComb.

AS for the OComb *with semi-bandit feedback*, one of the earliest works is from György et al. (2007) who consider the online shortest path problem; other instances of OComb such as m-sets and ranking selection are also considered by Uchiya et al. (2010) and Kale et al. (2010) with semi-bandit feedback. The state-of-the-art algorithm for OComb with semi-bandit is a variant of the OSMD algorithm, proposed by Audibert, Bubeck, and Lugosi (2014), that guarantees the expected regret to be at most $2\sqrt{nDT\log(D/n)}$. This upper-bound matches the order of a lower-bound (see e.g., Bubeck, Cesa-Bianchi, et al. (2012)). Note that OSMD runs efficiently only if the action set satisfies a special assumption and there exists an efficient optimization oracle (see Audibert, Bubeck, and Lugosi (2014) for more details); moreover, its performance depends on choices of potential functions. As discussed above, it still lacks a characterization for optimal choices of these inputs of OSMD, given an arbitrary instance of OComb. On the other hand, Neu and Bartók (2013) propose an FTPL-type algorithm, with losses estimators based on the Geometric Resampling technique, yielding a regret's upper-bound of order $O\left(n\sqrt{DT\log D}\right)$ and it provides guarantees on the running time (that is polynomial in terms of n) but only in expectation or either with high-probability.

> *For a general instance of* OComb *(under any feedback setting), it remains an open question to design an algorithm that guarantees an efficient running time (i.e., polynomial in terms of D and n) in worst-case scenarios while providing a good regret's upper-bound. Especially for the case of bandit feedback, there is still room for improvement in the regret guarantees of existing algorithms.*

Finally, we review the literature in one of the most important instances of the OComb framework: the **online shortest path problem** (OSP) (also called the path planning problem in the bandit literature)[23], i.e., where the learner's action set is the set of paths of a directed acyclic graph. OSP has a large range of applications, including the classical routing problem where it is needed to sequentially choose paths on a network to send packets (György et al. (2007)). Another application of OSP is the online recommendation systems where the strategy set of the learner (i.e., combinations/lists of recommended products) can be cast to a graph (see e.g., Kocák et al. (2014)). Definitions of OSP under full-information, semi-bandit and bandit

[22]However, OSP with bandit feedback (OSPBand)—which is our main focus in Chapter 9—is not explicitly considered in Combes et al. (2015) and it remains an open question whether any arbitrary instance of the OSPBand satisfies the condition such that CombEXP can be efficiently implemented (i.e., the convex hull of the action set can be represented by a polynomial number of linear equations and linear inequalities).

[23]Note that the term path planning problem is also used by other communities (e.g., in robotic motion planning) to refer to other classes of problems.

feedback settings are presented in Section 2.2.3. Additionally, we have also introduced in Section 7.2 another feedback setting called the OSP with side-observations. The OSP with full-information is studied by György et al. (2007) and Takimoto and Warmuth (2003) and Koolen et al. (2010). The semi-bandit setting is also considered by György et al. (2007).[24] On the other hand, OSP with bandit feedback (OSPBand) is discussed by Cesa-Bianchi and Lugosi (2012) (where it is addressed as path planning). The literature review of OSP with side-observations will be presented in Section 7.3.3. In the literature, it is common that OSP is not studied separately but often as an example of OComb and the most well-used learning policies in OSP either come from direct applications of algorithms in OComb or from their modified variants. Therefore, *the state-of-the-art regrets guarantees in* OSP *are the ones adopted from* OComb. However, unlike OComb, the graphical structure of OSP allows more efficient implementations of some algorithms. A useful technique is weight pushing, employed by György et al. (2007) and Takimoto and Warmuth (2003), that can be used to efficiently sample paths in Exp3-type algorithms (see our discussion in Section 2.2.3 for more details). Sakaue et al. (2018) also uses an extension of weight-pushing to efficiently compute estimated losses in a variant of ComBand algorithm for OSP with bandit feedback. These applications of weight-pushing help improve the running time in several Exp3-type algorithms; however, to guarantee that a variant of Exp3 can be efficiently implemented, it is needed to check whether the complexity of other steps can also be improved.

In OSP *with bandit feedback, although there exist* Exp3*-type algorithms that run efficiently (i.e., polynomial in terms of* D *and* n*) thanks to the weight pushing technique, their regret's guarantees can still be further improved. An open question (posed by Cesa-Bianchi and Lugosi (2012)) is to efficiently find an optimal exploration distribution of the* ComBand *algorithm when applying it to* OSP*. When designing new* Exp3*-type algorithms for* OSP*, implementability needs to be re-verified since it is not always trivial how the weight pushing technique can be integrated. Existing algorithms for* OSP *with semi-bandit feedback are the ones adopted from* OComb*, thus they encounter the same issues in implementation as in the literature of* OComb *(with semi-bandit feedback).*

7.3.3 Side-observations Feedback in MAB, OComb and OSP

In practical applications of MAB, OComb and OSP, there are situations where information in the feedback received by the learner are not captured completely by the standard feedback settings presented above. Therefore, other feedback models interpolating these settings are needed.

MAB with side-observations is first formulated by Mannor and Shamir (2011); the main difference of this model with the classical MAB is that at the end of each stage, besides the loss of the chosen action, losses of several other non-chosen actions are also revealed to the learner. The concept of *observation graphs*, denoted G^t_O, is introduced

[24] György et al. (2007) call this setting by the bandit feedback and refer to the bandit feedback (in our definition) as the restricted bandit feedback.

and elegantly represents this observability model: a vertex i is connected to a vertex j in the observation graph if by playing action i, the learner also observes the loss of action j. In the case where observation graphs are fixed, Mannor and Shamir (2011) propose the ExpBan algorithm guaranteeing a regret at most $O(\sqrt{\chi T \log D})$ where χ is the clique-partition number[25] of the observation graph. For the case where observations graphs may change through time, another algorithm, called ELP,[26] is proposed; it guarantees a regret upper-bound in the order of $O\left(\sqrt{\log D \sum_{t\in[T]} \alpha(G^t_O)}\right)$ if $G^t_O, t \in [T]$ are all undirected graphs, and in the order of $O\left(\sqrt{\log D \sum_{t\in[T]} \chi(G^t_O)}\right)$ for directed graphs. Here, $\alpha(G^t_O)$ is the independence number of G^t_O.[27] These results are then generalized by the series of works from Alon, Cesa-Bianchi, Dekel, et al. (2015), Alon, Cesa-Bianchi, Gentile, Mannor, et al. (2017), and Alon, Cesa-Bianchi, Gentile, and Mansour (2013) where two new algorithms are proposed: Exp3-SET and Exp3-DOM, corresponding to the uninformed setting (i.e., the learner only observes observation graphs after making decisions) and the informed setting (i.e., observation graphs are revealed before making decisions). The corresponding regret upper-bounds provided by Exp3-SET and Exp3-DOM are $\tilde{O}\left(\sqrt{\log D \sum_{t\in[T]} \text{mas}(G^t_O)}\right)$ and $\tilde{O}\left(\log D \sqrt{\log(DT) \sum_{t\in[T]} \alpha(G^t_O)}\right)$.[28] Note importantly that these algorithms involve the computation of maximum acyclic sub-graphs and dominating sets, which is *not* computational tractable in generic graphs. Extended results in the framework of online learning with graph-structured feedback (which is more general than MAB) are also presented in Alon, Cesa-Bianchi, Dekel, et al. (2015). Another extension is where observation graphs are never revealed completely to the learner (see A. Cohen et al. (2016)).

A more general model than MAB with side-observations is OComb with side-observations (SOComb), proposed by Kocák et al. (2014). A definition of SOComb is presented in Section 2.2.1. Importantly, although it has not yet been explicitly formulated in the literature, the OSP with side-observation model (SOOSP) that we introduce in Section 7.2 can be considered as an instance of SOComb. The concept of observation graphs in SOComb and SOOSP is adopted from the similar concept in MAB with side-observations discussed above. The main difference between the observations graphs G^t_O in SOOSP and that of MAB with side-observations is that the former captures side-observations between edges (in the graph G) whereas the latter captures side-observations between paths (i.e., actions/arms). Now, in SOComb, the standard algorithm is the FPL-IX algorithm (introduced by Kocák et al. (2014)), which could be applied directly to SOOSP. FPL-IX belongs to the family of follow-the-perturbed-leader algorithms and it involves an implicit exploration by geometric resampling. FPL-IX guarantees a regret at most $O\left(\log(nDT)n^{3/2}\sqrt{(d + C\sum_{t\in[T]} \alpha(G^t_O))(\log D + 1)}\right)$.[29] In

[25] It is defined as the smallest number of cliques into which the nodes can be partitioned.

[26] ELP stands for "Exponentially-weighted algorithm with Linear Programming".

[27] Briefly put, the independence number of a graph is the largest number of vertices without edges between them. See Section 8.2.2 for a formal definition.

[28] Here, $\text{mas}(G^t_O)$ is the size of the maximum acyclic sub-graph in G^t_O.

[29] Here, C is a constant chosen by Kocák et al. (2014).

expectation, FPL-IX is efficiently implementable if there exists an efficient oracle solving an optimization oracle at each stage; its expected running time is proved to be at most D times the running time of the oracle; a high-probability guarantee in the running time is also given by Kocák et al. (2014). Note also that Kocák et al. (2014) also considers another algorithm, called Exp3-IX, that runs in SOComb; however, its running time is exponential in terms of D, therefore, only the trivial case where $n = 1$ is analyzed.

The state-of-the-art algorithm for SOOSP *is* FPL-IX *(adopted from the literature of* SOComb*). However, in applying directly* FPL-IX *into* SOOSP*, we encounter three main issues: (i) the efficiency of* FPL-IX *is* not *guaranteed in worst-case scenarios and there is still room for improvement; (ii)* FPL-IX *requires that there exists an efficient oracle that solves an optimization problem at each stage; (iii)* FPL-IX *is designed for general cases of observation graphs and it still lacks analyses of several particular cases that may allow improvements in regret's guarantees.*

7.3.4 Learning Equilibria in Games

As mentioned, adversarial bandits (and extended frameworks such as adversarial OComb and OLO) can be formulated as repeated games between several players. An interesting situation that might arise is when all players apply online learning policies; particularly, where players do not know all parameters of the games and decisions of each player only depends on past observations on her payoffs (and not opponents' payoffs)—this is called the uncoupling ways of playing (see Cesa-Bianchi and Lugosi (2006)). In this situation, it is important to characterize the conditions under which the equilibria (and/or other game-theoretic solution concepts) of the game are *learnable*; in other words, the leading question is: when will sequences of uncoupling behaviors of the players converge toward the equilibria?

Although in this thesis, we do not focus on this question when analyzing online resource allocation games, we review here some notable results in learning equilibria in games since it is another important approach showing connections between online learning and game-theory. In the literature, equilibria are (often) said to be learnable in three (standard) different senses: (i) when the sequence of profile of player's mixed strategies (also called the actual sequence of plays) at stage t converges toward the set of equilibria as $t \to \infty$; (ii) the (marginal) empirical distributions of plays converges to the set of equilibria and (iii) the joint empirical frequencies of plays converge to the set of equilibria.[30] A somewhat disappointed result is that there exists instances of games in which none of the convergence described above occurs with Nash equilibria (see Hart and Mas-Colell (2003)). On the other hand, it is proven that by following exponential-weight strategies (Auer, Cesa-Bianchi, Freund, et al. (1995) and Freund and Schapire (1997)), the empirical distribution of plays converge to coarse correlated

[30]See Cesa-Bianchi and Lugosi (2006) and Faure et al. (2015) for more details

equilibria[31] of the game (see Hart and Mas-Colell (2000)). This result does not always provide interpretable predictions since coarse correlated equilbria may contain very non-rational strategies (see also the discussion in Heliou et al. (2017)).

Importantly, in some specific class of games, Nash equilibria are learnable. An example is the two-player zero-sum game (see Section 2.1 for a definition). In this class of games, the (marginal) empirical distributions of plays converge almost surely to the set of Nash equilibria (see Remark 7.4 in Cesa-Bianchi and Lugosi (2006)). On the other hand, Nash equilibria of zero-sum games are not learnable for the case of the joint empirical frequencies; but correlated equilibria are. Another class of games where results are available is potential games: Palaiopanos et al. (2017) proves the convergence towards Nash equilibrium of the multiplicative weight update rule; Heliou et al. (2017) shows the convergence of the actual sequence of play converges to Nash equilibrium (to an approximate equilibrium if the feedback is restricted).

It is interesting to analyze the conditions such that equilibria of resource allocation games are learnable when players use regret-minimization policies in the corresponding online versions. However, the set of techniques involved in this approach is very different from that of our objectives and results presented in this thesis. We do not study the equilibria learning problem here and leave it for future works.

. . . .

Summary: In this chapter, we considered online resource allocation games. As a case study, we defined the online CB game under a variety of feedback settings. We then showed that this game can be cast into an OComb or an OSP which allows us to use the corresponding regret-minimization algorithms/techniques. We also discussed several other online resource allocations games, including the online hide-and-seek game, and showed their conversions into the OSP framework. We formally defined the OSP with side-observations model and reviewed the literature of MAB, OComb, OSP and other related works as a preparation for our studies on online resource allocation games in the following chapters.

[31] This stable state notion is proposed by Moulin and Vial (1978), it is essentially weaker than the notion of Nash equilibrium.

Chapter 8

OSP with Side-Observations—Applications to the Online Semi-bandit CB Game and Beyond

Some of the ideas and results presented in this chapter have previously appeared in our publication Vu, Loiseau, Silva, and Tran-Thanh (2020).[a]

[a] A note on the terminology: the online shortest path problem, defined and studied in this thesis, is called the path planning problem in Vu, Loiseau, Silva, and Tran-Thanh (2020). These terms are often used interchangeably in the literature of bandit problems and online learning.

In this chapter, we start our analysis on the online shortest path problem and its applications to online resource allocation games with combinatorial structures; more specifically, we focus on the SOOSP model which we introduced in Section 7.2. We choose to study the SOOSP model because it allows a more refined[1] representation of the online semi-bandit CB game—the game we choose to use as the main motivation of this chapter. By casting this game into the SOOSP framework, we are able not only to exploit the structure of the learner's strategy set (by its graphical representation) but also to capture well all information available from the feedback that the learner observes at the end of each stage. Importantly, the SOOSP model is also useful for studying several other online resource allocation games and in fact, its scope of applications goes even further than these games. Therefore, our perspective is to first study the generic model of SOOSP and look for its solution (i.e., a regret-minimization algorithm) then we apply our findings to online resource allocation game with combinatorial structures that are cast to SOOSP, e.g., the online semi-bandit CB game, and analyze the derived benefits.

The study of the SOOSP model entails two key challenges. First, state-of-the-art algorithms in OComb with side-observations (SOComb), that are applicable to SOOSP, do not guarantee efficient implementations in worst-scenarios. Second, these algorithms are only analyzed in the generic setting of observation graphs and regret analyses for several particular cases of interest are omitted (these cases are encountered in some

[1] The feedback system in the SOOSP model, interpolates between semi-bandit feedback and full-information, is more refined and flexible than these two standard feedback settings (see Definition 7.2.1).

SOOSP instances corresponding to online resource allocation games). Our goal in this chapter is to design novel regret-minimization algorithms in SOOSP fixing the issues mentioned above. The remainders of this chapter are arranged as follows: In Section 8.1.1, we develop the conversion of the online semi-bandit CB game to the SOOSP model to show a motivational example and justify this model. In Section 8.1.2, we elaborate the state-of-the-art of the study of SOOSP (from the literature review on SOComb and SOOSP presented in Section 7.3.3) and define explicitly our objectives in studying the SOOSP model. In Section 8.2, we present a novel algorithm, called Exp3-OE, working in any generic SOOSP and analyze its performance as well as its running time. In Section 8.3, we turn our focus back to online resource allocation games (with combinatorial structures) that can be cast into SOOSP and analyze the application of the Exp3-OE algorithm to several of them, including the online semi-bandit CB game, the online hide-and-seek game and the online CB game under the full-information feedback setting.

8.1 Motivations and Challenges in SOOSP

In this section, we present the extra information that the learner can deduce in the online semi-bandit CB game; this serves as a motivation for us to study SOOSP. We also revisit the challenges and explicitly state our contributions in studying SOOSP.

8.1.1 The Online Semi-Bandit CB Game as an SOOSP

We first recall briefly the definition of the online semi-bandit CB game as follows: at each stage, a learner who has a budget of k troops plays a discrete CB game against an adversary across n battlefields ($k, n \in \mathbb{N} \backslash \{0\}$ are fixed and known by the learner); at the end of the stage, she receives limited feedback that is the gain (loss) she obtains from each battlefield and whether she wins, loses or gets a tie there; finally, the learner's objective is to minimize the expected regret. It is natural to use this model to capture several situations in practice; we refer the interested readers to Section 7.1.1 for some motivational examples.

From our discussion in Section 7.1.2, we know that any CB game can be cast as an online shortest path problem (OSP). In particular, the information that the learner receives at the end of each stage in the online semi-bandit CB game straightforwardly corresponds to the so-called *semi-bandit* feedback setting of OSPs, i.e., the learner observes the edges' losses belonging to her chosen path.[2] Furthermore, since the learner knows the winner-determination rule of the CB game (i.e., the learner wins a battlefields as long as her allocation is higher than the opponent's allocation), in the online semi-bandit CB game, she can deduce (without any extra cost) further information as follows:

> *If she allocates $z_t(i)$ troops to battlefield i and wins, she knows that if she had allocated more than $z_t(i)$ troops to i, she would also have won (and receives a loss 0 from*

[2] See Section 2.2.3 for a formal definition of OSP with semi-bandit feedback.

this battlefield).

If she knows the allocations are tie at battlefield i, she knows exactly the adversary's allocation to this battlefield and deduces all the losses she might have suffered if she had allocated differently to battlefield i.

If she allocates $z_t(i)$ troops to battlefield i and loses, she knows that if she had allocated less than $z_t(i)$ to i, she would also have lost (and received a loss $b_t(i)$ from this battlefield).

Recall that in casting online CB games into OSPs, we use the DAG $G_{k,n}$, defined in Definition 7.1.2, where each edge represents an allocation of troops to a battlefield. Therefore, the extra information deduced above implies that in the OSP corresponding to the online semi-bandit CB game, besides the losses of the edges belonging to the chosen path (i.e., the semi-bandit feedback), the learner also observes losses of some other edges that may not belong to the chosen path. This is the idea of the SOOSP model that we introduced in Section 7.2. We formally state the following proposition:

Proposition 8.1.1. *Any online semi-bandit CB game (where the learner allocates k troops across n battlefields) can be cast as an* SOOSP *(on the graph $G_{k,n}$).*

The following example illustrates the side-observations (and the observation graphs) in an instance of the online semi-bandit CB game, represented as an SOOSP.

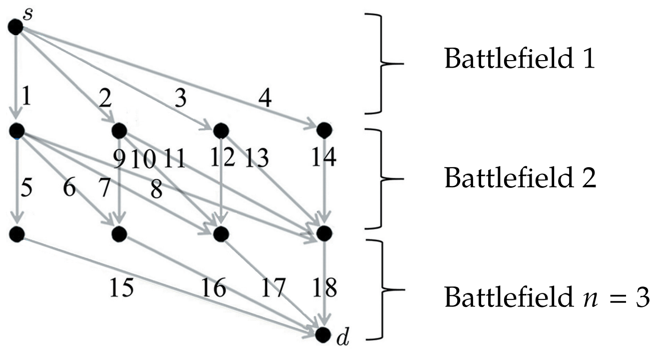


Figure 8.1: The graph $G_{3,3}$ corresponds to the CB game with $n = k = 3$. The edges are labeled from 1 to $E = 18$.

Example 8.1.2. *Consider an online semi-bandit CB game with $n = 3$ battlefields and that the learner has $k = 3$ troops. In Figure 8.1, we illustrate the graph $G_{3,3}$ corresponding to this instance of the game.*[3] *Now, at stage t, assume that the learner chooses to allocate 1 troop to each battlefield; on the graph $G_{3,3}$, this strategy is equivalent to choosing the path going through the edges* 2, 10 *and* 17*. We assume that the adversary chooses the allocations such that the outcomes of the game at this stage are the following: the learner loses in battlefield 1, wins in battlefield 2 and a tie occurs in battlefield 3. The learner trivially observes the losses of her chosen edges including* 2, 10 *and* 17 *(the losses are $b_t(1)$, 0 and $b_t(3)/2$ respectively); moreover, she can also deduce the side-observations:*

[3]This is actually the example that was already presented in Section 7.1.2: Figure 8.1 is exactly Figure 7.1 with the edges labeled from 1 to $E = 18$.

> *Edge 2 (corresponding to allocating 1 troop to battlefield 1) reveals that the loss on edge 1 (corresponding to allocating 0 troop to battlefield 1) is also $b_t(1)$ since she would also have lost this battlefield by choosing this edge.*
>
> *Edge 10 (corresponding to allocating 1 troop to battlefield 2) reveals that the losses on the edges 6, 7, 8, 11, and 13 (corresponding to allocating at least 1 troop to battlefield 2) are all 0 since she would also have won this battlefield by choosing these edges.*
>
> *Edge 17 (corresponding to allocating 1 troop to battlefield 3) reveals the losses on the edges 15, 16, 18 (corresponding to other possible allocations in battlefield 3) since she knows that the adversary also allocates 1 troop in this battlefield (a tie occurs here) and that she would have won or lost this battlefield if she had chosen these edges.*

The observation graph corresponding to the allocations chosen by the learner and the outcomes of the battlefields in this stage (also depending on the adversary's allocations) is illustrated in Figure 8.2. Note that since there are edges in $G_{k,n}$ that refer to the same allocation (e.g., the edges 5, 9, 12, and 14 in $G_{3,3}$ all refer to allocating 0 troops to battlefield 2), in all the observation graphs, the vertices corresponding to these edges are always connected.

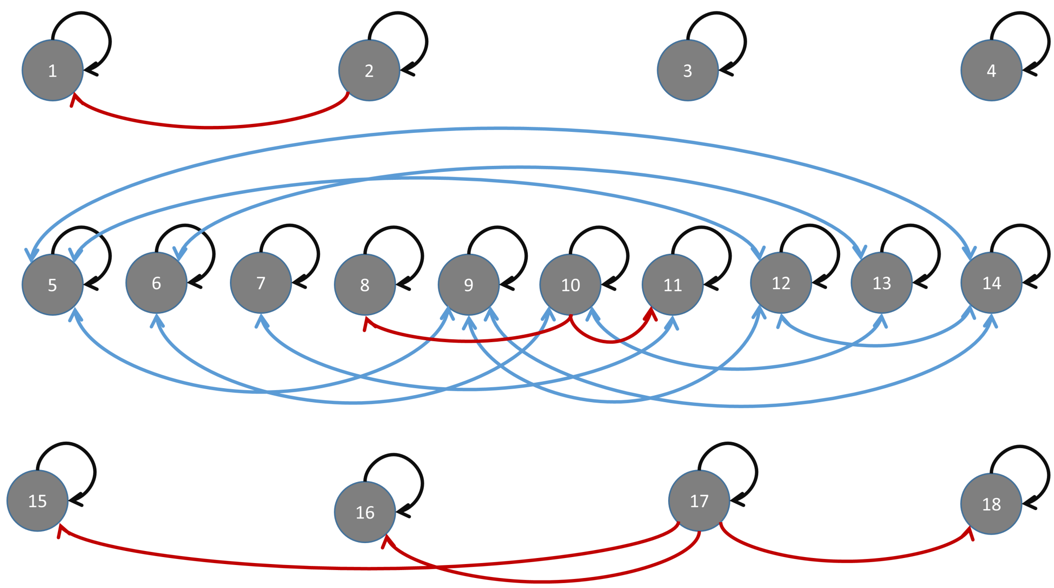


Figure 8.2: The observation graph corresponding to the online semi-bandit CB game given in Example 8.1.2. The (directed) red arrows illustrate the side-observations depending on the particular allocations of the learner and the outcomes of the battlefields; the vertices connected by (two-directional) blue arrows always mutually reveal one another (they are always connected).

8.1.2 Challenges in SOOSP and Our Contributions

As an instance of the online combinatorial optimization with side-observations framework (SOComb), the state-of-the-art regret-minimization algorithm in SOOSP is the

FPL-IX algorithm—a member of the class of Follow-the-Perturbed-Leader algorithms; it is proposed by Kocák et al. (2014) to study SOComb.[4] However, in applying directly FPL-IX into SOOSP, we encounter two main problems: (i) the efficiency of FPL-IX is only guaranteed with high-probability and its running time is still in order of $\Omega(T^{2/3})$ where T is the time horizon; (ii) FPL-IX requires that there exists an efficient oracle that solves an optimization problem at each stage. Both of these issues are incompatible with our aim of learning in the online semi-bandit CB game (and in other resource allocation games with combinatorial structures): although the probability that FPL-IX fails to terminate is small, this could lead to issues in implementing it in practice where the learner is obliged to quickly give a decision in each stage; moreover, it is unclear which oracle should be used in applying FPL-IX to this game.

In this chapter, we focus instead on another prominent class of OComb algorithms, the family of Exp3-type algorithms (see Section 2.2.2 for basic variants of Exp3). One of the key open questions in this field is how to design a variant of Exp3 with efficient running time and good regret guarantees for OComb problems in each feedback setting (see, e.g., Cesa-Bianchi and Lugosi (2012)). Our main contribution of this chapter is to propose an Exp3-type algorithm for SOOSPs that solves both of the aforementioned issues of FPL-IX and provides good regret guarantees; i.e., we give an affirmative answer to an important subset of the above-mentioned open problem. In more details, our proposed algorithm, called Exp3-OE, is applicable to any instance of SOOSP such that (i) Exp3-OE is always guaranteed to run efficiently, i.e., its running time is polynomial in terms of the number of edges of the graph in SOOSP and linear in T, without the need of any auxiliary oracle; (ii) Exp3-OE guarantees an upper-bound on the expected regret matching in order with the best benchmark in the literature (the FPL-IX algorithm). We also prove further improvements under additional assumptions on the observation graphs that have been so-far ignored in the literature.

8.2 Exp3-OE - An Efficient Algorithm for SOOSP

In this section, we present a new algorithm for SOOSP, called Exp3-OE (OE stands for Observable Edges). The guarantees on the expected regret of Exp3-OE in SOOSP is analyzed in Section 8.2.2. Moreover, Exp3-OE always runs efficiently in polynomial time in terms of E—the number of edges of G; this is discussed in Section 8.2.1. To prepare for the presentation of the Exp3-OE algorithm, we introduce two new definitions regarding side-observations in an SOOSP on a graph G (with the edges set $\mathcal{E}$ and the paths set $\mathcal{P}$):

$$\mathbb{O}_t(e) := \left\{ \boldsymbol{p} \in \mathcal{P} : \exists e' \in \boldsymbol{p}, e' \to e \right\}, \forall e \in \mathcal{E},$$
$$\mathbb{O}_t(\boldsymbol{p}) := \left\{ e \in \mathcal{E} : \exists e' \in \boldsymbol{p}, e' \to e \right\}, \forall \boldsymbol{p} \in \mathcal{P}.$$

Here, recall that $e' \to e$ indicates that by choosing a path containing the edge e', both the losses on the edges e' and e are revealed to the learner (in this case, we also say

[4] See Section 2.2.1 for a definition and see Section 7.3 for a literature review of SOComb.

that e' reveals e), i.e., in the corresponding observation graph, there is an edge from the vertex corresponding to e' to the vertex corresponding to e. Intuitively, $\mathbb{O}_t(e)$ is the set of all paths that, if chosen, reveal the loss on the edge e and $\mathbb{O}_t(\boldsymbol{p})$ is the set of all edges whose losses are revealed if the path $\boldsymbol{p}$ is chosen. Trivially, $\boldsymbol{p} \in \mathbb{O}(e) \Leftrightarrow e \in \mathbb{O}(\boldsymbol{p})$. Moreover, due to the semi-bandit feedback in SOOSP, if $\boldsymbol{p}^* \ni e^*$, then $\boldsymbol{p}^* \in \mathbb{O}_t(e^*)$ and $e^* \in \mathbb{O}_t(\boldsymbol{p}^*)$.

Apart from the results for SOOSP with general observation graphs, in this chapter, we additionally present several results under two particular classes of observations graphs, encountered in some practical instances (e.g., the online semi-bandit CB game and the online HS game), that provide more refined regret bounds compared to cases that were considered by Kocák et al. (2014):

(i) *symmetric* observation graphs where for each edge from v_e to $v_{e'}$, there also exists an edge from $v_{e'}$ to v_e (i.e., if $e \rightarrow e'$ then $e' \rightarrow e$); i.e., G^O_t is an undirected graph;

(ii) observation graphs that satisfy the following *Assumption* $(A1)$ which requires that if two edges belong to a path in G, then they cannot simultaneously reveal the loss of another edge:

Assumption $(A1)$: *For any $e \in \mathcal{E}$, if $e' \rightarrow e$ and $e'' \rightarrow e$, then $\nexists \boldsymbol{p} \in \mathcal{P} : \boldsymbol{p} \ni e', \boldsymbol{p} \ni e''$.*

Note also that although Definition 7.2.1 allows a non-oblivious adversary, in this chapter, for the sake of simplicity of the model and to compare with the state-of-the-art, we only focus on the SOOSP with an oblivious adversary (i.e., the generation of the losses at each stage do not depend on the past actions of the learner). In this setting, recall that the *expected regret* can be rewritten as $R_T := \mathbb{E}\left[\sum_{t \in [T]} L\left(\tilde{\boldsymbol{p}}_t\right)\right] - \min_{\boldsymbol{p} \in \mathcal{P}} \sum_{t \in [T]} L\left(\boldsymbol{p}\right)$. The results presented in this chapter involving the Exp3-OE algorithm can also be extended to the case of non-oblivious adversary.

Algorithm 10: The Exp3-OE Algorithm for SOOSP.

```
Input: T, η, β > 0, graph G
Initialize w_1(e) := 1, ∀e ∈ E
for t = 1 to T do
    Loss vector ℓ_t ∈ [0,1]^E is chosen adversarially (unobserved).
    x_t(p) := ∏_{e∈p} w_t(e) / ( ∑_{p'∈P} ∏_{e'∈p'} w_t(e') ), ∀p ∈ P.
    Use WPS Algorithm (Algorithm 4) to sample a path p̃_t ∈ P according
      to x_t(p̃_t) .
    Suffer the loss L_t(p̃_t) = ∑_{e∈p̃_t} ℓ_t(e).
    Observation graph G_t^O is generated and ℓ_t(e), ∀e ∈ O_t(p̃_t) are observed.
    ℓ̂_t(e) := ℓ_t(e) I_{e∈O_t(p̃_t)} / (q_t(e) + β), ∀e ∈ E, where q_t(e) := ∑_{p∈O_t(e)} x_t(p) is
      computed by Algorithm 11 (see Section 8.2.1).
    Update the weights w_{t+1}(e) := w_t(e) · exp(−η ℓ̂_t(e)).
```

The pseudo-code of Exp3-OE is given in Algorithm 10. As an Exp3-type algorithm, Exp3-OE relies on the average weights sampling where at stage t we update the weight $w_t(e)$ on each edge e by the exponential rule (line 9). For each path $\boldsymbol{p}$, we denote the path weight $w_t(\boldsymbol{p}) := \prod_{e\in\boldsymbol{p}} w_t(e)$ and rewrite the terms in line 4 of Algorithm 10 as follows:

$$x_t(\boldsymbol{p}) := \frac{\prod\limits_{e\in\boldsymbol{p}} w_t(e)}{\sum\limits_{\boldsymbol{p}'\in\mathcal{P}} \prod\limits_{e'\in\boldsymbol{p}'} w_t(e')} = \frac{w_t(\boldsymbol{p})}{\sum\limits_{\boldsymbol{p}'\in\mathcal{P}} w_t(\boldsymbol{p}')}, \forall \boldsymbol{p} \in \mathcal{P}. \tag{8.1}$$

Line 5 of Exp3-OE involves a sub-algorithm, called the WPS algorithm, that samples a path $\boldsymbol{p} \in \mathcal{P}$ with probability $x_t(\boldsymbol{p})$ (the sampled path is then denoted by $\tilde{\boldsymbol{p}}_t$) from any input $\{w_t(e), e \in \mathcal{E}\}$ at each stage t. This algorithm is based on a classical technique called *weight pushing* that was presented as Algorithm 4 in Section 2.2.3.

Compared to other instances of the Exp3-type algorithms, Exp3-OE has two major differences. First, at each stage t, the loss of each edge e is estimated by $\hat{\boldsymbol{\ell}}_t(e)$ (line 8) based on the term $q_t(e)$ and a parameter β. Intuitively, $q_t(e)$ is the probability that the loss on the edge e is revealed from playing the chosen path at t. Second, the implicit exploration parameter β added to the denominator allows us to "pretend to explore" in Exp3-OE without knowing the observation graph G_t^O before making the decision at stage t (the uninformed setting).[5] Unlike the standard Exp3, the loss estimator used in Exp3-OE is *biased*, i.e., for any $e \in \mathcal{E}$,

$$\begin{aligned}\mathbb{E}_t\left[\hat{\boldsymbol{\ell}}_t(e)\right] &= \sum_{\tilde{\boldsymbol{p}}\in\mathcal{P}} x_t(\tilde{\boldsymbol{p}}) \frac{\boldsymbol{\ell}_t(e)}{q_t(e)+\beta} \mathbb{I}_{\{e\in\mathbb{O}_t(\tilde{\boldsymbol{p}})\}} \\ &= \sum_{\tilde{\boldsymbol{p}}\in\mathbb{O}_t(e)} x_t(\tilde{\boldsymbol{p}}) \frac{\boldsymbol{\ell}_t(e)}{\sum\limits_{\boldsymbol{p}\in\mathbb{O}_t(e)} x_t(\boldsymbol{p})+\beta} \leq \boldsymbol{\ell}_t(e).\end{aligned} \tag{8.2}$$

Here, $\mathbb{E}_t$ denotes the expectation w.r.t. the randomness of choosing a path at stage t. Second, unlike standard Exp3 algorithms that keep track and update on the weight of each path,[6] the weight pushing technique is applied at line 5 (via the WPS algorithm) and line 8 (via Algorithm 11 in Section 8.2.1) where we work with edges weights instead of paths weights (recall that $E \ll P$).

8.2.1 Running Time Efficiency of the Exp3-OE Algorithm

First, let us recall the complexity of the WPS algorithm used in line 5 of the Exp3-OE algorithm; in other words, we recheck the running time of the weight pushing technique. For the WPS algorithm to work, for any vertex u in G, it is needed to compute the terms

[5]Intuitively, the use of the parameter β guarantees a lower-bound of the probability of being chosen of each path; this is called the implicit exploration scheme to distinguish with the explicit exploration scheme, i.e., mixing the "exploitation" sampling based on weights with a exploration distribution (see e.g., the ComBand algorithm, presented in Algorithm 2, for OComb with bandit feedback).

[6]In principle, each OSP problem can be rewritten as an MAB where each path corresponds to an arm; if we apply directly the standard Exp3 (Algorithm 1) into this MAB, it is needed to keep track and update the weight for each arm/path.

$H_t(s,u) := \sum_{\boldsymbol{p} \in \mathcal{P}_{s,u}} \prod_{e \in \boldsymbol{p}} w_t(e)$ and $H_t(u,d) := \sum_{\boldsymbol{p} \in \mathcal{P}_{u,d}} \prod_{e \in \boldsymbol{p}} w_t(e)$.[7] Intuitively, $H_t(u,v)$ is the aggregate weight of all paths from vertex u to vertex v at stage t. These terms can be computed recursively based on dynamic programming: we have presented in Section 2.2.3 the WP algorithm (Algorithm 3) that outputs $H_t(s,u), H_t(u,d), \forall u$ from any input $\{w_t(e), e \in \mathcal{E}\}$ in $O(E)$ time. Then, a path in G is sampled sequentially edge-by-edge based on these terms by the WPS algorithm. As a conclusion, the WP and WPS algorithms run efficiently in $O(E)$ time.

The final non-trivial step to efficiently implement Exp3-OE is to compute $q_t(e)$ in line 8, i.e., the probability that an edge e is revealed at stage t. Note that $q_t(e)$ is the sum of $|\mathbb{O}_t(e)| = O(P)$ terms; therefore, a direct computation is inefficient while a naive application of the weight pushing technique can easily lead to errors. To compute $q_t(e)$, we propose Algorithm 11, a non-straightforward application of weight pushing, in which we consecutively consider all the edges $e' \in \mathfrak{R}_t(e) := \{e' \in \mathcal{E} : e' \rightarrow e\}$. Then, we take the sum of the terms $x_t(\boldsymbol{p})$ of the paths $\boldsymbol{p}$ going through e' by the weight pushing technique while making sure that each of these terms $x_t(\boldsymbol{p})$ is included only once, even if $\boldsymbol{p}$ has more than one edge revealing e (this is a non-trivial step). In Algorithm 11, we denote by $\mathbb{C}(u)$ the set of the direct successors of any vertex $u \in \mathcal{V}$. We give a proof that Algorithm 11 outputs exactly $q_t(e)$ as defined in line 8 of Algorithm 10 in Appendix E.1. Algorithm 11 runs in $O(|\mathfrak{R}_t(e)|E)$ time; therefore, line 8 of Algorithm 10 can be done in at most $O(E^3)$ time.

Algorithm 11: Compute $q_t(e)$ of an edge e at stage t.

Input: $e \in \mathbb{O}_t(\tilde{\boldsymbol{p}}_t)$, set $\mathfrak{R}_t(e)$ and $w_t(\bar{e}), \forall \bar{e} \in \mathcal{E}$.

Output: $q_t(e)$.

Initialize $\bar{w}(\bar{e}) := w_t(\bar{e}), \forall \bar{e} \in \mathcal{E}$ and $q_t(e) := 0$.

Compute $H^*(s,d)$ by WP Algorithm (Algorithm 3) with inputs G and $\{w_t(\bar{e}), \bar{e} \in \mathcal{E}\}$.

for $e' \in \mathfrak{R}_t(e)$ **do**

  Compute $H(s,u), H(u,d), \forall u \in \mathcal{V}$ by WP Algorithm with inputs G and $\{\bar{w}(\bar{e}), \forall \bar{e} \in \mathcal{E}\}$.

  $K(e') := H(s,u_{e'}) \cdot w(e') \cdot H(v_{e'}, d)$; here, e' is the edge going from $u_{e'}$ to $v_{e'} \in \mathbb{C}(u_{e'})$.

  $q_t(e) := q_t(e) + K(e')/H^*(s,d)$.

  Update $\bar{w}(e') = 0$.

In conclusion, Exp3-OE runs in at most $O(E^3T)$ time, this guarantee works even for the worst-case scenario. For comparison, the FPL-IX algorithm runs in $O(E|\mathcal{V}|^2T)$ time in expectation and in $\tilde{O}(n^{1/2}E^{3/2}\log(E/\delta)T^{3/2})$ time with a probability at least $1-\delta$ for an arbitrary $\delta > 0$.[8] That is, FPL-IX might fail to terminate with a strictly

[7]Recall that the notation $\mathcal{P}_{u,v}$ denotes the set of all paths in G going from vertex u to vertex v.

[8]If one runs FPL-IX with Dijkstra's algorithm as the optimization oracle and with parameters chosen by Kocák et al. (2014).

positive probability[9] and it is not guaranteed to have efficient running time in all cases. Moreover, although the complexity bound (either in expectation or with high probability) of FPL-IX is slightly better in terms of E, the complexity bound of Exp3-OE improves that by a factor of $\sqrt{T}$. As is often the case in no-regret analysis, we consider the setting where T is significantly larger than other parameters of the problems; this is also consistent with the motivational applications of the online semi-bandit CB game presented in Section 7.1.1. Therefore, our contribution in improving the algorithm's running time in terms of T is relevant.

8.2.2 Performance of the Exp3-OE Algorithm

In this section, we present an upper-bound of the expected regret achieved by the Exp3-OE algorithm in SOOSP. For the sake of brevity, with $x_t(\boldsymbol{p})$ defined in (8.1), for any $t \in [T]$ and $e \in \mathcal{E}$, we denote:

$$r_t(e) := \sum\nolimits_{\boldsymbol{p} \ni e} x_t(\boldsymbol{p}) \text{ and } Q_t := \sum\nolimits_{e \in \mathcal{E}} r_t(e) / (q_t(e) + \beta).$$

Intuitively, $r_t(e)$ is the probability that the chosen path at stage t contains an edge e and Q_t is the summation over all the edges of the ratio of this quantity and the probability that the loss of an edge is revealed (plus β). We can bound the expected regret with this key term Q_t.

Theorem 8.2.1. *The expected regret of the* Exp3-OE *algorithm in the* SOOSP *satisfies:*

$$R_T \leq \log(P)/\eta + \left[\beta + (n \cdot \eta)/2\right] \cdot \sum\nolimits_{t \in [T]} Q_t. \tag{8.3}$$

A complete proof of Theorem 8.2.1 can be found in Appendix E.2 and has an approach similar to Alon, Cesa-Bianchi, Gentile, and Mansour (2013) and Cesa-Bianchi and Lugosi (2012) with several necessary adjustments to handle the new biased loss estimator in Exp3-OE. To see the relationship between the structure of the side-observations of the learner and the bound of the expected regret, we look for the upper-bounds of Q_t in terms of the observation graphs' parameters. This result relates to the some notions of the graphs defined as follows: an *independent (vertex) set* of a graph G is a subset of the vertices set $\mathcal{V}$ such that no two vertices in this subset is connected by an edge in G; and *the independence number* of G is simply the cardinality of the largest independent sets of G. Now, let us denote α_t to be the independence number[10] of G_t^O, we have the following statement.

Theorem 8.2.2. *Let us define* $M := \lceil 2E^2/\beta \rceil$, $N_t := \log\left(1 + \frac{M+E}{\alpha_t}\right)$ *and* $K_t := \log\left(1 + \frac{nM+E}{\alpha_t}\right)$. *Upper-bounds of* Q_t *in different cases of* G_t^O *are given in the following table:*

[9] A stopping criterion for FPL-IX can be chosen to avoid this issue but it raises the question on how one chooses the criterion such that the regret guarantees hold.

[10] The independence number of a directed graph is computed while ignoring the direction of the edges.

	Satisfies $(A1)$	Does not satisfy $(A1)$
Symmetric	α_t	$n\alpha_t$
Non-Symmetric	$1+2\alpha_t N_t$	$2n\,(1+\alpha_t K_t)$

A proof of this theorem is given in Appendix E.4. The main idea of this proof is based on several graph theoretic lemmas that are extracted from Alon, Cesa-Bianchi, Gentile, and Mansour (2013), Kocák et al. (2014), and Mannor and Shamir (2011). These lemmas establish the relationship between the independence number of a graph and the ratios of the weights on the graph's vertices that have similar forms to the key-term Q_t. The case where observation graphs are non-symmetric and do not satisfy Assumption $(A1)$ is the most general setting. Moreover, as showed in Theorem 8.2.2, the bounds of Q_t are improved if the observation graphs satisfy either the symmetry condition or Assumption $(A1)$. Intuitively, given the same independence numbers, a symmetric observation graph gives the learner more information than a non-symmetric one; thus, it yields a better bound on Q_t and the expected regret. On the other hand, Assumption $(A1)$ is a technical assumption that allows us to use different techniques in the proofs to obtain better bounds. These cases have not been explicitly analyzed in the literature while they are satisfied by several practical situations, including the online semi-bandit CB game (and the online Hide-and-Seek game considered in Section 8.3).

Finally, we give results on the upper-bounds of the expected regret, obtained by the Exp3-OE algorithm, presented as a corollary of Theorem 8.2.1 and Theorem 8.2.2.

Corollary 8.2.3. *In* SOOSP, *let α be an upper bound of $\alpha_t, \forall t \in [T]$; with appropriate choices of the parameters η and β, the expected regret of the* Exp3-OE *algorithm is:*[11]

(i) $R_T \leq \tilde{O}(n\sqrt{T\alpha\log(P)})$ *in the general cases.*

(ii) $R_T \leq \tilde{O}(\sqrt{nT\alpha\log(P)})$ *if Assumption $(A1)$ is satisfied by the observation graphs G^O_t, $\forall t \in [T]$.*

A proof of Corollary 8.2.3 and the choices of the parameters β and η (these choices are non-trivial) yielding these results is given in Appendix E.5. We can extract from this proof several more explicit results as follows:

(i) In the general case (i.e., Assumption $(A1)$ may not hold): if observations graphs are non-symmetric then $R_T \leq O\left(n\sqrt{T\alpha\log(P)[1+\log(\alpha+\alpha\log(\alpha)+E)]}\right)$; if they are symmetric then $R_T \leq (3/2)n\sqrt{T\alpha\log(P)}+\sqrt{nT\alpha}$.

(ii) If all observation graphs satisfy Assumption $(A1)$: if the observations graphs are non-symmetric then $R_T \leq O\left(\sqrt{nT\alpha\log(P)[1+2\log(1+E)]}\right)$; if they are all symmetric then $R_T \leq 2\sqrt{nT\alpha\log(P)}+\sqrt{T\alpha}$.

[11] Recall that $\tilde{O}$ is the variant of the asymptotic notation O that ignores the logarithmic factors (in terms of n and T).

Note that a trivial upper-bound of α_t is the number of vertices of the graph G_t^O which is E (the number of edges in G). In general, the more connected G_t^O is, the smaller α may be chosen; and thus the better the upper-bound of the expected regret is. In the (classical) semi-bandit setting, $\alpha_t = E, \forall t \in [T]$ and in the full-information setting, $\alpha_t = 1$, $\forall t \in [T]$. Finally, we also note that, if $P = \mathcal{O}(\exp(n))$ (this is typical in practice, including the CB and HS games), the bound in Corollary 8.2.3-(i) matches in order with the bounds (ignoring the logarithmic factors) given by the FPL-IX algorithm (see Kocák et al. (2014)). On the other hand, the form of the regret bound provided by the Exp3-IX algorithm (see Kocák et al. (2014)) does not allow us to compare directly with the bound of Exp3-OE in the general SOOSP. Exp3-IX is only analyzed by Kocák et al. (2014) when $n = 1$, i.e., $P = E$; in this case, we observe that the bound given by our Exp3-OE algorithm is better than that of Exp3-IX (by some multiplicative constants).

8.3 Exp3-OE in Online Resource Allocation Games

We now return to our initial motivation–the online resource allocation games and discuss the application of Exp3-OE to several examples in this class of games, including the online semi-bandit CB game (Section 8.3.1), the online hide-and-seek game (Section 8.3.2) and the online CB game with full-information feedback (Section 8.3.3).

8.3.1 Exp3-OE in the Online Semi-Bandit CB Game

In Section 8.1.1, we have shown that any instance of the online semi-bandit CB game (with k troops and n battlefields) can be cast to an SOOSP (on the corresponding graph $G_{k,n}$). Therefore, we can use the Exp3-OE algorithm as a regret-minimization algorithm for the learner in the online semi-bandit CB game. From Section 8.2.1 and the graphs $G_{k,n}$, we see that running Exp3-OE in the online semi-bandit CB game (with k troops and n battlefields) takes at most $\mathcal{O}(k^6 n^3 T)$ time. We remark again that Exp3-OE's running time is linear in T and efficient in all cases unlike when we run FPL-IX in the online semi-bandit CB game. Moreover, since there are edges in $G_{k,n}$ that refer to the same allocation, in all the observation graphs, the vertices corresponding to these edges are always connected. Therefore, an upper bound of the independence number α_t of G_t^O in the CB game is $\alpha_{\mathrm{CB}} = n(k+1) = \mathcal{O}(nk)$. Furthermore, we can verify that for any t, the observation graph G_t^O of the CB game always *satisfies Assumption* ($A1$) and it is *non-symmetric*. We can deduce the following result directly from Corollary 8.2.3:

Corollary 8.3.1. *In the* SOOSP *corresponding to an online semi-bandit CB games where the learner distributes k troops across n battlefields at each stage, the expected regret of the* Exp3-OE *algorithm satisfies:*[12]

$$R_T \leq \tilde{\mathcal{O}}(\sqrt{nT\alpha_{CB}\log(P)}) = \tilde{\mathcal{O}}(\sqrt{Tn^2k\min\{n-1,k\}}).$$

[12]Here, P is the number of paths in $G_{k,n}$ and α_{CB} is an upper-bound of the independence numbers of all observation graphs in the SOOSP.

From Corollary 8.3.1, we note that in the online semi-bandit CB games, the order of the regret bounds given by Exp3-OE is better than that of the FPL-IX algorithm by a factor of $\sqrt{n}$ (thanks to the fact that $(A1)$ is satisfied). Note that n corresponds to the number of battlefields in the CB game that can be large in applications; therefore, this improvement in the regret's guarantee is relevant. More explicitly, in the CB game, FPL-IX has a regret at most $O\left(\log(k^2n^2T)\sqrt{\log(k^2n)(k^2n^4+Cn^4kT)}\right) = \tilde{O}(\sqrt{Tn^4k})$ (C is a constant indicated by Kocák et al. (2014)) and Exp3-OE's regret bound is $O\left(\sqrt{n^2kT\cdot\min\{n-1,k\}[1+2\log(1+k^2n)]}\right)$. If $n-1\leq k$, we can rewrite this bound of Exp3-OE as $\tilde{O}(\sqrt{Tn^3k})$; otherwise, it can be rewritten as $\tilde{O}(\sqrt{Tn^2k^2})$.

Now, we additionally compare the regret guarantees given by our Exp3-OE algorithm and by the OSMD algorithm (see Audibert, Bubeck, and Lugosi (2014)). OSMD is the benchmark algorithm for OComb with semi-bandit feedback (i.e., OSMD ignores the side-observations); note that OSMD does not run efficiently in general. We observe that the expected regret's guarantees of Exp3-OE is better than OSMD in the online semi-bandit CB games if $O\left(n\cdot\log\left(n^3k^5\sqrt{T}\right)\right)\leq k$. Intuitively, Exp3-OE is better than OSMD in games where the learner's budget is sufficiently larger than the number of battlefields. We give a proof of this statement in Appendix E.7.

Next, we discuss the generalizability of our findings into other resource allocation games under the online setting. First, we note again that our conversion of the online CB game into OSP only depends on the strategy set of the learner (and not the specific rule and/or the payoff functions in each game); therefore, it can be straightforwardly applied to other resource allocation games such as the online LB game or the online HS game (as seen in Section 7.1.3). On the other hand, the possibility of deducing side-observations depends on the rule of the game; therefore, the specific type of side-observations in the online semi-bandit CB game might not be found in other games. For example, in the online setting of the discrete Lottery Blotto game with Tullock CSFs (i.e., the game with the same formulation as the online semi-bandit CB game but the Blotto-rule is replaced by the Tullock CSF (defined in (3.2)), even if the learner knows her gain/loss in a battlefield after allocating a certain number of troops, she cannot deduce precisely the losses of other alternative allocations (without observing the precise allocation of the adversary). Note that the model of SOOSP is general and also covers these cases where no side-observation is available; we can also run Exp3-OE in these cases and get the regret's guarantees as in Corollary 8.2.3 but only with $\alpha = E$ (e.g., in the online Lottery Blotto game, $\alpha = E = O(nk^2)$) while the efficiency of Exp3-OE is still maintained. On the other hand, the use of the Exp3-OE algorithm in an instance of the online HS game will be presented in the next section.

8.3.2 Exp3-OE in the Online Hide-and-Seek Game

In Section 7.1.3, we presented the formulation of the online HS game and showed that the instance where at each stage, the learner makes an n-search among k locations satisfying the k_0-coherence constraint ($n, k, k_0 \in \mathbb{N}$ are fixed and $1 \leq n \leq k$, $k_0 \in [0, k-1]$)

can be cast into an OSP by the use of a graph G_{k,n,k_0} (for example, the graph $G_{3,3,1}$ is given in Figure 7.2). Recall that in this case, the strategy set of the learner at each stage is:

$$S_{k,n,k_0} = \{z \in \{1,\dots,k\}^n : |z(i) - z(i+1)| \leq k_0, \forall i \in [n-1]\}.$$

Here, $z(i) = j \in [k]$ implies that in the n-search z, the learner has chosen location j as her i-th move. Moreover, for the sake of simplicity, we only consider the HS games with an oblivious adversary (note that our results can also be extended to the case of non-oblivious adversary). In this section, we address the game instance described above simply as the online HS game.

In the remainder of this section, we focus on the online HS game with the following *feedback setting*: at the end of stage t, the learner only observes the losses from the locations she chose in her n-search, and her objective is to minimize her expected regret over T. In the application of the spectrum sensing problem presented in Section 7.1.3, this feedback setting is equivalent to assuming that the learner (the secondary user) can measure the reliability/unreliability (the gain/loss) of the channels that she sensed; this can be done easily in practice. When converting the online HS game to OSP by using the graph G_{k,n,k_0} (see Section 7.1.3), this feedback correspond directly to the *semi-bandit feedback*. We then introduce an additional condition on how the hider/adversary assigns the losses on the locations in the online HS game:

($C1$) *At stage t, the adversary secretly assigns a loss $\boldsymbol{b}_t(j)$ to each location $j \in [k]$ (unknown to the learner). These losses are fixed throughout the n-search of the learner.*

Assuming that the learner knows that the adversary follows Condition ($C1$), from the feedback described above, she can deduce the following *side-observations*: within a stage, the loss at each location remains the same no matter when it is chosen among the n-search, i.e., knowing the loss of choosing location j as her i-th move, the learner knows all the loss if she chooses location j as her i'-th move for any $i' \neq i$. An example can be seen on the graph G_{k,n,k_0} corresponding to the case where $k = n = 3, k_0 = 1$, illustrated in Figure 7.2: the edges $1, 4, 6, 11$, and 13 represent that location 1 is chosen; thus, they mutually reveal each other. The semi-bandit feedback and side-observations as described above generate the observation graphs G^O_t. Under Condition ($C1$), for any $t \in [T]$, the observation graph G^t_O is *symmetric* and *does not satisfy* ($A1$); its the independence number is $\alpha_{\text{HS}} = k$.

Finally, we consider a relaxation of Condition ($C1$):

($C2$) *At stage t, the adversary assigns a loss $\boldsymbol{b}_t(j)$ on each location $j \in [k]$. For $i = 2,\dots,n$, after the learner chooses, say location j_i, as her i-th move, the adversary can observe that and change the losses $\boldsymbol{b}_t(j)$ for any location that has not been searched before by the learner,*[13] *i.e., she can change the losses $\boldsymbol{b}_t(j), \forall j \notin \{j_1,\dots,j_i\}$.*

By replacing Condition ($C1$) with Condition ($C2$), we can limit the side-observations of the learner: she can only deduce that if $i_1 < i_2$, the edges in G_{k,n,k_0} representing

[13] An interpretation is that by searching a location, the learner/seeker "discovers and secures" that location; therefore, the adversary/hider cannot change her assigned loss at that place.

choosing a location as the i_1-th move reveals the edges representing choosing that same location as the i_2-th move; but *not vice versa*. In this case, the observation graph G_t^O is non-symmetric; however, its independence number is still $\alpha_{\text{HS}} = k$ as in the HS games with Condition $(C1)$. We conclude by the following trivial proposition:

Proposition 8.3.2. *Any instance of the online HS game where the n-search of the learner satisfies the coherence constraint*[14] *and the losses are under Condition* $(C1)$ *(or Condition* $(C2)$*) can be cast as an* SOOSP.

Now, we can use the Exp3-OE algorithm in this game and observe that it only runs in at most $O(k^6n^3T)$ time (see our discussion on the running time of Exp3-OE in Section 8.2.1). Moreover, either with Conditions $(C1)$ or $(C2)$ we have the following corollary:

Corollary 8.3.3. *In the* SOOSP *corresponding to an online HS game with k locations and n-search under coherence constraint, the expected regret of the* Exp3-OE *is*

$$R_T \leq \tilde{O}(n\sqrt{T\alpha_{HS}\log(P)}) = \tilde{O}(\sqrt{Tn^3k}). \tag{8.4}$$

This result can be deduced directly from Corollary 8.2.3. Note that *the regret bound of* Exp3-OE in the HS game with Condition $(C1)$ (involving symmetric observation graphs) is slightly better[15] than that in the HS game with Condition $(C2)$. At a high-level, given the same scale on their inputs, from Corollary 8.3.1 and Corollary 8.3.3, we see that the expected regret's bounds of the Exp3-OE algorithm in the online semi-bandit CB game has the same order of magnitude as that of the HS game. This can be explained by the fact the independence numbers of the observation graphs in HS games are smaller than in CB games (by a multiplicative factor of n) but Assumption $(A1)$ is satisfied by the observation graphs of the CB games and not by the HS games (that improves the regret's guarantees in the HS games). Moreover, in the HS games with $(C1)$, the regret bounds of the Exp3-OE algorithm improves the bound of FPL-IX but they are still in the same order of the games' parameters (ignoring the logarithmic factors).[16] Finally, comparing with the OSMD algorithm, running in the HS games with Condition $(C1)$, the Exp3-OE has a better expected regret's guarantee if $O(n\log k_0) \leq k$; running in the HS games with Condition $(C2)$, it is better if $n \cdot \log k_0 \log(n^4k^5\sqrt{T}) \leq O(k)$. See Appendix E.7 for proof of this statement. Intuitively, the regret guarantees of Exp3-OE is better in the HS games where the total number of locations is sufficiently larger than the number of moves that the learner can make in each stage.

[14]Results on online HS games with other type of constraints can be obtained similarly.

[15]It is better by a multiplicative term of order $O\left(\sqrt{\log(nk)}\right)$ that is hidden in the $\tilde{O}$ notation in (8.4).

[16]More explicitly, in HS games with (C1), FPL-IX's regret is $O\left(\log(k^2n^2T)\sqrt{\log(k^2n)(k^2n^4+Cn^3kT)}\right)=\tilde{O}(Tn^3k)$ and Exp3-OE's regret is $O\left((3/2)\sqrt{n^3kT\log(k)}+\sqrt{nkT}\right)=\tilde{O}(Tn^3k)$ (similar results can be obtained for the HS games with Condition (C2)).

8.3.3 Exp3-OE in the Online Full-information CB Game

Finally, we consider the online CB game with the full-information feedback setting (see Section 7.1.1 for a formal definition). We first observe that this game can be cast directly as an OComb with full information in which the Hedge algorithm (proposed by Freund and Schapire (1997)) is an optimal algorithm, i.e., it provides a regret upper-bound matching the regret's lower-bound of this class of problems. However, the classical variant of Hedge runs inefficiently: its running time is polynomial in terms of in the number of actions, i.e., exponential in terms of the number of battlefields and players' budgets in the CB game.

On the other hand, we can also convert this game into an OSP with full-information feedback, i.e., after choosing a path, the learner observes the losses of *all* edges on the graph (see Section 7.1.2). There exist improved version of Hedge running in this instance of OSP where the weight pushing technique is applied to guarantee an efficient implementation (while providing optimal regret bounds). Therefore, at a high-level, there is no fundamental open question in studying the online CB game with full-information.

Despite this fact, for the sake of completeness, we show here how to apply our findings in the SOOSP model to the online CB game with full-information. Recall that our SOOSP model interpolates between the semi-bandit and the full-information feedback settings of OSP. Therefore, naturally, the online full-information CB game can also be cast as an SOOSP where all observation graphs are complete graphs and they have the independence number equal 1. Therefore, Exp3-OE can also be applied to this game. In this case, the term $q_t(e)$ in line 8 of Exp3-OE (see Algorithm 10)—i.e., the probability that an edge e is revealed at time t—is always equal to 1 due to the full-information feedback. Therefore, we have two trivial results:

> By setting $\beta = 0$ (i.e., in lines 8 and 9 of Algorithm 10, replace the estimated loss $\hat{\boldsymbol{\ell}}_t(e)$ by the real loss $\boldsymbol{\ell}_t(e)$ that is observed for any $e \in \mathcal{E}$) and $\gamma = \sqrt{n/T}$, from the proof of Theorem 8.2.1, we can deduce easily that Exp3-OE, running in the online full-information CB game, guarantees an expected regret:
>
> $$R_T \leq \mathcal{O}\left(\log(P)/\eta + \eta \cdot T\right) = \mathcal{O}(\sqrt{nT})$$
>
> Exp3-OE also runs efficiently in the online full-information CB game (in $\mathcal{O}(n^2 k^4 T)$ time) and the running time can even be further reduced by simply setting $q_t(e) = 1$ for any $e \in \mathcal{E}$, i.e., there is no need for Algorithm 11.

In fact, our Exp3-OE algorithm running with $\beta = 0$ and $q_t(e) = 1, \forall e$ can be considered as an efficient implementation (in the OSP) of the Hedge algorithm. As an unsurprising result, the regret upper-bound of Exp3-OE provided above matches exactly to the bounds obtained from applying Hedge to the OComb conversion of the online full-information CB game (that is optimal).

. . . .

Summary: In this chapter, we introduced a novel algorithm—the Exp3-OE algorithm—for the online shortest path problem with side-observations (SOOSP). Importantly, we designed Exp3-OE such that it is always efficiently implementable in SOOSP. Moreover, we proved that Exp3-OE provides regret guarantees matching to that of state-of-the-art algorithms in general cases of observation graphs and Exp3-OE's guarantees are better in several particular cases of interest. We applied our findings to several online resource allocation games that are cast into SOOSP problems including the online semi-bandit Colonel Blotto game and the online hide-and-seek game. We showed the benefits of using Exp3-OE in these games including notable improvements in both the running time and the regret guarantees.

Chapter 9

OSP with Bandit Feedback (OSPBand)—Applications to the Online Bandit CB Game

Some of the ideas and results presented in this chapter have previously appeared in our publication Vu, Loiseau, and Silva (2019b).[a] *The numerical experiments presented in this chapter have also appeared in this publication and the corresponding codes are available at* `https://github.com/dongquan11/BanditColonelBlotto`.

[a] A note on the terminology: the online shortest path problem, defined and studied in this thesis, is called the path planning problem in Vu, Loiseau, and Silva (2019b). These terms are often used interchangeably in the literature of bandit problems and online learning.

In this chapter, we turn our focus to a more restricted feedback setting of the online shortest path problem: the bandit case. Our motivation to study this problem comes from the online bandit CB game—an instance of online resource allocation games with combinatorial structures. In this variant of the CB game, at the end of each stage, the learner only observes the aggregate loss from all battlefields (without knowing precisely the loss in each battlefield nor the adversary's play). We presented its formal definition and discussed some motivational examples in Section 7.1. As we discussed in Chapter 7, any online CB bandit game can be cast into an OSP with bandit feedback (OSPBand). Therefore, with the aim of providing efficient algorithm that guarantees good payoffs in the online bandit CB game, we first study the OSPBand model.

State-of-the-art regret-minimization algorithms in OSPBand still have issues in implementation and there is still room for improvement in regret guarantees. Our objective of this chapter is twofold: (i) we design algorithms that improves regret guarantees in any instance of OSPBand while maintaining the efficiency in implementation, and (ii) we aim to apply these findings into the online bandit CB game to provide better learning policies for the learner. In particular, we focus on ComBand (defined in Section 2.2.2)—a standard algorithm of OComb with bandit feedback and aim to improve the regret guarantee and the running time of ComBand in OSPBand.

The outline and our contributions in this chapter are as follows: In Section 9.1, we review challenges in studying OSPBand and discuss several drawbacks of the ComBand algorithm. In Section 9.2, we propose a new algorithm, called Edge(μ), that is modified from ComBand in which the main upgrade is a quick computation of the involved co-occurrence matrix. Edge(μ) is always efficiently implemented in any arbitrary instance of OSPBand. On other hand, the regret's upper-bound provided by Edge(μ) depends directly on the input μ—the exploration distribution. Therefore, a question that naturally arises is to *efficiently* search for an exploration distribution that optimize the regret guarantee of Edge(μ) (and of ComBand). This, however, still remains an open question for medium and large-size instances (we discuss the involved challenges in Section 9.3). Without an optimal solution, state-of-the-art algorithms in OSPBand are often used with a simple exploration distribution; however, it does not provide a good regret guarantee in several cases. Therefore, we propose a fast method to compute an exploration distribution that can be used as the input of Edge(μ) to improve its regret guarantees. We conduct several numerical experiments to illustrate the improvements provided by Edge(μ) in the online bandit CB game, both in terms of the performance and the computation time; this is done in Section 9.4. Finally, in Section 9.5, we present an additional result: we construct instances of OSPBand and of the online CB bandit game that provide a regret lower-bound for any learning policy; we then compare this lower-bound with the regret upper-bound provided by Edge(μ).

9.1 Challenges in OSP with Bandit Feedback (OSPBand)

The model of OSP with bandit feedback (hereinafter, OSPBand) is an important instance of the OComb framework (under the bandit feedback setting, it is also called the combinatorial bandits)—see e.g., the discussion by Cesa-Bianchi and Lugosi (2012). However, the standard algorithms for OComb with bandit feedback and OSPBand have several drawbacks and there is still room for improvement. In this chapter, we focus on the ComBand algorithm—a standard algorithm in OComb with bandit feedback—that is also applicable to OSPBand. We refer the interested readers to Section 2.2.2 for a pseudo-code (Algorithm 2) and our detailed discussion on ComBand. Here, we briefly recall that for any instance of OComb with bandit feedback (and thus, any OSPBand), ComBand runs in exponential time in terms of the dimension of the action vectors (i.e., number of edges of the DAG in OSPBand) and it provides an expected-regret's guarantee given in Proposition 2.2.5. In particular, given an online bandit CB game where the learner allocates k troops across n battlefields at each stage, these results imply that if we apply directly ComBand into the OSPBand corresponding to this game (on the graph $G_{k,n}$ with $E = \mathcal{O}(nk^2)$ edges and $P = \mathcal{O}(2^{\min\{n-1,k\}})$ paths), the running time of ComBand is an exponential number in terms of n and k. Moreover, it provides an expected regret satisfying:

$$R_T \leq 2\sqrt{\left[\frac{2n}{E\lambda^*[M(\mu)]}+1\right] TE\log P} = \mathcal{O}\left(\sqrt{\frac{Tn\min\{n-1,k\}}{\lambda^*[M(\mu)]}+Tnk^2\min\{n-1,k\}}\right). \quad (9.1)$$

Here, we recall several notations: μ is a fixed distribution on the actions set (i.e., the paths set of OSPBand)—chosen as input of ComBand; $M(\mu)$ denotes the matrix $\mathbb{E}_{\boldsymbol{p}\sim\mu}\left[\boldsymbol{p}\boldsymbol{p}^\top\right]$ and $\lambda^*[M(\mu)]$ denotes the *smallest nonzero eigenvalue* of $M(\mu)$. From now on, we use the notation ComBand(μ) when it is needed to emphasize that ComBand takes a distribution μ as an input.

We encounter two key challenges in applying ComBand(μ) to OSPBand:

Challenge 1: Optimizing the Exploration Distribution. ComBand(μ) mixes an exploitation procedure (sampling based on weights that are updated according to an unbiased loss estimator) with an *exploration distribution* on the actions set (i.e., μ). In OSPBand, this guarantees that at each stage, each path $\boldsymbol{p}$ will be chosen with a probability at least $\mu(\boldsymbol{p})$.[1] The regret bound given by ComBand algorithm depends on the choice of the exploration distribution μ via the parameter $\lambda^*[M(\mu)]$. It is optimal to choose μ such that $\lambda^*[M(\mu)]$ is maximal. For OSPBand (and also for the online bandit CB game), an efficient method to find an optimal exploration distribution remains unknown. Several exploration distributions used in the literature such as the uniform distribution on a barycentric spanner of the action set (see Dani et al. (2008)) and the John's exploration (see Bubeck, Cesa-Bianchi, and Kakade (2012)) may derive good regret bounds but are either unavailable or cannot be efficiently constructed in a generic instance of OSPBand (see Section 7.3 for more details). On the other hand, the simplest exploration distribution (that is the most well-used in the literature) is the uniform distribution on the path set; however, in Cesa-Bianchi and Lugosi (2012), it is proven that this choice can lead to very bad regret guarantees for several cases of OSPBand: it can be exponential in terms of the number of edges. We aim to have a fast procedure finding distributions that provide better performance guarantees for ComBand-type algorithms when applying them to OSPBand.

Challenge 2: Implementation Issue. To obtain an implementation of ComBand(μ), it is needed at each stage to compute a co-occurrence matrix (see Section 9.2.1 for a definition)—which is basically a sum of an exponential number of semi positive definite matrices. The problem of finding an efficient and implementable procedure for this computation has not been solved completely. In the case of OSPBand, Sakaue et al. (2018) propose an algorithm, based on an extension of the weight-pushing technique,[2] that efficiently computes the co-occurrence matrix corresponding to the path sampling distribution for a particular instance of ComBand (where μ is an uniform distribution). However, this algorithm has a redundancy in representation (involving 5 sub-algorithms) and is still non-trivial to be implemented; moreover, it lacks the explicit analysis of the use of more generic exploration distributions. We desire a simpler and more general representation of this efficient algorithm.

In the following sections, we sequentially investigate the two challenges mentioned

[1] This is often referred to as the *explicit exploration scheme* to distinguish with the implicit exploration, i.e., guaranteeing a lower-bound of the probability of being chosen of each path by controlling the estimated losses (see e.g., the use of the parameter β in the Exp3-OE algorithm presented in Section 8.2).

[2] This is a dynamic programming techniques that are used to efficiently sample paths in Exp3-type algorithms; we have reviewed this technique in Section 2.2.3.

above: Challenge 2 is addressed in Section 9.2 and Challenge 1 will be considered in Section 9.3. Before doing that, we note that there exist other variants of ComBand that can also be applied to OSPBand (see Section 7.3 for a review on these algorithms); e.g., the CombEXP algorithm, proposed by Combes et al. (2015), improves the complexity of ComBand. However, OSPBand is not explicitly considered in Combes et al. (2015) and it remains an open question whether any arbitrary instance of the OSPBand satisfies the condition such that CombEXP can be efficiently implemented. Therefore, ComBand is still the state-of-the-art algorithm for our considering problems. Moreover, CombEXP also uses the uniform exploration distribution that is sub-optimal in OSPBand (see also Cesa-Bianchi and Lugosi (2012)); thus, the challenge in finding better exploration distributions is relevant.

9.2 Edge—An Efficient Algorithm for OSPBand

In this section, we present our solution for Challenge 2 described in the previous section; that is, we provide a variant of ComBand, with a simple representation, that runs efficiently in OSPBand. The results presented in this section are applicable to any instance of OSPBand with an arbitrary DAG G (with a source vertex s and a destination vertex d); we also use the notations $\mathcal{E}$ and $\mathcal{P}$ to denote the edges set and the set of all paths of G starting from s and ending at d, and denote $E := |\mathcal{E}|$ and $P = |\mathcal{P}|$. As a convention, henceforth, we often use μ to denote an arbitrary distribution on the set $\mathcal{P}$ such that its support spans $\mathcal{P}$.

First, recall that the *inefficiency* of the standard implementation of ComBand(μ) (see Algorithm 2 in Section 2.2.2) comes from three steps that are executed at each stage: the *weight-updating step* (line 9—Algorithm 2) that computes the weight embedded by ComBand(μ) on each path, the *sampling step* (line 5—Algorithm 2) that samples a path on G based on a distribution mixing between sampling from these P weights (with a normalization) and the exploration distribution μ, and the *computing the co-occurrence matrix step* (line 7—Algorithm 2) that is essential for estimating the losses of the paths.

As previously discussed, the use of weight pushing in improving the complexity of the weight-updating step and the sampling step of ComBand(μ) and other Exp3-type algorithms (when being applied to variants of OSP) can be found in several works in the literature, e.g., György et al. (2007) and Takimoto and Warmuth (2003). We have reviewed the weight pushing technique in Section 2.2.3 as Algorithm 3 and Algorithm 4. On the other hand, Sakaue et al. (2018) propose an application of weight pushing to compute the co-occurrence matrix in $\mathcal{O}(E^2T)$ time; particularly for ComBand in OSPBand with the Zero-suppressed Binary Decision Diagrams. This computation requires 5 sub-algorithms that involves heavy notations and unnecessary complexity for our setting. Moreover, the algorithm of Sakaue et al. (2018) is only presented in the case where the exploration distribution is chosen to be the uniform distribution; it is not discussed explicitly how it can be used effectively if ComBand has other distributions as inputs. In the next section, we use ideas similar to that of Sakaue et al. (2018) to derive a procedure to compute the co-occurrence matrices efficiently; our computation has a simpler and more general representation. Furthermore, this representation is also convenient when using co-occurrence matrices as tools for us to obtain other results (e.g., in improving the exploration distribution).

9.2.1 Co-occurrence Matrices Computation

Given a DAG G, for any distribution μ on the set $\mathcal{P}$ (here, $\mu(\boldsymbol{p})$ denotes the probability of sampling $\boldsymbol{p} \in \mathcal{P}$ from μ), we consider a matrix of the following form:

$$M(\mu) = \mathbb{E}_{\boldsymbol{p} \sim \mu(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top].$$

Hereinafter, we call $M(\mu)$ the *co-occurrence matrix corresponding to the distribution* μ. This term is adopted from Cesa-Bianchi and Lugosi (2012); intuitively, each entry $M(\mu)_{e_1,e_2}$ of this matrix ($e_1, e_2 \in \mathcal{E}$) is equal to the probability that a path sampled from μ contains both edges e_1 and e_2 (hence, a co-occurrence of e_1 and e_2). Note that $M(\mu)$ also depends on the structure of the graph under consideration; however, to lighten the notation, we do not explicitly include this into the notation of $M(\mu)$ (the involved graph is implicitly declared via μ). More importantly, when μ has a support that spans $\mathcal{P}$ (which is the condition to use μ as the exploration distribution of COMBAND), a direct computation of this matrix might involve a sum of $\Omega(P)$ terms, this is inefficient for our purpose. Therefore, we want to have a procedure such that for any given distribution μ on $\mathcal{P}$, $M(\mu)$ can be computed in polynomial time in terms of E. This remains an open question; however, we can provide solution for the distributions satisfying a special condition as follows:

Condition (WP): *A distribution μ on $\mathcal{P}$ is said to satisfy Condition (WP) if and only if there exists a set of weights $\tilde{w} := \{\tilde{w}(e) \geq 0, e \in \mathcal{E}\}$ such that for any path $\boldsymbol{p} \in \mathcal{P}$, we have $\mu(\boldsymbol{p}) = \prod\limits_{e \in \boldsymbol{p}} \tilde{w}(e) / \sum\limits_{\boldsymbol{p}' \in \mathcal{P}} (\prod\limits_{e' \in \boldsymbol{p}'} \tilde{w}(e'))$.*

Intuitively, if μ satisfies Condition (WP), we can efficiently sample a path from μ by weight pushing and μ can be represented by a E-dimensional (instead of P-dimensional) vector. Note that the uniform distribution on $\mathcal{P}$ (used by most of works in the literature) satisfies Condition (WP) (e.g., with the weights $w(e) = 1, \forall e$). Moreover, it might happens that one distribution can satisfy (WP) with different sets of weights; however, each set of weights $\tilde{w}$ only determines a unique distribution $\mu_{\tilde{w}}$ satisfying Condition (WP). Henceforth, we use the notation $\mu_{\tilde{w}}$ *to denote a distribution satisfying Condition (WP)* with the set of weights $\tilde{w} := \{\tilde{w}(e) \geq 0, e \in \mathcal{E}\}$. We also can observe trivially that if μ satisfies Condition (WP) with a set of positive weights, then it has the full support on $\mathcal{P}$.

Now, let us recall two notations: $\mathcal{P}_{u,v}$ denotes the set of all paths from vertex u to vertex v and $e_{[u,v]}$ denotes the edge going from vertex u to vertex v (if it exists). From Section 2.2.3, we know that for any set of weights $\{\tilde{w}(e), e \in \mathcal{E}\}$, it takes $O(E^2)$ time to compute all the terms $H(u,v) := \sum_{\boldsymbol{p} \in \mathcal{P}_{u,v}} \prod_{e \in \boldsymbol{p}} \tilde{w}(e)$ for any pair of vertices u, v in the graph G. This can be done by the WP algorithm (Algorithm 3). Based on these terms, we propose the CMat Algorithm (its pseudo code is given in Algorithm 12) that computes efficiently $M(\mu_{\tilde{w}})$.

Algorithm 12: The CMat Algorithm (for computing co-occurrence matrix $M(\mu_{\tilde{w}})$).

Input: $G = (\mathcal{V}, \mathcal{E})$ with source vertex s, destination vertes d; $\{\tilde{w}(e), \forall e \in \mathcal{E}\}$.
Output: The matrix $M(\mu_{\tilde{w}})$.

1 Compute $H(u,v) := \sum_{\boldsymbol{p} \in \mathcal{P}_{(u,v)}} \prod_{e \in \boldsymbol{p}} \tilde{w}(e), \forall u, v \in \mathcal{V}$ by WP algorithm (Algorithm 3).
2 **for** $e_1 = e_{[u_1,v_1]} \in \mathcal{E}$ **do**
3 $\quad M(\mu_{\tilde{w}})_{e_1,e_1} = \frac{H(s,u_1)\tilde{w}(e_1)H(v_1,d)}{H(s,d)}$.
4 $\quad$ **for** $e_2 = e_{[u_2,v_2]} \in \mathcal{E}, e_2 > e_1$ **do**
5 $\quad\quad M(\mu_{\tilde{w}})_{e_1,e_2} = \frac{H(s,u_1)\tilde{w}(e_1)H(v_1,u_2)\tilde{w}(e_2)H(v_2,d)}{H(s,d)}$.
6 **for** $e_1, e_2 \in \mathcal{E}, e_2 < e_1$ **do** $M(\mu_{\tilde{w}})_{e_1,e_2} = M(\mu_{\tilde{w}})_{e_2,e_1}$.

This algorithm runs in $O(E^2)$ time. The main intuition of Algorithm 12 is as follows: First, the probability that a path sampled from $\mu_{\tilde{w}}$ contains an edge e_1, i.e., the entry $M(\mu_{\tilde{w}})_{e_1,e_1}$, can be rewritten as

$$\sum_{\boldsymbol{p} \ni e_1} \mu_{\tilde{w}}(\boldsymbol{p}) = \frac{\sum_{\boldsymbol{p} \ni e_1} \prod_{e \in p} \tilde{w}(e)}{\sum_{\boldsymbol{p}' \in \mathcal{P}} \left(\prod_{e' \in \boldsymbol{p}'} \tilde{w}(e')\right)}. \tag{9.2}$$

On the other hand, by definition, $H(u,v)$ can be interpreted as the sum of the products of weights along all paths from u to v; note also that $H(s,d)$ is precisely the denominator in the right-hand-side of (9.2). Moreover, in G, if a path $\boldsymbol{p} \in \mathcal{P}$ contains an edge $e_1 = e_{[u_1,v_1]}$, then $\boldsymbol{p}$ also has to contain a path from node s to node u_1 and a path from node v_1 to node d. Hence, the numerator in the computation of $M(\mu_{\tilde{w}})_{e_1,e_1}$ in line 3 of Algorithm 12 is equal to the numerator in the right-hand-side of (9.2). Similarly, if a path $\boldsymbol{p}$ simultaneously contains the edges $e_1 = e_{[u_1,v_1]}$ and $e_2 = e_{[u_2,v_2]}$, then $\boldsymbol{p}$ also contains a path from node s to node u_1, a path from node v_1 to node u_2 and a path from node v_2 to node d. Therefore, the entry $M(\mu_{\tilde{w}})_{e_1,e_2}$ for $e_1 \neq e_2$ can be computed by line 5 of Algorithm 12. Finally, by definition, $M(\mu_{\tilde{w}})$ is a symmetric matrix; therefore, we can have the relation in line 6 of Algorithm 12.

9.2.2 The Edge(μ) Algorithm

In this section, we combine the techniques presented in the previous sections to create a modified variant of ComBand, called Edge, that runs efficiently in any OSPBand with an arbitrary DAG G. Unlike ComBand, the Edge algorithm works on edges instead of paths (hence, the name); in particular, at each stage t, it keeps a weight $w_t(e)$ on each edge $e \in \mathcal{E}$ (called the *edges weights*). We denote $w_t(\boldsymbol{p}) = \prod_{e \in \boldsymbol{p}} w_t(e), \forall \boldsymbol{p} \in \mathcal{P}$; we call these the *paths weights*. Let $\mu_{\tilde{w}}$ be a distribution on $\mathcal{P}$ satisfying Condition (WP) with a set of weights $\tilde{w}$. Then, for any $\boldsymbol{p} \in \mathcal{P}$, we introduce two new terms:

$$\nu_t(\boldsymbol{p}) := \frac{w_t(\boldsymbol{p})}{\sum_{\boldsymbol{p}' \in \mathcal{P}} w_t(\boldsymbol{p}')} = \frac{\prod_{e \in \boldsymbol{p}} w_t(e)}{\sum_{\boldsymbol{p}' \in \mathcal{P}} \left(\prod_{e' \in \boldsymbol{p}'} w_t(e')\right)} \quad \text{and } x_t(\boldsymbol{p}) = (1-\gamma)\nu_t(\boldsymbol{p}) + \gamma \mu_{\tilde{w}}(\boldsymbol{p}). \tag{9.3}$$

Accordingly, we define $M(\nu_t) := \mathbb{E}_{\boldsymbol{p} \sim \nu_t(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top]$ and $M(\mu_{\tilde{w}}) = \mathbb{E}_{\boldsymbol{p} \sim \mu_{\tilde{w}}(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top]$. We can use Algorithm 12 (with inputs $\{w_t(e), e \in \mathcal{E}\}$ and $\{\tilde{w}(e), e \in \mathcal{E}\}$) to compute these two matrices in $O(E^2)$ time (note that trivially, ν_t also satisfies Condition (WP)). On the other hand, due to (9.3), the co-occurrence corresponding to x_t, called C_t, satisfies

$$C_t := \mathbb{E}_{\boldsymbol{p} \sim x_t(\boldsymbol{p})}[\boldsymbol{p}\boldsymbol{p}^\top] = (1-\gamma)M(\nu_t) + \gamma M(\mu_{\tilde{w}}). \tag{9.4}$$

Therefore, we conclude that we can also efficiently compute C_t by Algorithm 12 in $O(E^2)$ time.

Now, we present a pseudo code of Edge in Algorithm 13, written with the input $\mu_{\tilde{w}}$. As for the case of ComBand(μ), we will use the notation Edge(μ) when it is needed to emphasize that this algorithm uses μ as an exploration distribution. In Edge, at each stage, a path is sampled efficiently (in $O(E^2)$ time) from the distribution x_t (lines 4, 5 and 6 of Algorithm 13). Moreover, in line 9, the co-occurrence matrix C_t is also computed efficiently in $O(E^2)$ time. Note that Edge also allows us more freedom to choose the exploration distribution (unlike the algorithm from Sakaue et al. (2018)). Moreover, as long as this distribution satisfies Condition (WP), the variant of the Edge algorithm taking $\mu_{\tilde{w}}$ as an input and using it as an exploration distribution runs efficiently in $O(E^2T)$ time. Note that a discussion on the choices of these exploration distributions will be given in Section 9.3.

Algorithm 13: The Edge($\mu_{\tilde{w}}$) Algorithm for OSPBand.

Input: $T \in \mathbb{N}, \gamma \in [0,1], \eta > 0$, graph G, distribution $\mu_{\tilde{w}}$ on $\mathcal{P}$.

1 $\forall e \in \mathcal{E}, w_1(e) := 1$.
2 **for** $t = 1, 2, \ldots, T$ **do**
3 Loss vector $\boldsymbol{\ell}_t \in [0,1]^E$ is chosen adversarially (unobserved).
4 Sample β from Bernoulli distribution $\mathcal{B}(\gamma)$.
5 **if** $\beta = 0$ **then** sample a path $\tilde{\boldsymbol{p}}_t \sim \nu_t(\tilde{\boldsymbol{p}}_t)$ using WPS Algorithm (Algorithm 4) with $\{w_t(e), e \in \mathcal{E}\}$ as inputs.
6 **else** sample a path $\tilde{\boldsymbol{p}}_t \sim \mu_{\tilde{w}}(\tilde{\boldsymbol{p}}_t)$ using WPS Algorithm with $\{\tilde{w}(e), e \in \mathcal{E}\}$ as inputs.
7 Suffer and observe the loss $L(\boldsymbol{p}_t) = (\boldsymbol{\ell}_t)^\top \tilde{\boldsymbol{p}}_t \leq 1$.
8 Compute $C_t := \mathbb{E}_{\boldsymbol{p} \sim x_t}[\boldsymbol{p}\boldsymbol{p}^\top]$ based on (9.4) and Algorithm 12.
9 Estimate loss $\hat{\boldsymbol{\ell}}_t = \left(\boldsymbol{\ell}_t(\tilde{\boldsymbol{p}}_t)^\top\right) C_t^{-1} \tilde{\boldsymbol{p}}_t$.
10 $\forall e \in \mathcal{E}, w_{t+1}(e) := w_t(e) \cdot e^{-\eta \hat{\boldsymbol{\ell}}_t(e)}$.

We conclude this section with the following proposition (note that theses results work with OSPBand instances involving non-oblivious adversaries).

Proposition 9.2.1. *In any instance of* OSPBAND *(with a time horizon T) on an arbitrary DAG G where there are P paths, E edges and the length of the longest paths is n,* EDGE$(\mu_{\tilde{w}})$ *runs efficiently in $O(E^2T)$ time; moreover, with the same choices of γ and η, the expected regret of* EDGE$(\mu_{\tilde{w}})$ *is equal to that of* COMBAND$(\mu_{\tilde{w}})$ *and satisfies*

$$R_T \leq 2\sqrt{\left[\frac{2n}{E\lambda^*[M(\mu_{\tilde{w}})]}+1\right]TE\log(P)}.$$

In particular, in the OSPBAND *corresponding to the online bandit CB game with k troops and n battlefields,* EDGE$(\mu_{\tilde{w}})$ *runs in $O(n^2k^4T)$;*[3] *the expected regret's bound provided by* EDGE$(\mu_{\tilde{w}})$ *in this case is:*

$$R_T \leq O\left(\sqrt{\frac{Tn\cdot\min\{n-1,k\}}{\lambda^*[M(\mu_{\tilde{w}})]}+Tnk^2\min\{n-1,k\}}\right).$$

9.3 Improving Exploration Distributions Used in the EDGE Algorithm

In this section, we study the problem of selecting a distribution to input into EDGE and use it as the exploration distribution such that EDGE guarantees a good regret bound. Now, recall that the notation $\lambda^*[M]$ denotes the smallest non-zero eigenvalue of a matrix M. From results presented in Proposition 9.2.1, the larger $\lambda^*[M(\mu)]$ is, the smaller the upper-bound regret guaranteeing by EDGE(μ) becomes. Therefore, given an arbitrary instance of OSPBAND (or an instance of the online bandit CB game), our aim is to find a distribution μ such that $\lambda^*[M(\mu)]$ is as large as possible (while we can still guarantee that EDGE(μ) runs efficiently).

Formally, let us label the paths in $\mathcal{P}$ by $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_P$, we consider an eigenvalue-optimization problem as follows (its search space is P-dimensional):

$$\text{maximize} \quad \lambda^*\left[\sum_{i=1}^{P} x_i \cdot \left[\boldsymbol{p}_i\boldsymbol{p}_i^\top\right]\right] \tag{9.5}$$

$$\text{subject to} \quad \boldsymbol{x}\in[0,1]^P, \sum_{i=1}^{P} x_i = 1, \tag{9.6}$$

$$\{\boldsymbol{p}_i : x_i > 0\} \text{ spans } \mathcal{P}. \tag{9.7}$$

A popular choice of the literature (e.g., György et al. (2007) and Sakaue et al. (2018)) on the exploration distribution when applying variants of COMBAND in OSPBAND is the uniform distribution on the path set, denoted μ_{uni}. Trivially, μ_{uni} satisfies (9.6)-(9.7) and also Condition (WP) (i.e., EDGE(μ_{uni}) is efficiently implementable). However, there exists an instance[4] of OSPBAND such that μ_{uni} yields the eigenvalue $\lambda^*[M(\mu_{\text{uni}})]$ that may be of order $\Omega(P^{-1})$, which implies that EDGE(μ_{uni}) (and COMBAND(μ_{uni})) can only guarantee a regret upper-bound that is exponentially large in terms of the number of edges. In spite of this, due to its "popularity" in the literature, we often use μ_{uni} as a benchmark for our findings in this section.

[3] This is in contrast with *COMBAND*(μ) that runs in $O(\exp(n)T)$.

[4] An example of such case is presented by Cesa-Bianchi and Lugosi (2012).

9.3.1 Optimizing Exploration Distributions By Semi-Definite Programming

To find an optimal exploration distribution to use in EDGE (and COMBAND) for OSPBAND, we want to solve the optimization problem (9.5)-(9.7) whose objective function relates to the eigenvalues of matrices. One natural approach in solving (9.5)-(9.7) is to cast it into a semi-definite programming problem (SDP). This approach is also suggested by Cesa-Bianchi and Lugosi (2012). A formulation of an SDP that is equivalent to (9.5)-(9.7) can be found in Appendix F.1. In principle, given a graph G, this SDP can be solved exactly to find a distribution μ on the corresponding paths set that maximizes $\lambda^*[M(\mu)]$. However, in practice, this SDP formulation still cannot be solved efficiently due to the fact that the feasible set still has dimension P and that it contains a constraint relating to a summation of P terms. In our simulation, standard SDP solvers[5] take a long running time to solve this SDP problem even with small instances and they easily run into computationally memory issues with moderate instances. In the next section, we change the perspective and propose a fast method that can quickly find a "good" feasible solution of (9.5)-(9.7) , although it might not be the optimal solution.

9.3.2 Derivative-free Optimization and Change of Search Space

Given an instance of OSPBAND on a graph G, the challenge now is to find a fast method providing an exploration distribution μ to be used in EDGE(μ) that guarantees a sufficiently good regret-bound. Moreover, it is desired to be able to efficiently sample a path from μ (line 6 of Algorithm 13) and to efficiently compute the matrix $M(\mu)$ in order to compute C_t (line 8 of Algorithm 13). Note that a sufficient condition for this is that μ satisfies Condition (WP).

To efficiently find such a distribution, first, we note that the primary issue with the problem (9.5)-(9.7) is the dimension of its search-space (i.e., the set of points satisfying (9.6)-(9.7))—it is an exponential number in terms of the number of edges of G. To mitigate this issue, we reformulate to reduce the dimension of the search space. We consider the following problem whose search space is only E-dimensional:

$$\max_{\tilde{w} \in (0,\infty)^E} \lambda^*(M(\mu_{\tilde{w}})). \tag{9.8}$$

Here, given $\tilde{w} \in (0,\infty)^E$, recall that the notation $\mu_{\tilde{w}}$ denotes a distribution on the paths set such that $\mu_{\tilde{w}}(\boldsymbol{p}) = \prod_{e \in \boldsymbol{p}} \tilde{w}(e) / \sum_{\boldsymbol{p}' \in \mathcal{P}} (\prod_{e' \in \boldsymbol{p}'} \tilde{w}(e'))$ for any $\boldsymbol{p} \in \mathcal{P}$; in other words, it satisfies Condition (WP) with the set of weights $\tilde{w} = \{\tilde{w}(e) > 0, e \in \mathcal{E}\}$. This choice of $\tilde{w}$ also guarantees that $\mu_{\tilde{w}}$ has a full support on $\mathcal{P}$ (i.e., $\mu_{\tilde{w}}$ satisfies (9.7)). More importantly, for each feasible solution of (9.8), say $\boldsymbol{w}^*$, we can construct a corresponding feasible solution of (9.5)-(9.7); additionally, the objective function of (9.8) at $\boldsymbol{w}^*$ equals to that of (9.5) at $x_i := \mu_{\boldsymbol{w}^*}(\boldsymbol{p}_i), i \in \{1, 2, \dots, P\}$. The construction of $\mu_{\boldsymbol{w}^*}$ is actually in $O(P)$ time, but we do not need to explicitly do so in order to run EDGE algorithm with $\mu_{\boldsymbol{w}^*}$. Instead, since $\mu_{\boldsymbol{w}^*}$ is guaranteed to satisfy Condition (WP), we can use the WP

[5] CVXOPT solver, available at *https://cvxopt.org/* and Mosek solver *https://www.mosek.com/*, both use primal-dual interior points methods.

Algorithm to sample efficiently a path from μ_{w^*} and use Algorithm 12 to compute efficiently $M(\mu_{w^*})$. Therefore, we can solve (9.8) to (implicitly) find an exploration distribution and use it efficiently in EDGE.

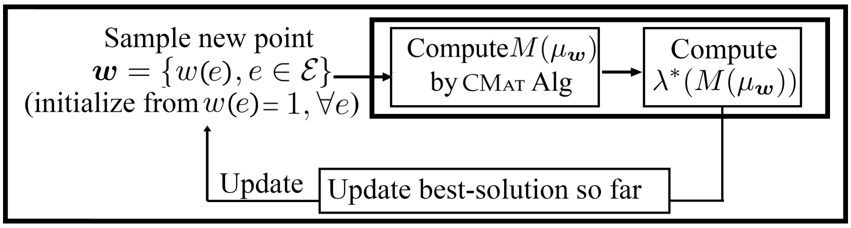


Figure 9.1: Diagram illustrating the derivative-free optimization for improving exploration distributions of the EDGE algorithm.

Although (9.8) reduces significantly the dimension of the search space compared to (9.5)-(9.6), this formulation loses the structure that allows us to apply standard convex optimization algorithms.[6] Therefore, in this section, we use a *derivative-free* algorithm to heuristically solve (9.8). Despite the fact that the solution found by this method may not be optimal, we can still guarantee that this solution is at least as good as the uniform distribution that is often used in the state-of-the-art algorithms (we initialize our algorithm with μ_{uni}). Moreover, although the search space in (9.8) may not cover the whole search space in (9.5)-(9.6), the solution found in (9.8) (which might correspond to a sub-optimal for (9.5)-(9.6)) is guaranteed to be efficiently embedded with EDGE; on the other hand, even if we found an optimal solution of (9.5)-(9.6), it does not guarantee to be efficiently usable in EDGE (nor in COMBAND). A diagram explaining the intuition of our method to solve (9.8) can be found in Figure 9.1. We can use any derivative-free optimization solver that goes with specific strategies of sampling new points and justifying the current-best solution.

We denote by μ_{free} the distribution corresponding to the solution of (9.8) found by our derivative-free method and note that $\lambda^*(M(\mu_{\text{free}})) \geq \lambda^*(M(\mu_{\text{uni}}))$.[7] Finally, as a trivial corollary of Proposition 9.2.1, we have:

Proposition 9.3.1. *In any instance of* OSPBAND *on an arbitrary DAG G,* EDGE(μ_{free}) *runs efficiently in $\mathcal{O}(E^2T)$ time and guarantees an expected regret*

$$R_T \leq 2\sqrt{\left[\frac{2n}{E\lambda^*[M(\mu_{free})]}+1\right] TE\log(P)}.$$

In particular, in the OSPBAND *corresponding to the online bandit CB game with k troops and n battlefields,* EDGE(μ_{free}) *also runs in $\mathcal{O}(n^2k^4T)$ and guarantees an expected regret*

$$R_T \leq \mathcal{O}\left(\sqrt{\frac{Tn\cdot\min\{n-1,k\}}{\lambda^*[M(\mu_{free})]} + Tnk^2\min\{n-1,k\}}\right).$$

[6]The function giving the smallest non-zero eigenvalue of a matrix $M(\mu_{\tilde{w}})$ from an input $\tilde{w}$ is not known to be convex or concave.

[7]This is due to the fact that we take $w(e)=1, \forall e \in \mathcal{E}$ (corresponding to μ_{uni}) as the initialization point.

9.4 Numerical Evaluation

We conduct several experiments to evaluate the performance of Edge and measure the effect of optimizing the exploration distribution.[8] In these experiments, without loss of generality, a learner, having k troops at each stage, plays repeatedly a discrete CB game on n battlefields against a single adversary who has k_A troops (unknown to the learner). We define a special adversary, called the *extreme-strong adversary*: an adversary having $k_A = (n-1)(k+1)+(k-1)$ troops, she "blocks" $n-1$ battlefields (each has a value equal to $\varepsilon/(n-1)$ in any stage) by allocating $k+1$ troops to them; and allocates $k-1$ troops to a certain battlefield i with value $b_t(i) = 1-\varepsilon$ (unknown to the learner). In this case, the losses on all paths are always 1 except for the single path representing that the learner allocates all k troops to battlefield i; this path yields the loss ε. We choose this adversary to follow an example in Cesa-Bianchi and Lugosi (2012) illustrating why μ_{uni} fails to guarantee a good regret bound in OSP. The algorithms need to "explore" the low-loss path as soon as possible to reduce the regret.

We use the Zoopt solver[9] (see Y.-R. Liu et al. (2017)) embedded with the sRACOS algorithm (Hu et al. (2017)) as the derivative-free optimization solver to heuristically solve (9.8)—its output is called μ_{free}. Our experiments run on an Intel Core i5-7300U CPU@ 2.60GHz and 8.00GB RAM. Each instance is run 5 times and the average results are reported.

In the first experiment, we compare the running time between ComBand and Edge and the results confirm that ComBand takes exponential time while Edge runs in polynomial time in terms of k and n; these results are reported in Figure 9.2 (the numbers of edges and paths in the corresponding $G_{k,n}$ are also reported for the sake of comparison).

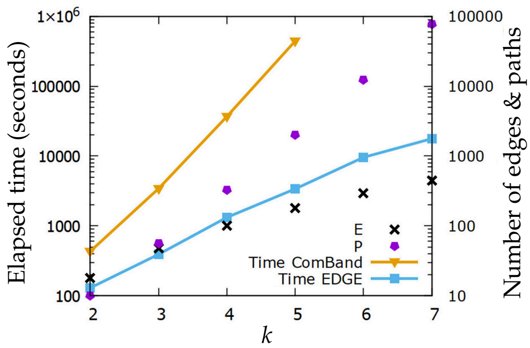
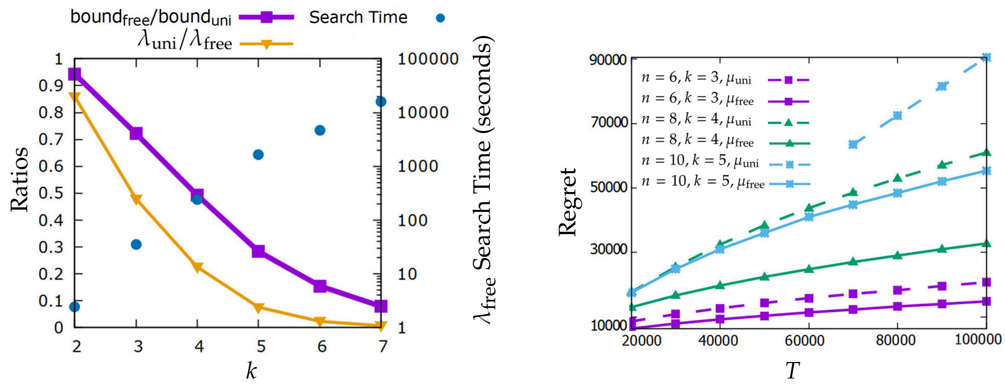


Figure 9.2: ComBand(μ_{free}) vs Edge(μ_{free}); $n = 2k$, $T = 40000$ fixed.

Next, we compare the performance of Edge when it uses μ_{uni} and μ_{free} as the exploration distribution. Figure 9.3(a) (y-axes is drawn with log-scale) illustrates the

[8] Our code is given at *https://github.com/dongquan11/BanditColonelBlotto*.

[9] Available at *https://zoopt.readthedocs.io/en/latest/*. We run it in 100E iterations; this stopping criterion is recommended by Hu et al. (2017); moreover, this criterion is enough to solve (9.8) optimally in our experiments with small instances ($k, n \leq 3$).



(a) $n = 2k$, $T = 40000$ fixed.

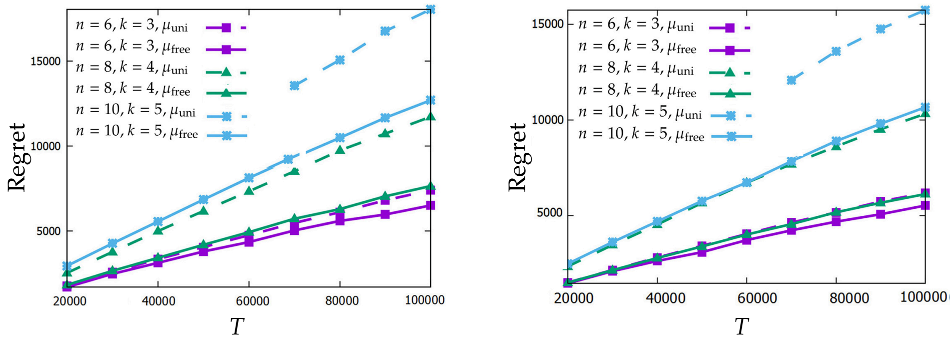
(b) The actual regrets.

Figure 9.3: Performances evaluation of Edge(μ_{uni}) and Edge(μ_{free}) in the online bandit CB game with the extreme-strong adversary.

trade-off between the time spent to find μ_{free} and the improvement in the eigenvalues and the upper-bounds predicted by Proposition 9.3.1. Note that the smaller the ratios bound$_{free}$/bound$_{uni}$ and $\lambda_{uni}/\lambda_{free}$ are, the more improvement that Edge(μ_{free}) provides compared to Edge(μ_{uni}). Finally, we compare the performance of Edge(μ_{uni}) and Edge(μ_{free}) by their actual regrets (see Figure 9.3(b)). Note that to efficiently compute the best hindsight loss (it is non-trivial), we apply Algorithm 7 (see Section 5.3.1) that finds the best response against a set of allocations of the adversary. We observe that the actual regret of Edge(μ_{free}) is better than Edge(μ_{uni}); as k increases, the difference between these regrets also increases. For example, for instance $k = 3, n = 6$ and $T = 10^5$, the ratio (Regret$_{uni}$ − Regret$_{free}$)/Regret$_{uni}$ equals 28% while this ratio of instance $k = 5$, $n = 10$, $T = 10^5$ is 38%. Note that Edge(μ_{free}) can run with larger instances (in k, n) but we choose not to report here since Edge(μ_{uni}) is unavailable in these instances (it requires extremely large T).[10]

Besides the extreme-strong adversary, we also consider two other adversary's strategies: the *uniform-adversary* (resp. the *battlefields-wise adversary*) who at each stage t repeatedly draws a battlefield by the uniform distribution (resp. draws a battlefield i with a probability $b_t(i)/\sum_{j\in[n]} b_t(j)$) then allocates one troop to that battlefield until she runs out of troops (the budget of the adversary at each stage is k_A). Note that for this experiment, the battlefields' values $b_t(i)$ are generated uniformly from $[0, 8]$ then they are normalized to guarantee that $\sum_{i\in[n]} b_t(i) = 1$. For each instance with different parameters k, n and adversary's strategies, we run each algorithm Edge(μ_{uni}) and Edge(μ_{free}) 5 times and the average results of their actual regret are reported in Figure 9.4. We notice that, at a high-level, the results from these cases are similar to that of the extreme-strong adversary case.

[10]Both ComBand(μ) and Edge(μ) require that $\gamma \leq 1$ to run; therefore, the regret bounds given in Proposition 9.2.1 and Proposition 2.2.5 can only be obtained if $T \geq \left[n \log(P)\right] / \left[(\lambda^*[M(\mu)])^2 \left(\frac{E}{n} + \frac{2}{\lambda^*[M(\mu)]}\right)\right]$ (parameters tuned by Cesa-Bianchi and Lugosi (2012)). When μ_{uni} is very small, it requires extremely large time horizon T to run Edge(μ_{uni}) that is impractical.



(a) Against uniform-adversary (b) Against battlefield-wise adversary

Figure 9.4: Actual regrets of Edge(μ_{uni}) and Edge(μ_{free}) with other types of adversaries.

9.5 A Regret Lower-Bound in the Online Bandit CB Game

In the previous sections, we have focused on the analysis of the regret upper-bounds guaranteed by algorithms in the classes of OSPBand and the online bandit CB game. To provide further justification of our findings, in this chapter, we consider the problem of characterizing regret lower-bounds in these classes of problems.

In Audibert, Bubeck, and Lugosi (2014), an instance of OComb with bandit feedback—where the learner's action set is $S \subset \{0,1\}^E$ and that $\|z\|_1 \leq n, \forall z \in S$—is analyzed and it is showed that for any learning policy, the expected regret suffered by the learner is at least $\Omega(n\sqrt{ET})$. We can follow this work and design an instance of OSPBand that yields a similar result. Formally, we have the following proposition:

Proposition 9.5.1. *There exists an instance of* OSPBand *on a graph, where the number of edges is E and the length of the longest paths is n, such that*

$$\inf_{\textit{strategies}} \sup_{\textit{adversaries}} R_T = \Omega\left(n\sqrt{ET}\right).$$

Here, the inf and sup are taken over all possible learning policies of the learner and all (feasible) strategies of the adversary.[11] A proof of this proposition is given in Appendix F.2. From this result, we observe that if there exists an exploration distribution μ such that $\lambda^*[M(\mu)]$ is of the order $\Omega(1/E)$, then Edge(μ) (and ComBand(μ)) yields a regret upper-bound that is tight with this lower bound. It remains an open question to characterize which class of OSPBand satisfying that such an exploration distribution exists.

For the particular class of OSPBand corresponding to the online bandit CB game (i.e., playing on the graph $G_{k,n}$ and the losses are generated according to the Blotto-rule—see Section 7.1.1), it remains an open question whether there exists an instance

[11]More precisely, the sup is taken over the class of all possible processes that generate the loss vector $\boldsymbol{\ell}_t \in [0,1]^E$ from measurable functions of the past information (i.e., $\tilde{\boldsymbol{p}}_s, \boldsymbol{\ell}_s, x_s, \forall s \in \{1,2,\ldots,t-1\}$).

such that the regret suffered by the learner always yields a lower-bound with the same order as in Proposition 9.5.1. In this thesis, we can only prove a somewhat weaker result as follows:

Proposition 9.5.2. *There exists an instance of the online bandit CB game with* $2n$ *battlefields* ($n \in \mathbb{N} \backslash \{0\}$) *such that in the corresponding* OSPBand, we have

$$\inf_{\text{strategies}} \sup_{\text{adversaries}} R_T = \Omega\left(n\sqrt{T}\right). \tag{9.9}$$

A proof of this proposition is given in Appendix F.3. The main idea of the proof is to design a special instance of the online bandit CB game (including the choices on number of battlefields, the budgets of the learner and the adversary, the battlefields' values) that is similar to an instance of OComb with bandit feedback studied by Dani et al. (2008) (for finding a lower-bound of OComb). We observe that in this special game instance, for any exploration distribution μ on the graph $G_{k,n}$, the gap between the regret lower-bound in (9.9) and the upper-bound provided by Edge(μ) is always at least of the order $\Omega(\min\{n-1, k\})$.[12] We conjecture that it is possible to find another instance of the online bandit CB game yielding a larger lower-bound on the regret; we leave this question for future studies. Finally, for the sake of comparison, note that *if there exists* an instance of the online bandit CB game yielding the result in Proposition 9.5.1, the corresponding regret lower-bound is of the order $\Omega(kn^{3/2}\sqrt{T})$; that is larger than the one in (9.9) by a factor of $k\sqrt{n}$.

. . . .

Summary: In this chapter, we presented the Edge algorithm, a variant of the well-known ComBand algorithm, that runs efficiently in any online shortest path problem with bandit feedback (OSPBand). We also proposed a fast method to compute exploration distributions to use as inputs in Edge that improves the expected regret guarantees. We applied these findings to the online bandit CB game, cast into an OSPBand. We presented regret lower-bounds for OSPBand with a generic graph and for the set of OSPBand instances corresponding to online bandit CB games. The results in this chapter not only extend the scope of application of the online CB game in practice (even for large instances) but also contribute to the literature of OSPBand.

[12] Here, recall that k is the budget of the learner at each stage.

Part III

CONCLUSIONS AND PERSPECTIVES

In this part of the manuscript, we present an overview of the results obtained in the previous chapters and then discuss several related open challenges as well as directions for potential future work.

Conclusions of the Thesis

In this thesis, we chose to investigate resource allocation games and we primarily focused on the Colonel Blotto game (CB game) as a case study. We present below an overview of the results obtained in several prominent instances of this class of games under the two main settings: the offline setting and the online learning setting.

The Offline Setting

In Part I of the thesis, we modeled resource allocation games as one-shot complete-information games and focused on analyzing players' behaviors in a game-theoretic perspective. In particular, we studied several resource allocation games belonging to the family of Blotto games including the generalized CB game, the discrete CB game, the generalized Lottery Blotto game (LB game) and the generalized-rule CB game. We addressed the challenge of searching for strategies that can be efficiently constructed such that they provably guarantee good payoffs for players in these Blotto games. We obtained the following results:

(i) We defined the generalized CB game with a formulation that is more general than that of any other work in the literature. We proposed a simply-constructed class of strategies (the IU strategies) yielding approximate equilibria of this game and proved that the approximation error in using the IU strategies (relative to the magnitude of players' payoffs) quickly decreases as the number of battlefields increases. Therefore, this error is considerably negligible in the generalized CB game with a large number of battlefields, showing the practicality of the IU strategies. These results are presented in Chapter 3 and Chapter 4.

(ii) We studied the (constant-sum) discrete CB game. In this game, we efficiently constructed a class of approximate equilibria by extending the ideas of the IU strategies with a non-trivial round-up process. We characterized the involved error in terms of the game's parameters. We showed, by numerical experiments, the benefits in running time when using the proposed approximate equilibria (in comparison with exact equilibria computations that remain inefficient in implementation) and the trade-off between these benefits and the approximation errors in the players' payoffs. These are the main results of Chapter 5.

(iii) We analyzed the generalized LB game—a natural extension of the CB game—and showed that the IU strategies are also approximate equilibria of this game. We applied this result to the LB games with the power-form and logit-form CSFs (two of the most common CSFs in the literature) and showed that the involved approximation errors are negligible under a condition on the number of battlefields and the parameters of the CSFs. Finally, we presented several initial results on the (constant-sum) generalized-rule CB game (GR-CB game). As a tool to study this game, we characterized the exact equilibria (in all parameters' configurations) of the all-pay auction with favoritism. We used these results to propose an efficient heuristic algorithm computing a set of distributions that approximates the optimal univariate distributions of the GR-CB game. Based on them, we then constructed a class of approximate equilibria of the GR-CB game under an assumption. These results are given in Chapter 6.

From the analyses leading to these results, we inspected the difficulty in the (exact) equilibrium characterization of the CB game: the correlation between allocations of each player towards different battlefields. Essentially, this comes from the budget constraints of the CB game—the kind of constraints that also appears in other resource allocation games. The IU strategies and other extended ideas are our method to bypass this difficulty in the CB game (and other Blotto games), with a trade-off involving small errors. We conjecture that this approach might be applied similarly in other resource allocation games with similar challenges. On the other hand, although the idea behind the class of IU strategies is fairly simple and straightforward, we found that proving meaningful results from this idea is often non-trivial. Moreover, we observed that a modification in the formulation of the CB game, even small, can lead to new issues and it requires a different set of techniques to solve them. Therefore, one really needs careful consideration in implementing this approach on a given game. Finally, in constructing the results mentioned above, we obtained several interesting side-results, including a complete characterization of exact equilibria of all-pay auctions with favoritism and an efficient algorithm, based on dynamic programming, to compute best-responses in the discrete CB game.

The Online Learning Setting

In Part II of the thesis, we formulated the class of online resource allocation games to capture situations where players play repeatedly a game without knowing all information when making decisions. We addressed the question of how to play in online resource allocation games with combinatorial structures to obtain a good guarantee on payoffs while maintaining an efficient implementability. To answer this question, we conducted a regret-minimization analysis on the online discrete CB game and several other online resource allocation games (with combinatorial structures). We obtained the following results:

(i) We defined the online (discrete) CB game and the online hide-and-seek game (HS game) and showed that they can be cast into online shortest path problems (OSP). To the best of our knowledge, these are the first online learning models formulated for these games. We discussed similar conversions of several other online resource allocation games with combinatorial structures. We also explicitly defined a novel online learning model: OSP with side-observations (SOOSP). These results are given in Chapter 7.

(ii) We studied the the SOOSP model and showed that it captures well the online semi-bandit CB game. We designed a novel algorithm, called Exp3-OE, that runs efficiently in any generic instance of SOOSP, i.e., its running time is polynomial in terms of the number of edges of the graph in SOOSP. Moreover, in several cases of interest, we proved that Exp3-OE improves the regret guarantees in comparison with state-of-the-art algorithms for SOOSP. We applied these findings to the online semi-bandit CB game and the online HS game, being cast into SOOSP, and showed the benefits of this algorithm: it improves the regret guarantee and always runs in polynomial time in terms of the games' parameters. These results are presented in Chapter 8.

(iii) We studied the OSP with bandit feedback (OSPBand) model. We designed an algorithm, called Edge, that is a modified version of the classical ComBand algorithm. Edge runs much more efficiently than ComBand in any instance of OSPBand and it has a simpler representation than CombD—another efficient algorithm in the literature. Moreover, we designed and analyzed Edge with more choices of exploration distributions to use as inputs than CombD. We introduced a procedure, based on derivative-free optimization, that quickly provides a distribution μ_{free} such that when using Edge (or ComBand) with μ_{free} as an exploration distribution, regret guarantees are improved compared to other set-ups of ComBand used in the literature. We applied Edge to the online bandit CB game and conducted numerical experiments confirming the improvements in performance and implementability benefited from Edge. These results are given in Chapter 9.

In the analyses leading to these results, we showed how to exploit structures of online resource allocation games, such as the online CB game and the online HS game, to make important connections with several online learning frameworks and to improve learning policies. The scope of applications of our obtained results go beyond merely solving online resource allocation games and they also contribute to the literature of OSP in general.

Future Work

In this section, we discuss new questions and directions that our work opens up, starting with the ones directly relating to the results obtained in this thesis, and then moving to other interesting extensions and broader classes of problems.

Direct Extensions of the Obtained Results

In the offline setting, naturally, the leading open question concerns the exact equilibrium of the generalized CB game: can we prove (or disprove) its existence and if it exists, how to construct (or at least characterize) it? In our humble opinion, this question is challenging (see also the discussion on this problem by Kovenock and Roberson (2015)). From the results obtained in Chapter 4, we conjecture that the game with a large number of battlefields might be an easier case to start the study on exact equilibrium. This remark comes from our observation that when this parameter tends to infinity, the IU strategies—our proposed approximate equilibria of this game—get closer to obtain optimality (the approximation error, relative to the magnitude of players' payoffs, tends to zero). Another direction would be to extend the equilibrium construction, based on the copula theory, of Roberson (2006) (for the restricted case of constant-sum CB games with homogeneous battlefields) to the generalized CB game. In other Blotto games, characterizing the exact equilibria is also a fundamental open question.

For approximate equilibria in Blotto games, it would be interesting to find other strategies that also guarantee good payoffs for the players and compare them to our solutions given in this thesis. In particular, our proposed approximate equilibria in the generalized LB games are based on the analysis of the generalized CB game (adding an extra term to the approximation error), and a relevant open question is to find other solutions for the LB game without using its connection to the CB game. On the other hand, in the generalized-rule CB game, our results are only obtained under an assumption on the existence of a positive solution of an equation (Equation (6.17) in Chapter 6). In our numerical simulations, however, this assumption appears to always hold, and the solution seems to be unique. A direct open challenge in this game is to prove theoretically these (numerical) results. This helps us to be able to obtain similar results without the necessity of that assumption and also can lead to improvements in the approximation error. Another direction that may extend further the scope of applications of the GR-CB game is to analyze situations where players may freely choose pre-allocations and effectiveness of resources (possibly under certain constraints or with a cost) instead of considering them as exogenous factors as in our model.

In the online learning perspective, our results open up several fundamental questions. In the SOOSP model (also for the online semi-bandit CB game and the online HS game), there is still a gap between the regret upper-bound provided by our proposed Exp3-OE algorithm and available regret lower-bounds in the literature. To tighten this gap, we can either improve even further Exp3-OE or look for better lower bounds. Besides this direction, we are also interested in looking for conditions on observations graphs (other than the symmetry property and Assumption (A1) studied in Chapter 8) such that when applying Exp3-OE to SOOSP satisfying such conditions, its regret guarantee is improved.

In the OSPBand model (also for the online bandit CB game), the fundamental question of finding an efficient method to compute the (exact) optimal exploration distribution to use in our proposed Edge algorithm (and the ComBand algorithm) still remains open. In Chapter 9, we proposed a heuristic procedure that efficiently searches among the distributions satisfying Assumption (WP)—which technically does not cover the set of all usable distributions. However, interestingly, in our numerical experiments for graphs having particular symmetries corresponding to the online bandit CB game, it appears that the heuristic can find very near-optimal exploration distributions. This raises two questions as follows: Can we prove that there exists a distribution in the E-dimensional search-space defined by Assumption (WP) yielding optimality? Which condition of the graphs in OSPBAND is sufficient for this property to hold? A better regret lower-bound for OSPBAND (and for the online bandit CB game) is another important challenge.

Finally, recall that in the online learning setting, we focused specifically on online resource allocation games with combinatorial structures, with the online *discrete* CB game as a case study. An obvious extension is to consider games with continuous strategy sets, e.g., the online version of the generalized CB game. One natural direction we can follow is to exploit techniques from the literature of multi-armed bandits with continuum-arm (see e.g., R. Agrawal (1995)) and/or bandits in metric spaces (see e.g., Kleinberg et al. (2008)). However, in the particular case of the online generalized CB game, it might be non-trivial to apply the results from these models because the losses suffered by the learner are generated by discontinuous functions due to the rule of the game.

Other Research Directions

Beyond the direct extensions mentioned above, the framework of resource allocation games that we introduced in this thesis can lead to other research directions and more broadly defined problems. One of the interesting situations appearing in practice is resource allocation games where cooperative players can form coalitions. For example, a security problem where N defenders want to cooperate to fight against a strong attacker can be modeled as a CB game (with more than two players) with the setup as follows: A set of N (selfish) players with small budgets compete against one opponent with a larger budget. The "small" players win a battlefield if their aggregate allocations is higher than the "big" player's allocation and in that case, the value of this battlefield is then shared among the small players. Each "small" player wants to optimize her cooperation with others to gain as much as possible while the "big" player needs to predict such coalitions and fights against them. Our results in the two-player CB game may serve as bases for studying fundamental questions in this cooperative model such as how important is each player to certain coalitions or what is the optimal payoff that each player can achieve. Another applicable setting that may be considered, when cooperation is allowed in resource allocation games, is where players in a team can communicate but with a cost, and an interesting question is to find conditions under which players have incentives to communicate.

Recall that in this thesis, we use the online learning perspective to model and study sequential plays in resource allocation games with incomplete information (by the regret-minimization analysis). An alternative approach is to model them as repeated incomplete information games (e.g., by the model of Aumann et al. (1995)). For instance, types of players can be defined by their budgets that will be drawn randomly (and kept as private information). Another interesting setting is the LB game where contest success functions (CSFs) are drawn randomly from a given set, players may know this set but not precisely the drawn CSFs. In these cases, a Bayes-Nash equilibrium analysis could be an important contribution. Another direction is where players have asymmetric information about the games and the main objective is to determine conditions in which a player has an incentive to reveal his information. An example is the setting where players, without knowledge of her opponents' budgets, can choose to pre-allocate before the game starts; at a high-level, this pre-allocation is equivalent to revealing partial information about their budget. Initial analyses of this approach appeared in some works in the literature, e.g., Chandan et al. (2020).

Learning equilibria in games is another worth-mentioning direction that relates to online learning techniques (see also our discussion in Section 7.3.4). On the one hand, recall that in the literature of the (one-shot) constant-sum discrete CB game (which can be easily transformed to a zero-sum game), it remains impractical to implement the algorithms computing the exact equilibria in large-scale instances (see our discussion in Chapter 5). On the other hand, when zero-sum games are played repeatedly, a well-known result is that the (marginal) empirical distributions of plays converge almost surely to the set of Nash equilibria. Therefore, it is interesting to look for a method, based on this convergence, to compute the equilibria of the discrete CB game, then compare its complexity (depending on the related rate of convergence) with that of other available methods in the literature. Another fundamental question is to study the convergence of the actual sequence of plays when players doing regret-minimization policies in online resource allocation games. However, it seems that when following this direction, the structure of resource allocation games does not provide any particular advantage and the question in this case is as challenging as that in a general setting.

Finally, we discuss an interesting class of problems having a close relation with the resource allocation games considered in this thesis: multi-item auctions—simply defined as situations where a bidder needs to decide the bids for several simultaneously running auctions (selling different items) in order to optimize the aggregate payoffs. In a multi-item auction requiring integer bids, the strategy set of a bidder (with a virtual budget which is equal to the aggregate of the values that she assigns on the items) is very similar to that of a player in the discrete CB game. Therefore, as in the online (discrete) CB game, an online multi-item auction with integer bids, where a learner plays and receives a stream of limited feedback, can also be cast into an online shortest path problem. The question that arises is how a bidder can exploit particular auction rules to improve her performance (e.g., in the first-price and second-price auctions, side-observations might also exist and can be exploited).

Other settings in multi-item auctions are also interesting to study. There exist applications where a bidder is provided a total budget to distribute throughout the stages and toward many simultaneously running auctions; for example, a marketing campaign is invested with a certain budget and is allowed to freely choose how to distribute it in a certain time duration. This situation involves not only the correlation between the decisions in different auctions but also the correlation between decisions in different stages. A possible approach is to model this problem into the multi-armed bandit with knapsacks framework (see e.g., Badanidiyuru et al. (2013)). Another crucial problem in online multi-item auctions is the optimization of the auctioneer. An interesting question, in the mechanism design perspective, is the following: given the freedom in choosing the auction rule, what is an optimal choice of an auctioneer such that when all bidders conduct a specific no-regret policy, it brings the highest profit?

List of Publications

Quezada, Franco, Céline Gicquel, Safia Kedad-Sidhoum, and Dong Quan Vu (2020). "A Multi-stage Stochastic Integer Programming Approach for a Multi-echelon Lot-sizing Problem with Returns and Lost Sales". In: *Computers & Operations Research* 116, p. 104865.

Vu, Dong Quan, Patrick Loiseau, and Alonso Silva (2018a). "A Simple and Efficient Algorithm to Compute Epsilon-Equilibria of Discrete Colonel Blotto Games". In: *Proceedings of the 17th International Conference on Autonomous Agents and Multi Agent Systems (AAMAS)*, pp. 2115–2117.

— (2018b). "Efficient Computation of Approximate Equilibria in Discrete Colonel Blotto Games". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence (IJCAI-ECAI)*, pp. 519–526.

— (2019a). *Approximate Equilibria in Non-constant-sum Colonel Blotto and Lottery Blotto Games with Large Numbers of Battlefields*. arXiv: 1910.06559 [cs.GT].

— (2019b). "Combinatorial Bandits for Sequential Learning in Colonel Blotto Games". In: *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*.

Vu, Dong Quan, Patrick Loiseau, Alonso Silva, and Long Tran-Thanh (2020). "Path Planning Problems with Side Observations—When Colonels Play Hide-and-Seek". In: *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*.

References

Abernethy, Jacob, E Hazan, and Alexander Rakhlin (2008). "Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization". In: *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pp. 263–273.

Abernethy, Jacob and Alexander Rakhlin (2009). "Beating the Adaptive Bandit with High Probability". In: *Information Theory and Applications Workshop*. IEEE, pp. 280–289.

Adamo, Tim and Alexander Matros (2009). "A Blotto Game with Incomplete Information". In: *Economics Letters* 105.1, pp. 100–102.

Agrawal, Rajeev (1995). "The continuum-armed bandit problem". In: *SIAM journal on control and optimization* 33.6, pp. 1926–1951.

Agrawal, Shipra and Nikhil R Devanur (2014). "Bandits with concave rewards and convex knapsacks". In: *Proceedings of the 15th ACM conference on Economics and Computation (EC)*, pp. 989–1006.

Ahmadinejad, Amir Mahdi, Sina Dehghani, Mohammad Taghi Hajiaghayi, Brendan Lucier, Hamid Mahini, and Saeed Seddighin (2016). "From Duels to Battlefields: Computing Equilibria of Blotto and Other Games". In: *Proceedings of the 13th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 369–375.

Alcalde, José and Matthias Dahm (2007). "Tullock and Hirshleifer: a meeting of the minds". In: *Review of Economic Design* 11.2, pp. 101–124.

Alon, Noga, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren (2015). "Online learning with feedback graphs: Beyond bandits". In: *JMLR Workshop and Conference Proceedings*. Vol. 40. Microtome Publishing.

Alon, Noga, Nicolo Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir (2017). "Nonstochastic multi-armed bandits with graph-structured feedback". In: *SIAM Journal on Computing* 46.6, pp. 1785–1826.

Alon, Noga, Nicolo Cesa-Bianchi, Claudio Gentile, and Yishay Mansour (2013). "From bandits to experts: A tale of domination and independence". In: *Advances in Neural Information Processing Systems 26 (NIPS)*, pp. 1610–1618.

Alpern, Steve, Vic Baston, and Shmuel Gal (2008). "Network search games with immobile hider, without a designated searcher starting point". In: *International Journal of Game Theory* 37.2, pp. 281–302.

— (2009). "Searching symmetric networks with utilitarian-postman paths". In: *Networks: An International Journal* 53.4, pp. 392–402.

Altman, Eitan, Thomas Boulogne, Rachid El-Azouzi, Tania Jiménez, and Laura Wynter (2006). "A survey on networking games in telecommunications". In: *Computers & Operations Research* 33.2, pp. 286–311.

Amir, Nadav (2018). "Uniqueness of optimal strategies in Captain Lotto games". In: *International Journal of Game Theory* 47.1, pp. 55–101.

Audibert, Jean-Yves and Sébastien Bubeck (2009). "Minimax Policies for Adversarial and Stochastic Bandits". In: *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*.

— (2010). "Regret bounds and minimax policies under partial monitoring". In: *Journal of Machine Learning Research* 11.Oct, pp. 2785–2836.

Audibert, Jean-Yves, Sébastien Bubeck, and Gábor Lugosi (2011). "Minimax policies for combinatorial prediction games". In: *Proceedings of the 24th Annual Conference on Learning Theory (COLT)*, pp. 107–132.

— (2014). "Regret in Online Combinatorial Optimization". In: *Mathematics of Operations Research* 39.1, pp. 31–45.

Audibert, Jean-Yves, Rémi Munos, and Csaba Szepesvári (2009). "Exploration–exploitation tradeoff using variance estimates in multi-armed bandits". In: *Theoretical Computer Science* 410.19, pp. 1876–1902.

Auer, Peter (2002). "Using confidence bounds for exploitation-exploration trade-offs". In: *Journal of Machine Learning Research* 3.Nov, pp. 397–422.

Auer, Peter, Nicolo Cesa-Bianchi, and Paul Fischer (2002). "Finite-time analysis of the multiarmed bandit problem". In: *Machine learning* 47.2-3, pp. 235–256.

Auer, Peter, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire (1995). "Gambling in a rigged casino: The adversarial multi-armed bandit problem". In: *focs*. IEEE, p. 322.

— (2002). "The nonstochastic multiarmed bandit problem". In: *SIAM journal on computing* 32.1, pp. 48–77.

Aumann, Robert J (1985). "What is game theory trying to accomplish?" In: *Frontiers of Economics, edited by K. Arrow and S. Honkapohja*. Citeseer.

Aumann, Robert J, Michael Maschler, and Richard E Stearns (1995). *Repeated games with incomplete information*. MIT press.

Avenhaus, Rudolf and Morton John Canty (1996). *Compliance quantified: An introduction to data verification*. Cambridge University Press.

Avenhaus, Rudolf and D Marc Kilgour (2004). "Efficient distributions of arms-control inspection effort". In: *Naval Research Logistics (NRL)* 51.1, pp. 1–27.

Badanidiyuru, Ashwinkumar, Robert Kleinberg, and Aleksandrs Slivkins (2013). "Bandits with knapsacks". In: *IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, pp. 207–216.

Bahl, Harish C, Larry P Ritzman, and Jatinder ND Gupta (1987). "Or practice—determining lot sizes and resource requirements: A review". In: *Operations Research* 35.3, pp. 329–345.

Banos, Alfredo et al. (1968). "On pseudo-games". In: *The Annals of Mathematical Statistics* 39.6, pp. 1932–1945.

Baron, David P and John A Ferejohn (1989). "Bargaining in legislatures". In: *American political science review* 83.4, pp. 1181–1206.

Baye, M. R., Dan Kovenock, and C. G. De Vries (1994). "The solution to the Tullock rent-seeking game when R> 2: Mixed-strategy equilibria and mean dissipation rates". In: *Public Choice* 81.3-4, pp. 363–380.

Baye, M. R., Dan Kovenock, and C. G. de Vries (1996). "The all-pay auction with complete information". In: *Economic Theory* 8.2, pp. 291–305.

Beck, Anatole and DJ Newman (1970). "Yet more on the linear search problem". In: *Israel journal of mathematics* 8.4, pp. 419–429.

Behnezhad, Soheil, Avrim Blum, Mahsa Derakhshan, MohammadTaghi HajiAghayi, Mohammad Mahdian, Christos H Papadimitriou, Ronald L Rivest, Saeed Seddighin, and Philip B Stark (2018). "From battlefields to elections: Winning strategies of blotto and auditing games". In: *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, pp. 2291–2310.

Behnezhad, Soheil, Sina Dehghani, Mahsa Derakhshan, Mohammad Taghi Haji Aghayi, and Saeed Seddighin (2017). "Faster and Simpler Algorithm for Optimal Strategies of Blotto Game". In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 369–375.

Bertsekas, Dimitri P. (2017). *Dynamic Programming and Optimal Control*. 4th ed. Vol. 1. Athena Scientific.

Besson, Lilian and Emilie Kaufmann (2018). "What Doubling Tricks Can and Can't Do for Multi-Armed Bandits". In: *arXiv preprint arXiv:1803.06971*.

Bhattacharya, Sourabh, Tamer Başar, and Maurizio Falcone (2014). "Surveillance for security as a pursuit-evasion game". In: *Proceedings of the 5th International Conference on Decision and Game Theory for Security (GameSec)*, pp. 370–379.

Blackett, Donald W (1958). "Pure strategy solutions of Blotto games". In: *Naval Research Logistics Quarterly* 5.2, pp. 107–109.

Borel, E (1921). "La théorie du jeu et les équations intégrales à noyau symétrique". In: *Comptes rendus de l'Académie des Sciences* 173.1304-1308, p. 58.

Borel, E and J Ville (1938). *Application de la théorie des probabilités aux jeux de hasard*. original edition by Gauthier-Villars, Paris, 1938; reprinted at the end of Théorie mathématique du bridge à la portée de tous, by E. Borel & A. Chéron, Editions Jacques Gabay, Paris.

Bostock, F.A. (1984). "On a discrete search problem on three arcs". In: *SIAM Journal on Algebraic Discrete Methods* 5.1, pp. 94–100.

Brassard, Gilles and Paul Bratley (1996). *Fundamentals of Algorithmics*. Prentice-Hall, Inc.

Bubeck, Sébastien, Nicolo Cesa-Bianchi, et al. (2012). "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: *Foundations and Trends® in Machine Learning* 5.1, pp. 1–122.

Bubeck, Sébastien, Nicolo Cesa-Bianchi, and Sham M Kakade (2012). "Towards minimax policies for online linear optimization with bandit feedback". In: *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*. Vol. 23. Microtome, pp. 41–1.

Bubeck, Sébastien, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári (2011). "X-armed bandits". In: *Journal of Machine Learning Research* 12.May, pp. 1655–1695.

Bubeck, Sébastien and Aleksandrs Slivkins (2012). "The Best of Both Worlds: Stochastic and Adversarial Bandits". In: *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, pp. 42.1–42.23.

Cesa-Bianchi, Nicolo and Gábor Lugosi (2006). *Prediction, learning, and games*. Cambridge university press.

— (2012). "Combinatorial bandits". In: *Journal of Computer and System Sciences* 78.5, pp. 1404–1422.

Challet, Damien and Yi-Cheng Zhang (1998). "On the minority game: Analytical and numerical studies". In: *Physica A: Statistical Mechanics and its applications* 256.3-4, pp. 514–532.

Chandan, Rahul, Keith Paarporn, and Jason R Marden (2020). "When showing your hand pays off: Announcing strategic intentions in Colonel Blotto games". In: *arXiv preprint arXiv:2002.11648*.

Che, Yeon-Koo and Ian Gale (2000). "Difference-Form Contests and the Robustness of All-Pay Auctions". In: *Games and Economic Behavior* 30.1, pp. 22–43.

Chia, Pern Hui (2012). "Colonel Blotto in web security". In: *The 11th Workshop on Economics and Information Security, WEIS Rump Session*, pp. 141–150.

Chien, Su Fong, Charilaos C Zarakovitis, Qiang Ni, and Pei Xiao (2019). "Stochastic Asymmetric Blotto Game Approach for Wireless Resource Allocation Strategies". In: *IEEE Transactions on Wireless Communications*.

Clark, Derek J and Christian Riis (1998a). "Competition over more than one prize". In: *The American Economic Review* 88.1, pp. 276–289.

— (1998b). "Contest success functions: an extension". In: *Economic Theory* 11.1, pp. 201–204.

— (2000). "Allocation efficiency in a competitive bribery game". In: *Journal of Economic Behavior & Organization* 42.1, pp. 109–124.

Cohen, Alon, Tamir Hazan, and Tomer Koren (2016). "Online Learning with Feedback Graphs Without the Graphs". In: *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 811–819.

Cohen, Chen and Aner Sela (2007). "Contests with ties". In: *The BE Journal of Theoretical Economics* 7.1.

Combes, Richard, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. (2015). "Combinatorial bandits revisited". In: *Advances in Neural Information Processing Systems 28 (NIPS)*, pp. 2116–2124.

Corchón, Luis C (2007). "The theory of contests: a survey". In: *Review of Economic Design* 11.2, pp. 69–100.

Corchón, Luis C and Matthias Dahm (2010). "Foundations for contest success functions". In: *Economic Theory* 43.1, pp. 81–98.

Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein (2009). *Introduction to Algorithms, Third Edition*. The MIT Press.

Dagan, Arnon and Shmuel Gal (2008). “Network search games, with arbitrary searcher starting point”. In: *Networks: An International Journal* 52.3, pp. 156–161.

Dani, Varsha, Sham M Kakade, and Thomas P Hayes (2008). “The Price of Bandit Information for Online Optimization”. In: *Advances in Neural Information Processing Systems 21 (NIPS)*, pp. 345–352.

De Guenin, Jacques (1961). “Optimum distribution of effort: An extension of the Koopman basic theory”. In: *Operations Research* 9.1, pp. 1–7.

Domingos, Pedro and Matt Richardson (2001). “Mining the network value of customers”. In: *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 57–66.

Dresher, Melvin (1962). *A sampling inspection problem in arms control agreements: A game-theoretic analysis*. Tech. rep. Rand Corp Santa Monica CA.

Duffy, John and Alexander Matros (2015). “Stochastic asymmetric Blotto games: Some new results”. In: *Economics Letters* 134, pp. 4–8.

Dziubiński, Marcin (2013). “Non-symmetric discrete General Lotto games”. In: *International Journal of Game Theory* 42.4, pp. 801–833.

Elton, Edwin J, Martin J Gruber, and Manfred W Padberg (1976). “Simple criteria for optimal portfolio selection”. In: *The journal of Finance* 31.5, pp. 1341–1357.

Faure, Mathieu, Pierre Gaillard, Bruno Gaujal, and Vianney Perchet (2015). “Online learning and game theory. a quick overview with recent results and applications”. In: *ESAIM: Proceedings and Surveys* 51, pp. 246–271.

Flood, Merrill M (1972). “The hide and seek game of Von Neumann”. In: *Management Science* 18.5-part-2, pp. 107–109.

Freund, Yoav and Robert E Schapire (1997). “A decision-theoretic generalization of on-line learning and an application to boosting”. In: *Journal of computer and system sciences* 55.1, pp. 119–139.

Friedman, Lawrence (1958). “Game-Theory Models in the Allocation of Advertising Expenditures”. In: *Operations Research* 6.5, pp. 699–709.

Fu, Qiang (2006). “A theory of affirmative action in college admissions”. In: *Economic Inquiry* 44.3, pp. 420–428.

Fu, Qiang and Zenan Wu (2019). “Contests: Theory and topics”. In: *Oxford Research Encyclopedia of Economics and Finance*.

Ganesh, K, Sanjay Mohapatra, RA Malairajan, and M Punniyamoorthy (2015). *Resource Allocation Problems in Supply Chains*. Emerald Group Publishing.

Garivier, Aurélien and Olivier Cappé (2011). “The KL-UCB algorithm for bounded stochastic bandits and beyond”. In: *Proceedings of the 24th Annual Conference on Learning Theory (COLT*, pp. 359–376.

Garnaev, A (2007). “Find a hidden “treasure””. In: *Naval Research Logistics (NRL)* 54.1, pp. 109–114.

Gerchinovitz, Sébastien (2013). “Sparsity regret bounds for individual sequences in online linear regression”. In: *Journal of Machine Learning Research* 14.Mar, pp. 729–769.

Gerchinovitz, Sébastien and Tor Lattimore (2016). "Refined lower bounds for adversarial bandits". In: *Advances in Neural Information Processing Systems 29 (NIPs)*, pp. 1198–1206.

Gittins, John, Kevin Glazebrook, and Richard Weber (2011). *Multi-armed bandit allocation indices*. John Wiley & Sons.

Gross, Oliver (1950). *The symmetric Blotto game*. Tech. rep. US Air Force Project RAND Research Memorandum.

Gross, Oliver and Robert Wagner (1950). *A continuous Colonel Blotto game*. Tech. rep. RAND project air force Santa Monica CA.

György, András, Tamás Linder, Gábor Lugosi, and György Ottucsák (2007). "The online shortest path problem under partial monitoring". In: *Journal of Machine Learning Research* 8.Oct, pp. 2369–2403.

Hajimirsaadeghi, M. and N. B. Mandayam (2017). "A dynamic colonel Blotto game model for spectrum sharing in wireless networks". In: *Proceedings of the 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 287–294.

Hannan, James (1957). "Approximation to Bayes risk in repeated play". In: *Contributions to the Theory of Games* 3, pp. 97–139.

Hart, Sergiu (2008). "Discrete Colonel Blotto and General Lotto games". In: *International Journal of Game Theory* 36.3, pp. 441–460.

— (2016). "Allocation games with caps: from Captain Lotto to all-pay auctions". In: *International Journal of Game Theory* 45.1-2, pp. 37–61.

Hart, Sergiu and Andreu Mas-Colell (2000). "A simple adaptive procedure leading to correlated equilibrium". In: *Econometrica* 68.5, pp. 1127–1150.

— (2003). "Uncoupled dynamics do not lead to Nash equilibrium". In: *American Economic Review* 93.5, pp. 1830–1836.

Heliou, Amélie, Johanne Cohen, and Panayotis Mertikopoulos (2017). "Learning with bandit feedback in potential games". In: *Advances in Neural Information Processing Systems 30 (NIPS)*, pp. 6369–6378.

Helmbold, David P and Manfred K Warmuth (2009). "Learning permutations with exponential weights". In: *Journal of Machine Learning Research* 10.Jul, pp. 1705–1736.

Herrera, Juliver Gil and Juan Felipe Botero (2016). "Resource allocation in NFV: A comprehensive survey". In: *IEEE Transactions on Network and Service Management* 13.3, pp. 518–532.

Hillman, Arye L. and John G. Riley (1989). "Politically contestable rents and transfers". In: *Economics & Politics* 1.1, pp. 17–39.

Hirshleifer, Jack (1989). "Conflict and rent-seeking success functions: Ratio vs. difference models of relative success". In: *Public choice* 63.2, pp. 101–112.

Hoeffding, W (1963). "Probability inequalities for sums of bounded random variables". In: *Journal of the American Statistical Association*, 58, pp. 13–30.

Hohzaki, Ryusuke (2007). "An inspection game with multiple inspectees". In: *European Journal of Operational Research* 178.3, pp. 894–906.

— (2016). "Search games: Literature and survey". In: *Journal of the Operations Research Society of Japan* 59.1, pp. 1–34.

Hortala-Vallve, Rafael and Aniol Llorente-Saguer (2012). "Pure strategy Nash equilibria in non-zero sum colonel Blotto games". In: *International Journal of Game Theory* 41.2, pp. 331–343.

Hu, Yi-Qi, Hong Qian, and Yang Yu (2017). "Sequential Classification-Based Optimization for Direct Policy Search." In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2029–2035.

Hussain, Hameed et al. (2013). "A survey on resource allocation in high performance distributed computing systems". In: *Parallel Computing* 39.11, pp. 709–736.

Ibaraki, Toshihide and Naoki Katoh (1988). *Resource allocation problems: algorithmic approaches*. MIT press.

Kakade, Sham M, Shai Shalev-Shwartz, and Ambuj Tewari (2008). "Efficient Bandit Algorithms for Online Multiclass Prediction". In: *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pp. 440–447.

Kalai, Adam and Santosh Vempala (2005). "Efficient algorithms for online decision problems". In: *Journal of Computer and System Sciences* 71.3, pp. 291–307.

Kale, Satyen, Lev Reyzin, and Robert E Schapire (2010). "Non-stochastic Bandit Slate Problems". In: *Advances in Neural Information Processing Systems 23 (NIPS)*, pp. 1054–1062.

Karimi, Behrooz, SMT Fatemi Ghomi, and JM Wilson (2003). "The capacitated lot sizing problem: a review of models and algorithms". In: *Omega* 31.5, pp. 365–378.

Kaufmann, Emilie et al. (2018). "On Bayesian index policies for sequential resource allocation". In: *The Annals of Statistics* 46.2, pp. 842–865.

Kaufmann, Emilie, Nathaniel Korda, and Rémi Munos (2012). "Thompson sampling: An asymptotically optimal finite-time analysis". In: *Proceedings of the 23rd International Conference on Algorithmic Learning Theory (ALT)*. Springer, pp. 199–213.

Kim, Bara and Jeongsim Kim (2019). "Existence of a unique Nash equilibrium for an asymmetric lottery Blotto game with weighted majority". In: *Journal of Mathematical Analysis and Applications* 479.1, pp. 1403–1415.

Kim, Geofferey Jiyun, Jerim Kim, and Bara Kim (2018). "A lottery Blotto game with heterogeneous items of asymmetric valuations". In: *Economics Letters* 173, pp. 1–5.

Kim, Jeongsim and Bara Kim (2017). "An asymmetric lottery Blotto game with a possible budget surplus and incomplete information". In: *Economics Letters* 152, pp. 31–35.

Kirkegaard, Rene (2012). "Favoritism in asymmetric contests: Head starts and handicaps". In: *Games and Economic Behavior* 76.1, pp. 226–248.

Kitahara, Minoru and Ryo Ogawa (2010). "All-pay auctions with handicaps". In:

Kleinberg, Robert D (2005). "Nearly tight bounds for the continuum-armed bandit problem". In: *Advances in Neural Information Processing Systems 18 (NIPS)*, pp. 697–704.

Kleinberg, Robert D, Aleksandrs Slivkins, and Eli Upfal (2008). "Multi-armed bandits in metric spaces". In: *Proceedings of the 40th annual ACM symposium on Theory of computing*, pp. 681–690.

Klumpp, Tilman, Kai A Konrad, and Adam Solomon (2019). "The dynamics of majoritarian Blotto games". In: *Games and Economic Behavior* 117, pp. 402–419.

Klumpp, Tilman and Mattias K Polborn (2006). "Primaries and the New Hampshire effect". In: *Journal of Public Economics* 90.6-7, pp. 1073–1114.

Knuth, Donald E. (1976). "Big Omicron and Big Omega and Big Theta". In: *SIGACT News* 8.2, pp. 18–24.

Kocák, Tomáš, Gergely Neu, Michal Valko, and Rémi Munos (2014). "Efficient Learning by Implicit Exploration in Bandit Problems with Side Observations". In: *Advances in Neural Information Processing Systems 27 (NIPS)*, pp. 613–621.

Kohli, Pushmeet, Michael Kearns, Yoram Bachrach, Ralf Herbrich, David Stillwell, and Thore Graepel (2012). "Colonel Blotto on Facebook: the effect of social relations on strategic interaction". In: *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 141–150.

Konrad, Kai A (2002). "Investment in the absence of property rights; the role of incumbency advantages". In: *European Economic Review* 46.8, pp. 1521–1537.

Konrad, Kai A and Dan Kovenock (2009). "Multi-battle contests". In: *Games and Economic Behavior* 66.1, pp. 256–274.

Koolen, Wouter M, Manfred K Warmuth, Jyrki Kivinen, et al. (2010). "Hedging Structured Concepts". In: *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*. Citeseer, pp. 93–105.

Koopman, B. O. (1956a). "The Theory of Search. I. Kinematic Bases". In: *Operations Research* 4.3, pp. 324–346.

— (1956b). "The theory of search. II. Target detection". In: *Operations research* 4.5, pp. 503–531.

— (1957). "The theory of search: III. The optimum distribution of searching effort". In: *Operations research* 5.5, pp. 613–626.

Korda, Nathaniel, Emilie Kaufmann, and Remi Munos (2013). "Thompson sampling for 1-dimensional exponential family bandits". In: *Advances in neural information processing systems*, pp. 1448–1456.

Kovenock, Dan and Brian Roberson (2010). "Conflicts with multiple battlefields". In:

— (2011). "A Blotto game with multi-dimensional incomplete information". In: *Economics Letters* 113.3, pp. 273–275.

— (2012). "Coalitional Colonel Blotto games with application to the economics of alliances". In: *Journal of Public Economic Theory* 14.4, pp. 653–676.

— (2015). "Generalizations of the General Lotto and Colonel Blotto games". CESifo Working Paper No. 5291.

Kulpa, Wladyslaw (1997). "The Poincaré-Miranda theorem". In: *The American Mathematical Monthly* 104.6, pp. 545–550.

Kvasov, Dmitriy (2007a). "Contests with limited resources". In: *Journal of Economic Theory* 136.1, pp. 738–748.

— (2007b). "Contests with limited resources". In: *Journal of Economic Theory* 136.1, pp. 738–748.

Lai, TzeLeung and Herbert Robbins (1985). "Asymptotically efficient adaptive allocation rules". In: *Advances in applied mathematics* 6.1, pp. 4–22.

Laslier, J. F. (2002). "How two-party competition treats minorities". In: *Review of Economic Design* 7.3, pp. 297–307.

— (2005). "Party objectives in the "divide a dollar" electoral competition". In: *Social Choice and Strategic Decisions*, pp. 113–130.

Laslier, J. F. and N. Picard (2002). "Distributive politics and electoral competition". In: *Journal of Economic Theory* 103.1, pp. 106–130.

Li, Sanxi and Jun Yu (2012). "Contests with endogenous discrimination". In: *Economics Letters* 117.3, pp. 834–836.

Littlestone, Nick and Manfred K Warmuth (1989). *The weighted majority algorithm*. University of California, Santa Cruz, Computer Research Laboratory.

Liu, Yu-Ren, Yi-Qi Hu, Hong Qian, Yang Yu, and Chao Qian (2017). "ZOOpt/ZOOjl: Toolbox for Derivative-Free Optimization". In: *arXiv preprint arXiv:1801.00329*.

Lombardi, Michele and Michela Milano (2012). "Optimal methods for resource allocation and scheduling: a cross-disciplinary survey". In: *Constraints* 17.1, pp. 51–85.

Lugosi, Gábor, Shie Mannor, and Gilles Stoltz (2008). "Strategies for prediction under imperfect monitoring". In: *Mathematics of Operations Research* 33.3, pp. 513–528.

Macdonell, Scott T. and Nick Mastronardi (2015). "Waging simple wars: a complete characterization of two-battlefield Blotto equilibria". In: *Economic Theory* 58.1, pp. 183–216.

Mahajan, Aditya and Demosthenis Teneketzis (2008). "Multi-armed bandit problems". In: *Foundations and applications of sensor management*. Springer, pp. 121–151.

Maillard, Odalric-Ambrym, Rémi Munos, and Gilles Stoltz (2011). "A finite-time analysis of multi-armed bandits problems with kullback-leibler divergences". In: *Proceedings of the 24th Annual Conference On Learning Theory (COLT)*, pp. 497–514.

Mannor, Shie and Ohad Shamir (2011). "From Bandits to Experts: On the Value of Side-Observations". In: *Advances in Neural Information Processing Systems 24 (NIPS)*, pp. 684–692.

Markowitz, Harry (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, pp. 77–91.

Maschler, Michael (1966). "A price leadership method for solving the inspector's non-constant-sum game". In: *Naval research logistics quarterly* 13.1, pp. 11–33.

Masucci, Antonia Maria and Alonso Silva (2014). "Strategic resource allocation for competitive influence in social networks". In: *Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 951–958.

— (2015). "Defensive resource allocation in social networks". In: *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, pp. 2927–2932.

Menard, Pierre and Aurélien Garivier (2017). "A minimax and asymptotically optimal algorithm for stochastic bandits". In: *Proceedings of the 28th International Conference on Algorithmic Learning Theory (ALT)*. Vol. 76. Proceedings of Machine Learning Research. PMLR, pp. 223–237.

Morgenstern, Oskar and John Von Neumann (1953). *Theory of games and economic behavior*. Princeton university press.

Morris, R.C. (1962). *Studies in search for a conscious evader*. Tech. rep. Massachusetts Inst. of Tech Lexington Lincoln Lab.

Moulin, Hervé and J-P Vial (1978). "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". In: *International Journal of Game Theory* 7.3-4, pp. 201–221.

Myerson, Roger B (1991). *Game Theory: Analysis of Conflict*. Harvard University Press.

— (1993). "Incentives to cultivate favored minorities under alternative electoral systems". In: *American Political Science Review* 87.4, pp. 856–869.

Nakai, Teruhisa (1986). "A search game with one object and two searchers". In: *Journal of applied probability* 23.3, pp. 696–707.

Nash, John (1950). "Equilibrium points in N-person games". In: *Proceedings of the national academy of sciences* 36.1, pp. 48–49.

— (1951). "Non-cooperative games". In: *Annals of mathematics*, pp. 286–295.

— (1953). "Two-person cooperative games". In: *Econometrica: Journal of the Econometric Society*, pp. 128–140.

Navda, Vishnu, Aniruddha Bohra, Samrat Ganguly, and Dan Rubenstein (2007). "Using channel hopping to increase 802.11 resilience to jamming attacks". In: *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. IEEE, pp. 2526–2530.

Nemirovski, A.S. and D.B. Yudin (1983). *Problem Complexity and Method Efficiency in Optimization*. A Wiley-Interscience publication. Wiley.

Nesterov, Y. and A. Nemirovsky (1994). *Interior-pointpolynomial methods in convexprogramming*. Tech. rep. Philadelphia, PA,

Neu, Gergely (2015). "Explore no more: Improved high-probability regret bounds for non-stochastic bandits". In: *Advances in Neural Information Processing Systems*, pp. 3168–3176.

Neu, Gergely and Gábor Bartók (2013). "An efficient algorithm for learning with semi-bandit feedback". In: *Proceedings of the 24th International Conference on Algorithmic Learning Theory (ALT)*. Springer, pp. 234–248.

Osborne, Martin J and Ariel Rubinstein (1994). *A course in game theory*. MIT press.

Osório, António (2013). "The lottery Blotto game". In: *Economics Letters* 120.2, pp. 164–166.

Paarporn, Keith, Rahul Chandan, Mahnoosh Alizadeh, and Jason R Marden (2019). "Characterizing the interplay between information and strength in Blotto games". In: *arXiv preprint arXiv:1909.03382*.

Palaiopanos, Gerasimos, Ioannis Panageas, and Georgios Piliouras (2017). "Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos". In: *Advances in Neural Information Processing Systems*, pp. 5872–5882.

Pastine, Ivan and Tuvana Pastine (2012). "Incumbency advantage and political campaign spending limits". In: *Journal of Public Economics* 96.1-2, pp. 20–32.

Perchet, Vianney (2011). "Internal regret with partial monitoring: Calibration-based optimal algorithms". In: *Journal of Machine Learning Research* 12.Jun, pp. 1893–1921.

Pita, James, Milind Tambe, Christopher Kiekintveld, Shane Cullen, and Erin Steigerwald (2011). "GUARDS—innovative application of game theory for national airport security". In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.

Powell, Robert (2009). "Sequential, nonzero-sum "Blotto": Allocating defensive resources prior to attack". In: *Games and Economic Behavior* 67.2, pp. 611–615.

Primo, David M (2007). "A comment on Baron and Ferejohn (1989): The open rule equilibrium and coalition formation". In: *Public Choice* 130.1-2, pp. 129–135.

Rakhlin, Alexander, Jacob Abernethy, A Agarwal, P Bartlett, E Hazan, and A Tewari (2009). *Lecture notes on online learning*.

Rakhlin, Alexander and Karthik Sridharan (2016). "BISTRO: An Efficient Relaxation-Based Method for Contextual Bandits". In: *Proceedings of The 33rd International Conference on Machine Learning (ICML)*, pp. 1977–1985.

Reichert, Corinne (2020). "Europe allows Huawei for 5G through security guidelines". In: *Cnet*.

Richardson, Matt and Pedro Domingos (2002). "Mining knowledge-sharing sites for viral marketing". In: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 61–70.

Rinott, Yosef, Marco Scarsini, and Yaming Yu (2012). "A Colonel Blotto gladiator game". In: *Mathematics of Operations Research* 37.4, pp. 574–590.

Roberson, Brian (2006). "The Colonel Blotto Game". In: *Economic Theory* 29.1, pp. 1–24.

Roberson, Brian and Dmitriy Kvasov (2012). "The non-constant-sum Colonel Blotto game". In: *Economic Theory* 51.2, pp. 397–433.

Robson, ARW (2005). *Multi-item contests. Australian National University*. Tech. rep. Working Paper.

Sakaue, Shinsaku, Masakazu Ishihata, and Shin-ichi Minato (2018). "Efficient Bandit Combinatorial Optimization Algorithm with Zero-suppressed Binary Decision Diagrams". In: *International Conference on Artificial Intelligence and Statistics*, pp. 585–594.

Savage, Leonard J (1972). *The foundations of statistics*. Courier Corporation.

Schwartz, Galina, Patrick Loiseau, and Shankar S Sastry (2014). "The heterogeneous Colonel Blotto game". In: *Proceedings of the 7th International Conference on Network Games, Control and Optimization (NetGCoop)*, pp. 232–238.

Shubik, Martin and Robert James Weber (1981). "Systems defense games: Colonel Blotto, command and control". In: *Naval Research Logistics Quarterly* 28.2, pp. 281–287.

Siegel, Ron (2009). "All-pay contests". In: *Econometrica* 77.1, pp. 71–92.

— (2014). "Asymmetric contests with head starts and nonmonotonic costs". In: *American Economic Journal: Microeconomics* 6.3, pp. 59–105.

Skaperdas, Stergios (1996). "Contest success functions". In: *Economic theory* 7.2, pp. 283–290.

Sleator, Daniel Dominic and Robert Endre Tarjan (1985). "Self-adjusting binary search trees". In: *Journal of the ACM (JACM)* 32.3, pp. 652–686.

Snyder, James M (1989). "Election goals and the allocation of campaign resources". In: *Econometrica: Journal of the Econometric Society*, pp. 637–660.

Stevis-Gridneff, Matina (2020). "E.U. Recommends Limiting, but Not Banning, Huawei in 5G Rollout". In: *The New York Times*.

Stone, Lawrence D (1976). *Theory of optimal search*. Elsevier.

Subelman, Eduardo J (1981). "A hide–search game". In: *Journal of Applied Probability* 18.3, pp. 628–640.

Takimoto, Eiji and Manfred K Warmuth (2003). "Path kernels and multiplicative updates". In: *Journal of Machine Learning Research* 4.Oct, pp. 773–818.

Tharumarajah, A. (2001). "Survey of resource allocation methods for distributed manufacturing systems". In: *Production Planning & Control* 12.1, pp. 58–68.

Thomas, Caroline (2017). "N-dimensional Blotto game with heterogeneous battlefield values". In: *Economic Theory*, pp. 1–36.

Thompson, William R (1933). "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". In: *Biometrika* 25.3/4, pp. 285–294.

Tran-Thanh, Long, Archie Chapman, Alex Rogers, and Nicholas R Jennings (2012). "Knapsack based optimal policies for budget–limited multi–armed bandits". In: *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*.

TSA (2011). "TSA — Transportation Security Administration — U.S. Department of Homeland Security". In: *TSA*.

Tullock, Gordon (1980). "Efficient Rent Seeking". In: *Lockard A.A., Tullock G. (eds) (2001). Efficient Rent Seeking Chronicle of an Intellectual Quagmire.* Springer, Boston, MA (Preprint of Tullock, Gordon (1980) in Toward a Theory of the Rent-Seeking Society. Efficient Rent Seeking.)

Uchiya, Taishi, Atsuyoshi Nakamura, and Mineichi Kudo (2010). "Algorithms for adversarial bandit problems with multiple plays". In: *Proceedings of the 21st International Conference on Algorithmic Learning Theory (ALT)*. Springer, pp. 375–389.

Vaart, A. W. van der (1998). *Asymptotic Statistics*. Cambridge University Press.

Van der Vaart, Aad W (2000). *Asymptotic statistics*. Vol. 3. Cambridge university press.

Vandenberghe, Lieven and Stephen Boyd (1996). "Semidefinite programming". In: *SIAM review* 38.1, pp. 49–95.

Vidal, Rene, Omid Shakernia, H Jin Kim, David Hyunchul Shim, and Shankar Sastry (2002). "Probabilistic pursuit-evasion games: theory, implementation, and experimental evaluation". In: *IEEE transactions on robotics and automation* 18.5, pp. 662–669.

Viro, O Ya, OA Ivanov, N Yu Netsvetaev, and VM Kharlamov (2008). *Elementary topology*. American Mathematical Soc.

Von Neumann, John (1928). "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1, pp. 295–320.

— (1953). "A certain zero-sum two-person game equivalent to the optimal assignment problem". In: *Contributions to the Theory of Games* 2.0, pp. 5–12.

Wang, Qingsi and Mingyan Liu (2016). "Learning in hide-and-seek". In: *IEEE/ACM Transactions on Networking* 24.2, pp. 1279–1292.

Warmuth, Manfred K, Wouter M Koolen, and David P Helmbold (2011). "Combining Initial Segments of Lists". In: *Proceedings of the 22nd International Conference on Algorithmic Learning Theory (ALT)*. ALT'11. Springer-Verlag, pp. 219–233.

Warmuth, Manfred K and Dima Kuzmin (2008). "Randomized online PCA algorithms with regret bounds that are logarithmic in the dimension". In: *Journal of Machine Learning Research* 9.Oct, pp. 2287–2320.

Weinstein, J (2005). *Two notes on the Blotto game. Northwestern University*.

Wittman, Michael D (2011). *Solving the Blotto game: A computational approach*. Tech. rep. MIT working paper.

El-Yaniv, Ran (1998). "Competitive solutions for online financial problems". In: *ACM Computing Surveys (CSUR)* 30.1, pp. 28–69.

Yucek, Tevfik and Huseyin Arslan (2009). "A survey of spectrum sensing algorithms for cognitive radio applications". In: *IEEE communications surveys & tutorials* 11.1, pp. 116–130.

Zermelo, Ernst (1913). "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels". In: *Proceedings of the 5th international congress of mathematicians*. Vol. 2, pp. 501–504.

Appendix

Appendix A

Supplementary Materials for Chapter 4 on the $\mathcal{CB}_n$ Game

A.1 Preliminary Lemmas

Lemma A.1. *Given a game $\mathcal{CB}_n$ (or $\mathcal{LB}_n$), for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we have:*

(i) $\lambda^*_A, \lambda^*_B > 0$ *and* $\gamma^* = \lambda^*_A/\lambda^*_B$.

(ii) *For any* $i \in [n]$, *we have* $\mathbb{E}[A^S_{\gamma^*,i}] = \frac{1}{2}\frac{v^B_i}{\lambda^*_B}$, $\mathbb{E}[A^W_{\gamma^*,i}] = \left(\frac{v^A_i}{\lambda^*_A}\right)^2 \frac{\lambda^*_B}{2v^B_i}$, $\mathbb{E}[B^S_{\gamma^*,i}] = \frac{1}{2}\frac{v^A_i}{\lambda^*_A}$ *and* $\mathbb{E}[B^W_{\gamma^*,i}] = \left(\frac{v^B_i}{\lambda^*_B}\right)^2 \frac{\lambda^*_A}{2v^A_i}$.

(iii) $X^A = \sum_{i\in[n]} \mathbb{E}[A^*_i]$ *and* $X^B = \sum_{i\in[n]} \mathbb{E}[B^*_i]$.

(iv) *For any* $i \in [n]$, A^*_i *and* B^*_i *have a constant upper-bound; particularly,*

$$\mathbb{P}\left(A^*_i \le 2X^B\right) = \mathbb{P}\left(B^*_i \le 2X^B\right) = 1.$$

Proof.

(i) The positivity of λ^*_A and λ^*_B follows from the positivity of γ^* and the definitions of λ^*_A and λ^*_B in (4.6) and (4.7). By dividing (4.6) by (4.7) and combining with (4.5), we trivially have that $\gamma^* = \lambda^*_A/\lambda^*_B$.

(ii) These results come directly from the definitions of the distributions $F_{A^S_{\gamma^*,i}}$, $F_{A^W_{\gamma^*,i}}$ $F_{B^S_{\gamma^*,i}}$ and $F_{B^W_{\gamma^*,i}}$.

(iii) We multiply both sides of (4.7) by X^A/λ_B^* and both sides of (4.6) by X^B/λ_A^* then using the fact that $\gamma^* = \lambda_A^*/\lambda_B^*$ to obtain the following:

$$X^A = \sum_{j\in\Omega_A(\gamma^*)} \frac{1}{2}\frac{v_j^B}{\lambda_B^*} + \sum_{j\notin\Omega_A(\gamma^*)} \left(\frac{v_j^A}{\lambda_A^*}\right)^2 \frac{\lambda_B^*}{2v_j^B}, \tag{10}$$

$$X^B = \sum_{j\in\Omega_A(\gamma^*)} \left(\frac{v_j^B}{\lambda_B^*}\right)^2 \frac{\lambda_A^*}{2v_j^A} + \sum_{j\notin\Omega_A(\gamma^*)} \frac{1}{2}\frac{v_j^A}{\lambda_A^*}. \tag{11}$$

Combining with (ii), we deduce that $X^A = \sum_{i\in[n]} \mathbb{E}[A_i^*]$ and $X^B = \sum_{i\in[n]} \mathbb{E}[B_i^*]$.

(iv) If $i \in \Omega_A(\gamma^*)$, we have $A_i^* = A_{\gamma^*,i}^S$ and $B_i^* = B_{\gamma^*,i}^W$. Recalling Definition 4.1.4, we have that $\mathbb{P}\left(A_i^S \le v_i^B/\lambda_B^*\right) = 1$ and $\mathbb{P}\left(B_i^W \le v_i^B/\lambda_B^*\right) = 1$. On the other hand, from (10), we deduce

$$X^B \ge X^A \ge \sum_{j\in\Omega_A(\gamma^*)} \frac{v_j^B}{2\lambda_B^*} \ge \frac{v_i^B}{2\lambda_B^*}.$$

Therefore,

$$\mathbb{P}(A_i^S \le 2X^B) \ge \mathbb{P}\left(A_i^S \le v_i^B/\lambda_B^*\right) = 1,$$
$$\text{and } \mathbb{P}(B_i^W \le 2X^B) \ge \mathbb{P}(B_i^W \le v_i^B/\lambda_B^*) = 1.$$

We conclude that for any $i \in \Omega_A(\gamma^*)$, A_i^*, B_i^* are bounded by $2X^B$. If $i \notin \Omega_A(\gamma^*)$, we have $A_i^* = A_{\gamma^*,i}^W$ and $B_i^* = B_{\gamma^*,i}^S$. Recalling Definition 4.1.4, we have that $\mathbb{P}\left(A_i^W \le v_i^A/\lambda_A^*\right) = 1$ and $\mathbb{P}\left(B_i^S \le v_i^A/\lambda_A^*\right) = 1$. On the other hand, from (11), we deduce

$$X^B \ge \sum_{j\notin\Omega_A(\gamma^*)} \frac{v_j^A}{2\lambda_A^*} \ge \frac{v_i^A}{2\lambda_A^*}.$$

Therefore,

$$\mathbb{P}(A_i^W \le 2X^B) \ge \mathbb{P}\left(A_i^W \le v_i^A/\lambda_A^*\right) = 1,$$
$$\text{and } \mathbb{P}(B_i^S \le 2X^B) \ge \mathbb{P}(B_i^S \le v_i^A/\lambda_A^*) = 1.$$

We conclude that for $i \notin \Omega_A(\gamma^*)$, A_i^*, B_i^* are also bounded by $2X^B$.

□

Proposition 4.1.6. *Under Assumption $(A0)$, for any game $\mathcal{CB}_n$, there exist positive constants $\underline{\gamma}, \bar{\gamma}, \underline{\lambda}, \bar{\lambda}$, that do not depend on n, such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and its corresponding λ_A^*, λ_B^*, we have $\underline{\gamma} \le \gamma^* \le \bar{\gamma}$ and $\underline{\lambda} \le \lambda_A^*, \lambda_B^* \le \bar{\lambda}$.*

Proof. Let $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we consider the following cases:

Case 1: If $0 < \gamma^* < \min_{i\in[n]}\left\{\frac{v_i^A}{v_i^B}\right\}$**.** In this case, $\Omega_A(\gamma^*) = [n]$, and since γ^* is a solution of (4.5), we deduce:

$$\gamma^* = \frac{X^B}{X^A}\frac{\sum_{i=1}^n v_i^B}{\sum_{i=1}^n \frac{(v_i^B)^2}{v_i^A}} \geq \frac{X^B}{X^A}\frac{n\frac{\underline{w}}{n\overline{w}}}{n\frac{\left(\frac{\overline{w}}{n\underline{w}}\right)^2}{\frac{\underline{w}}{n\overline{w}}}} = \frac{X^B}{X^A}\left(\frac{\underline{w}}{\overline{w}}\right)^4.$$

Here, the inequality comes directly from (4.1).

Case 2: If $\gamma^* \geq \max_{i\in[n]}\left\{\frac{v_i^A}{v_i^B}\right\}$**.** In this case, $\Omega_A(\gamma^*) = \emptyset$, and since γ^* is a solution of (4.5), we deduce:

$$\gamma^* = \frac{X^B}{X^A}\frac{\sum_{i=1}^n \frac{(v_i^A)^2}{v_i^B}}{\sum_{i=1}^n v_i^A} \leq \frac{X^B}{X^A}\left(\frac{\overline{w}}{\underline{w}}\right)^4.$$

Case 3: If $\exists i, j : \frac{v_i^A}{v_i^B} \leq \gamma^* < \frac{v_j^A}{v_j^B}$**.** In this case, we have $\gamma^* \in \left[\left(\frac{\underline{w}}{\overline{w}}\right)^2, \left(\frac{\overline{w}}{\underline{w}}\right)^2\right]$; this is trivially deduced from (4.1), .

In conclusion, we conclude the bounds of γ^* as in the statement of Proposition 4.1.6; here, we denote $\bar{\gamma} := \max\left\{\frac{X^B}{X^A}\left(\frac{\overline{w}}{\underline{w}}\right)^4, \left(\frac{\overline{w}}{\underline{w}}\right)^2\right\} = \frac{X^B}{X^A}\left(\frac{\overline{w}}{\underline{w}}\right)^4$ and $\underline{\gamma} := \min\left\{\frac{X^B}{X^A}\left(\frac{\underline{w}}{\overline{w}}\right)^4, \left(\frac{\underline{w}}{\overline{w}}\right)^2\right\}$.

On the other hand, from the definition of λ_A^* in (4.6), we deduce

$$\begin{aligned}
\lambda_A^* &\geq \frac{(\gamma^*)^2}{2X^B}\sum_{i\in\Omega_A(\gamma^*)}\left(\frac{\underline{w}}{n\overline{w}}\right)^2\frac{1}{\frac{\overline{w}}{n\underline{w}}} + \frac{1}{2X^B}\sum_{i\notin\Omega_A(\gamma^*)}\frac{\underline{w}}{n\overline{w}}\\
&\geq \min\left\{\frac{(\gamma^*)^2}{2X^B}, \frac{1}{2X^B}\right\}\cdot\sum_{i\in[n]}\frac{1}{n}\left(\frac{\underline{w}}{\overline{w}}\right)^3\\
&\geq \min\left\{\frac{(\gamma^*)^2}{2X^B}, \frac{1}{2X^B}\right\}\cdot\left(\frac{\underline{w}}{\overline{w}}\right)^3.
\end{aligned}$$

Similarly, we have the upper-bound

$$\lambda_A^* \leq \max\left\{\frac{(\gamma^*)^2}{2X^B}, \frac{1}{2X^B}\right\}\cdot\left[\sum_{i\in\Omega_A(\gamma^*)}\frac{1}{n}\left(\frac{\overline{w}}{\underline{w}}\right)^3 + \sum_{i\notin\Omega_A(\gamma^*)}\frac{1}{n}\left(\frac{\overline{w}}{\underline{w}}\right)^3\right] = \max\left\{\frac{(\gamma^*)^2}{2X^B}, \frac{1}{2X^B}\right\}\cdot\left(\frac{\overline{w}}{\underline{w}}\right)^3.$$

Similarly, we can prove that $\min\left\{\frac{1}{2X^A}, \frac{1}{2\gamma^{*2}X^A}\right\}\left(\frac{\underline{w}}{\overline{w}}\right)^3 \leq \lambda_B^* \leq \max\left\{\frac{1}{2X^A}, \frac{1}{2\gamma^{*2}X^A}\right\}\left(\frac{\overline{w}}{\underline{w}}\right)^3$; therefore,

$$\min\left\{\frac{\gamma^{*2}}{2X^B}, \frac{1}{2X^B}, \frac{1}{2X^A}, \frac{1}{2(\gamma^*)^2X^A}\right\}\left(\frac{\underline{w}}{\overline{w}}\right)^3 \leq \lambda_A^*, \lambda_B^* \leq \max\left\{\frac{\gamma^{*2}}{2X^B}, \frac{1}{2X^B}, \frac{1}{2X^A}, \frac{1}{2(\gamma^*)^2X^A}\right\}\left(\frac{\overline{w}}{\underline{w}}\right)^3.$$

Since $\gamma^* \in [\underline{\gamma}, \bar{\gamma}]$, λ_A^* and λ_B^* are bounded in $[\underline{\lambda}, \bar{\lambda}]$, where

$$\underline{\lambda} := \min\left\{\frac{\underline{\gamma}^2}{2X^B}, \frac{1}{2X^B}, \frac{1}{2X^A}, \frac{1}{2\bar{\gamma}^2X^A}\right\}\left(\frac{\underline{w}}{\overline{w}}\right)^3,$$

$$\bar{\lambda} := \max\left\{\frac{\bar{\gamma}^2}{2X^B}, \frac{1}{2X^B}, \frac{1}{2X^A}, \frac{1}{2\underline{\gamma}^2 X^A}\right\}\left(\frac{\bar{w}}{\underline{w}}\right)^3.$$

□

Finally, we prove a trivial result that will be used quite often in the remainder of this section.

Lemma A.2. *For any* $\hat{\varepsilon} > 0$ *and* $\hat{C} \geq 1$*, we have that* $(\ln(\hat{C}) + 1)\ln\left(\frac{1}{\min\{\hat{\varepsilon}, 1/e\}}\right) \geq \ln\left(\frac{\hat{C}}{\hat{\varepsilon}}\right)$.

Proof. Case 1: If $\hat{\varepsilon} < 1/e$. In this case, we have $\ln(1/\hat{\varepsilon}) > 1$; therefore,

$$(\ln(\hat{C})+1)\ln\left(\frac{1}{\min\{\hat{\varepsilon}, 1/e\}}\right) = (\ln(\hat{C}) + 1)\ln\left(\frac{1}{\hat{\varepsilon}}\right) = \ln(\hat{C})\ln\left(\frac{1}{\hat{\varepsilon}}\right) + \ln\left(\frac{1}{\hat{\varepsilon}}\right) > \ln\left(\frac{\hat{C}}{\hat{\varepsilon}}\right).$$

Case 2: If $\hat{\varepsilon} \geq 1/e$. We have $\ln(1/\hat{\varepsilon}) \leq 1$; therefore,

$$(\ln(\hat{C})+1)\ln\left(\frac{1}{\min\{\hat{\varepsilon}, 1/e\}}\right) = (\ln(\hat{C})+1)\ln\left(\frac{1}{1/e}\right) = \ln(\hat{C})+1 \geq \ln(\hat{C})+\ln\left(\frac{1}{\hat{\varepsilon}}\right) = \ln\left(\frac{\hat{C}}{\hat{\varepsilon}}\right).$$

□

A.2 Proof of Theorem 4.2.3

First note that in the remainders of this section, for any bounded, non-negative random variable Z (i.e., $\exists C > 0 : \mathbb{P}(Z \in [0, C]) = 1$), any measurable function g on $\mathbb{R}$, we write $\int_0^\infty g(x)\mathrm{d}F_Z(x)$ instead of $\int_0^C g(x)\mathrm{d}F_Z(x)$ if there is no need to emphasize the bounds of Z. For the sake of notation, we also denote by $A_{=0}$ the event $\left\{\sum_{j\in[n]} A_j^* = 0\right\}$ and by $A_{>0}$ its complement event, that is $\left\{\sum_{j\in[n]} A_j^* > 0\right\}$. Similarly, we denote by $B_{=0}$ the event $\left\{\sum_{j\in[n]} B_j^* = 0\right\}$ and by $B_{>0}$ the event $\left\{\sum_{j\in[n]} B_j^* > 0\right\}$.

Recall the notation $F_{A_i^n}$ and $F_{B_i^n}$ as the univariate marginal distributions corresponding to battlefield $i \in [n]$ of the $\text{IU}_A^{\gamma^*}$ and $\text{IU}_B^{\gamma^*}$ strategies (the corresponding random variables are denoted A_i^n and B_i^n).From the definition of the IU^{γ^*} strategy (via Algorithm 5), for any $x \geq 0$ and $i \in [n]$, we have:

$$\begin{aligned} F_{A_i^n}(x) &= \mathbb{P}\left(\left\{A_i^n \leq x\right\} \bigcap A_{=0}\right) + \mathbb{P}\left(\left\{A_i^n \leq x\right\} \bigcap A_{>0}\right) \\ &= \mathbb{P}\left(A_{=0}\right) + \mathbb{P}\left(\left\{\frac{A_i^* \cdot X^A}{\sum_{j\in[n]} A_j^*} \leq x\right\} \bigcap A_{>0}\right). \end{aligned} \tag{12}$$

Here, we have used the fact that if $\sum_{j\in[n]} A_j^* = 0$ (i.e., when $A_{=0}$ happens), then $A_i^n = 0$ by definition and thus, $\mathbb{P}(A_i^n \leq x) = 1$ and $\mathbb{P}\left(\left\{A_i^n \leq x\right\} \cap A_{=0}\right) = \mathbb{P}(A_{=0})$. Similarly to (12), for any $x \geq 0$ and $i \in [n]$,

$$F_{B_i^n}(x) = \mathbb{P}\left(B_{=0}\right) + \mathbb{P}\left(\left\{\frac{B_i^* \cdot X^B}{\sum_{j\in[n]} B_j^*} \leq x\right\} \bigcap B_{>0}\right). \tag{13}$$

For the random variables A_i^n and B_i^n ($i \in [n]$), we prepare a lemma stating several useful results as follows (its proof is given in Appendix A.3).

Lemma A.3. *For any n and $i \in [n]$, we have*

(i) $\mathbb{P}(A_i^n = 0) = \mathbb{P}(A_i^* = 0)$ *and* $\mathbb{P}(B_i^n = 0) = \mathbb{P}(B_i^* = 0)$.

(ii) $\mathbb{P}(A_i^n = x) = \mathbb{P}(B_i^n = y) = 0$ *for any* $x \in (0, \infty)\backslash\{X^A\}$ *and* $y \in (0, \infty)\backslash\{X^B\}$.

(iii) $\mathbb{P}(A_i^n = X^A) \leq \left(1 - \frac{\lambda}{\overline{\lambda}} \frac{\underline{w}^2}{\overline{w}^2}\right)^{n-1}$ *and* $\mathbb{P}(B_i^n = X^B) \leq \left(1 - \frac{\lambda}{\overline{\lambda}} \frac{\underline{w}^2}{\overline{w}^2}\right)^{n-1}$.

Intuitively, Result (ii) states that the function $F_{A_i^n}$ (resp. $F_{B_i^n}$) is continuous on $(0, X^A)$ (resp. $(0, X^B)$). The discontinuity of $F_{A_i^n}$ (resp. $F_{B_i^n}$) at X^A (resp. at X^B) is due to the normalization step involved in the definition of the IU$^{\gamma^*}$ strategy; note that the probability that $A_i^n = X^A$ (resp. $B_i^n = X^B$) quickly tends to zero when n increases as has been shown in Result (iii). Finally, Result (i) shows that in some cases, $F_{A_i^n}$ and $F_{B_i^n}$ may be discontinuous at 0. This is due to the fact that the functions $F_{A_i^*}$ and $F_{B_i^*}$ may be discontinuous at 0. Moreover, recall that we chose the assignments of the outputs in line 3 and 7 of Algorithm 5 to be allocating zero to every battlefield, i.e., the mass at 0 of $F_{A_i^n}$ and $F_{B_i^n}$ is added by a (negligibly small) positive probability. While other assignments do not affect our results, they make $F_{A_i^n}$ (resp. $F_{B_i^n}$) be discontinuous at some points differing from 0 and X^A (resp. X^B), e.g., if in line 3 of Algorithm 5, we assign $x_i^A = X^A/n$, the distribution $F_{A_i^n}$ would also be discontinuous at the point X^A/n. Our choice of assignments provides more convenience in our analysis since we have to consider their discontinuity at 0 in any case.

Finally, with all the preparation steps mentioned above, we are ready to prove Theorem 4.2.3.

Theorem 4.2.3.

(i) In any game $\mathcal{CB}_n$, there exists a positive number $\varepsilon = \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the following inequalities hold for any pure strategy $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ of players A and B:

$$\Pi^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*}) \leq \Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + \varepsilon W^A, \tag{4.13}$$

$$\Pi^B(\mathrm{IU}_A^{\gamma^*}, \boldsymbol{x}^B) \leq \Pi^B(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + \varepsilon W^B. \tag{4.14}$$

(ii) There exists a constant $C^* > 0$ such that for any $\varepsilon \in (0, 1]$ and in any game $\mathcal{CB}_n$ with $n \geq C^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, (4.13) and (4.14) hold for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, any pure strategy $\boldsymbol{x}^A$, $\boldsymbol{x}^B$ of players A and B.

Proof. In this section, we first give a proof of Result (ii) of Theorem 4.2.3. Result (i) will be deduced from (ii). We first look for the condition on n such that (4.13) holds for any pure strategy $\boldsymbol{x}^A$ of player A. The proof that (4.14) holds for any pure strategy of player B under the same condition can be done similarly and thus is omitted.

First, we write explicitly the payoffs of player A when player B plays the IU$_B^{\gamma^*}$ strategy and player A plays either the pure strategy $\boldsymbol{x}^A$ or the IU$_A^{\gamma^*}$ strategy:

$$\Pi^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*}) = \alpha \sum_{i=1}^{n} w_i^A \mathbb{P}(B_i^n = x_i^A) + \sum_{i=1}^{n} w_i^A \mathbb{P}(B_i^n < x_i^A), \tag{14}$$

$$\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) = \alpha \sum_{i=1}^n w_i^A \mathbb{P}(B_i^n = A_i^n) + \sum_{i=1}^n w_i^A \mathbb{P}(B_i^n < A_i^n)$$
$$= \alpha \sum_{i=1}^n \int_0^\infty w_i^A \mathbb{P}(B_i^n = x) \mathrm{d}F_{A_i^n}(x) + \sum_{i=1}^n \int_0^\infty w_i^A \mathbb{P}(B_i^n < x) \mathrm{d}F_{A_i^n}(x). \quad (15)$$

We then prepare a useful lemma, its proof is given in Appendix A.4. Intuitively, this lemma shows that as n is large enough, we can prove (4.13) without the need of analyzing separately the case where players get tie allocations (that is our results hold regardless of the tie-breaking-rule parameter α).

Lemma A.4. *Given $\varepsilon \in (0,1]$, there exists a constant $C_0^* > 0$ (that does not depend on ε) such that for any $n \geq C_0^* \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, for any game $\mathcal{CB}_n$ and $\gamma^* \in \mathcal{S}_n^{(4.5)}$ the following inequality is a sufficient condition of (4.13):*

$$\sum_{i=1}^n v_i^A F_{B_i^n}\left(x_i^A\right) \leq \sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) + \frac{\varepsilon}{2}. \quad (16)$$

In the remainders of the proof, we focus on (16) and look for the condition of n such that it holds; this will be done in the following five steps. After that, from Lemma A.4, we can conclude that (4.13) also holds with the corresponding condition on n.

Step 1: Prove that $\{F_{A_i^*}\}_i$ is optimal against $\{F_{B_i^*}\}_i$.

Lemma A.5. *In any game $\mathcal{CB}_n$, for any pure strategy $\boldsymbol{x}^A$ of player A and $\gamma^* \in \mathcal{S}_n^{(4.5)}$, we have*

$$\sum_{i=1}^n v_i^A F_{B_i^*}\left(x_i^A\right) \leq \sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^*}(x) \mathrm{d}F_{A_i^*}(x). \quad (17)$$

The proof of Lemma A.5 is given in Appendix A.5. This lemma can be interpreted as follows: if the allocation of player B to battlefield i follows the distribution $F_{B_i^*}$, then it is optimal for player A to play such that her allocation at this battlefield follows $F_{A_i^*}$ (we do not know if it is possible to construct a mixed strategy such that player A's allocation at battlefield i follows $F_{A_i^*}$ for all $i \in [n]$; however, this does not affect our results in this work). Using this lemma, we will analyze the validity of (16) by proving that, as $n \to \infty$, the terms in (16) respectively converge toward the terms in (17). To do this, we consider the next step.

Step 2: Prove that $F_{A_i^n}$ and $F_{B_i^n}$ uniformly converge toward $F_{A_i^*}$ and $F_{B_i^*}$ as n increases.

Lemma A.6. *For any $\varepsilon_1 \in (0,1]$, there exists $C_1 > 0$ (that does not depend on ε_1) such that for any $n \geq C_1 \varepsilon_1^{-2} \ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right)$ and $i \in [n]$,*

$$\sup_{x \in [0,\infty)} \left|F_{A_i^n}(x) - F_{A_i^*}(x)\right| \leq \varepsilon_1 \quad \textit{and} \quad \sup_{x \in [0,\infty)} \left|F_{B_i^n}(x) - F_{B_i^*}(x)\right| \leq \varepsilon_1. \quad (18)$$

A proof of this lemma is given in Appendix A.6. The main intuition of this result comes from the fact that A^n_i (resp. B^n_i) is the normalization of A^*_i, $i \in [n]$ (except for the special cases of the events $A_{=0}$ and $B_{=0}$) and the use of concentration inequalities on the random variables $\sum_{j\in[n]} A^*_j$ (and $\sum_{j\in[n]} B^*_j$). In this work, we apply the Hoeffding's inequality (Theorem 2, Hoeffding (1963)) to obtain the rate of convergence indicated here in Lemma A.6.

Step 3: Prove that the left-hand-side of (16) converges toward the left-hand-side of (17). Take C_1 as indicated in Lemma A.6, we define $C^*_1 := 16 C_1(\ln(4)+1)$ and deduce that $\frac{C^*_1}{\varepsilon^2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1\left(\frac{4}{\varepsilon}\right)^2 \ln\left(\frac{1}{\min\{\frac{\varepsilon}{4},\frac{1}{e}\}}\right)$.[13] Therefore, take $\varepsilon_1 := \varepsilon/4$, for any $n \geq C^*_1\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have $n \geq C_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$; apply Lemma A.6, for any pure strategy $\boldsymbol{x}^A$ of player A, we have

$$\begin{aligned}\left|\sum_{i=1}^n v^A_i F_{B^n_i}(x^A_i) - \sum_{i=1}^n v^A_i F_{B^*_i}(x^A_i)\right| &\leq \sum_{i=1}^n v^A_i \sup_{x\in[0,\infty)}\left|F_{B^n_i}(x) - F_{B^*_i}(x)\right| \\ &\leq \sum_{i=1}^n v^A_i \frac{\varepsilon}{4} = \frac{\varepsilon}{4}. \end{aligned} \tag{19}$$

Step 4: Prove that the right-hand-side of (16) converges toward the right-hand-side of (17). We consider the difference of the involved terms as follows.

$$\begin{aligned}&\left|\sum_{i=1}^n \int_0^\infty v^A_i F_{B^n_i}(x)\mathrm{d}F_{A^n_i}(x) - \sum_{i=1}^n \int_0^\infty v^A_i F_{B^*_i}(x)\mathrm{d}F_{A^*_i}(x)\right| \\ \leq &\sum_{i=1}^n v^A_i \int_0^\infty \left|F_{B^n_i}(x) - F_{B^*_i}(x)\right| \mathrm{d}F_{A^n_i}(x) + \sum_{i=1}^n v^A_i\left|\int_0^\infty F_{B^*_i}(x)\mathrm{d}F_{A^n_i}(x) - \int_0^\infty F_{B^*_i}(x)\mathrm{d}F_{A^*_i}(x)\right|.\end{aligned} \tag{20}$$

Let us define $C^*_2 := C_1 \cdot 64(\ln(8)+1)$ (again, C_1 is the constant indicated in Lemma A.6), we have that $C^*_2\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1\left(\frac{8}{\varepsilon}\right)^2\ln\left(\frac{1}{\min\{\frac{\varepsilon}{8},\frac{1}{e}\}}\right)$.[14] Therefore, take $\varepsilon_1 := \varepsilon/8$, for any $n \geq C^*_2\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have $n \geq \varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$ and by Lemma A.6, we have

$$\sum_{i=1}^n v^A_i \int_0^\infty \left|F_{B^n_i}(x) - F_{B^*_i}(x)\right|\mathrm{d}F_{A^n_i}(x) \leq \sum_{i=1}^n v^A_i \int_0^\infty \frac{\varepsilon}{8}\mathrm{d}F_{A^n_i}(x) = \sum_{i=1}^n v^A_i\frac{\varepsilon}{8}. \tag{21}$$

[13] This is due to $C^*_1 \cdot \varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) = C_1\left(\frac{4}{\varepsilon}\right)^2 \cdot (\ln(4)+1)\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1 \cdot \left(\frac{4}{\varepsilon}\right)^2\ln\left(\frac{4}{\varepsilon}\right)$; here, we have applied Lemma A.2 with $\hat{\varepsilon} := \varepsilon$ and $\hat{C} := 4$; moreover, $\frac{\varepsilon}{4} = \min\{\frac{\varepsilon}{4},\frac{1}{e}\}$ since $\varepsilon \leq 1$; thus, we can rewrite $\ln\left(\frac{4}{\varepsilon}\right) = \ln\left(\frac{1}{\min\{\varepsilon/4,1/e\}}\right)$.

[14] This is due to $C^*_2 \cdot \varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) = C_1\left(\frac{8}{\varepsilon}\right)^2 \cdot (\ln(8)+1)\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1 \cdot \left(\frac{8}{\varepsilon}\right)^2\ln\left(\frac{8}{\varepsilon}\right)$; here, we have applied Lemma A.2 with $\hat{\varepsilon} := \varepsilon$ and $\hat{C} := 8$; moreover, $\frac{\varepsilon}{8} = \min\{\frac{\varepsilon}{8},\frac{1}{e}\}$ since $\varepsilon \leq 1$; thus, we can rewrite $\ln\left(\frac{8}{\varepsilon}\right) = \ln\left(\frac{1}{\min\{\varepsilon/8,1/e\}}\right)$.

Now, we need to find an upper-bound of the second term in the right-hand-side of (20). To do this, we present a lemma, called Lemma A.7 (stated below), that is based on the portmanteau lemma (see, e.g., Van der Vaart (2000)) regarding the weak convergence of a sequence of measures. Note importantly that by a direct application of the portmanteau lemma (since $F_{B_i^*}$ is Lipschitz continuous and from Lemma A.6, $F_{A_i^n}$ uniformly converges to $F_{A_i^*}$), we can prove that $\int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^n}(x)$ converges toward $\int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^*}(x)$ as $n \to \infty$; however, note that the convergence rate obtained by doing this is large due to the fact that the Lipschitz constant of $F_{B_i^*}$ (that is λ_A^*/v_i^A) increases as n increases. To obtain a better convergence rate as indicated in Lemma A.7, we exploit the properties of the involved functions that allow us to use the telescoping sum trick (see Appendix A.7 for more details).

Lemma A.7. *For any $\varepsilon_2 \in (0, 1]$, there exists a constant $C_2 > 0$ (that does not depend on ε_2) such that for any $n \geq C_2 \cdot \varepsilon_2^{-2} \ln\left(\frac{1}{\min\{\varepsilon_2, 1/e\}}\right)$ and $i \in [n]$, we have*

$$\left|\int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^n}(x) - \int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^*}(x)\right| \leq \varepsilon_2. \tag{22}$$

The proof of Lemma A.7 is given in Appendix A.7. Based on this constant C_2, we define $C_3^* := 8^2 C_2(\ln 8 + 1)$. Now, take $\varepsilon_2 := \varepsilon/8$, we have that [15]

$$C_3^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq C_2 \varepsilon_2^{-2} \ln\left(\frac{1}{\min\{\varepsilon_2, 1/e\}}\right);$$

therefore, $n \geq C_3^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \implies n \geq C_2 \varepsilon_2^{-2} \ln\left(\frac{1}{\min\{\varepsilon_2, 1/e\}}\right)$; by Lemma A.7, we deduce

$$\left|\int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^n}(x) - \int_0^\infty F_{B_i^*}(x)\mathrm{d}F_{A_i^*}(x)\right| \leq \varepsilon/8.$$

Combine this with (20) and (21), for any $n = \max\{C_2^*, C_3^*\}\varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, we have

$$\left|\sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^n}(x)\mathrm{d}F_{A_i^n}(x) - \sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^*}(x)\mathrm{d}F_{A_i^*}(x)\right| \leq \sum_{i=1}^n v_i^A \varepsilon/8 + \sum_{i=1}^n v_i^A \varepsilon/8 = \frac{\varepsilon}{4}. \tag{23}$$

<u>**Step 5:**</u> **Conclusion.** For any $n \geq \max\{C_1^*, C_2^*, C_3^*\}\varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$ and any pure strategy $\boldsymbol{x}^A$ of player A, we conclude that

$$\sum_{i=1}^n v_i^A F_{B_i^n}\left(x_i^A\right) \leq \sum_{i=1}^n v_i^A F_{B_i^*}\left(x_i^A\right) + \frac{\varepsilon}{4} \qquad \text{(from (19))}$$

[15] Once again, apply Lemma A.2, $C_3^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) = C_2\left(\frac{8}{\varepsilon}\right)^2 (\ln(8) + 1) \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq C_2\left(\frac{8}{\varepsilon}\right)^2 \ln\left(\frac{8}{\varepsilon}\right)$; moreover, we have $\varepsilon_2 := \frac{\varepsilon}{8} = \min\left\{\frac{\varepsilon}{8}, \frac{1}{e}\right\}$.

$$\leq \sum_{i=1}^{n} \int_0^\infty v_i^A F_{B_i^*}(x) \mathrm{d}F_{A_i^*}(x) + \frac{\varepsilon}{4} \qquad \text{(from (17))}$$
$$\leq \sum_{i=1}^{n} \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} \qquad \text{(from (23))}$$
$$= \sum_{i=1}^{n} \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) + \frac{\varepsilon}{2}.$$

This is exactly (16); thus, apply Lemma A.4 and denote $C^*_{(4.13)} := \max\{C^*_0, C^*_1, C^*_2, C^*_3\}$, we have proved that (4.13) holds for any $n \geq C^*_{(4.13)} \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$. Similarly, we can prove that there exists a constant $C^*_{(4.14)}$ such that (4.14) holds for any $n \geq C^*_{(4.14)} \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$. Finally, define $C^* := \max\{C^*_{(4.13)}, C^*_{(4.14)}\}$, we conclude the proof for Result (ii).

Now, to obtain Result (i), we prove that Result (ii) implies Result (i). Note that the constant C^* found in the Result (ii) does not depend on neither n nor ε. Moreover, the function

$$\xi \colon (0, \infty) \longrightarrow (0, \infty)$$
$$\tilde{\varepsilon} \mapsto C^* \tilde{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\tilde{\varepsilon}, 1/e\}}\right).$$

is continuous and increases to infinity when ε tends to zero. Therefore, for any $n \geq 1$, there exists an $\varepsilon > 0$ such that $n = C^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$. Now, apply Result (ii), (4.13) and (4.14) hold in the game $\mathcal{CB}_n$ for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and pure strategies $\boldsymbol{x}^A, \boldsymbol{x}^B$. We conclude the proof by notice that if $\varepsilon \geq 1/e$, we have $n = C^* \varepsilon^{-2}$ and thus $\varepsilon = \sqrt{n/C^*} = \mathcal{O}(n^{-1/2})$; on the other hand, if $\varepsilon < \frac{1}{e}$, we have $\ln\left(\frac{1}{\varepsilon}\right) > 1$ that induces $n = C^* \varepsilon^{-2} \ln\left(\frac{1}{\varepsilon}\right) \geq C^* \varepsilon^{-2} \geq \frac{C^*}{\varepsilon}$, thus, $\frac{1}{\varepsilon} \leq \frac{n}{C^*}$. We deduce that

$$\varepsilon = \sqrt{\frac{C^*}{n} \ln\left(\frac{1}{\varepsilon}\right)} \leq \sqrt{\frac{C^*}{n} \ln\left(\frac{n}{C^*}\right)} = \tilde{\mathcal{O}}(n^{-1/2}).$$

□

A.3 Proof of Lemma A.3

(i) Assuming $A_i^* = 0$, if $\sum_{j \neq i} A_j^* = 0$ then $A_i^n = 0$ (due to line 3 of Algorithm 5) and if $\sum_{j \neq i} A_j^* > 0$ then $A_i^n = A_i^* / \sum_{j \in [n]} A_j^* = 0$. Reversely, assuming $A_i^n = 0$, then regardless whether $\sum_{j \in [n]} A_j^* = 0$ or $\sum_{j \in [n]} A_j^* > 0$, we have $A_i^* = 0$. Therefore, $A_i^n = 0 \Leftrightarrow A_i^* = 0$ for any n and $i \in [n]$. Similarly, we can prove that

$$B_i^n = 0 \Leftrightarrow B_i^* = 0.$$

(ii) The results are trivial in cases where $x > X^A$ and $y > X^B$ due to the definition of A_i^n and B_i^n (that guarantees that with probability 1, $A_i^n \leq X^A$ and $B_i^n \leq X^B$).

In the following, we consider the case where $x \in (0, X^A)$. For any n, $i \in [n]$, we denote $Z_i := \sum_{j\neq i} A_j^*$ and obtain:

$$
\begin{aligned}
&\mathbb{P}(A_i^n = x)\\
=&\mathbb{P}\left(\{A_i^n = x\} \bigcap A_{>0}\right) && \text{(since } x > 0)\\
=&\mathbb{P}\left(\left\{A_i^* = \frac{x}{X^A} \sum\nolimits_{j\in[n]} A_j^*\right\} \bigcap A_{>0}\right)\\
=&\mathbb{P}\left(\left\{A_i^* \left(1 - \frac{x}{X^A}\right) = \frac{x}{X^A} \sum\nolimits_{j\neq i} A_j^*\right\} \bigcap A_{>0}\right)\\
=&\mathbb{P}\left(\left\{A_i^* = \frac{Z_i \cdot x}{X^A - x}\right\} \bigcap A_{>0}\right) && \text{(note that } X^A - x > 0)\\
\leq&\mathbb{P}(\{A_i^* = Z_i = 0\} \cap A_{>0}) + \int_{z>0} \mathbb{P}\left(A_i^* = \frac{z \cdot x}{X^A - x}\right) \mathrm{d}F_{Z_i}(z)\\
\leq&\mathbb{P}(A_{=0} \cap A_{>0}) + \int_{z>0} 0 \, \mathrm{d}F_{Z_i}(z)\\
=&0.
\end{aligned}
$$

Here, the second-to-last inequality comes from the fact that $\frac{zx}{X^A - x} > 0$, $\forall z > 0$, $\forall x \in (0, X^A)$ and $\mathbb{P}(A_i^* = a) = 0$ for any $a > 0$. Similarly, we can prove that $\mathbb{P}(B_i^n = y) = 0$ for any $y \in (0, X^B)$.

(iii) We have

$$
\begin{aligned}
\mathbb{P}(A_i^n = X^A) = \mathbb{P}&\left(\left\{A_i^* = \sum_{j\in[n]} A_j^*\right\} \bigcap A_{>0}\right)\\
&\leq \mathbb{P}\left(\sum_{j\neq i} A_j^* = 0\right)\\
&= \prod_{j\neq i} \mathbb{P}\left(A_j^* = 0\right).
\end{aligned}
$$

Here, the last equality comes from the fact that $A_j^*, j \in [n]$ are non-negative and independent.

Now, if there exists $j \neq i$ such that $j \in \Omega_A(\gamma^*)$, then $\mathbb{P}(A_j^* = 0) = 0$ due to the fact that $A_j^* = A_{\gamma^*,j}^S$ and the definition of $A_{\gamma^*,j}^S$ (see (4.8)). In this case, $\prod_{j\neq i} \mathbb{P}\left(A_j^* = 0\right) = 0$. On the other hand, if $j \notin \Omega_A(\gamma^*)$ for any $j \neq i$, then $A_j^* = A_{\gamma^*,j}^W$ for $j \neq i$; therefore,

$$
\begin{aligned}
\prod_{j\neq i} \mathbb{P}\left(A_j^* = 0\right) = \prod_{j\neq i} \left[\left(\frac{v_j^B}{\lambda_B^*} - \frac{v_j^A}{\lambda_A^*}\right) \Big/ \frac{v_j^B}{\lambda_B^*}\right] &= \prod_{j\neq i} \left(1 - \frac{v_j^A}{v_j^B} \frac{\lambda_B^*}{\lambda_A^*}\right)\\
&\leq \left(1 - \frac{\underline{\lambda}}{\bar{\lambda}} \frac{\frac{\underline{w}}{n\bar{w}}}{\frac{\bar{w}}{n\underline{w}}}\right)^{n-1} = \left(1 - \frac{\underline{\lambda}}{\bar{\lambda}} \frac{\underline{w}^2}{\bar{w}^2}\right)^{n-1}.
\end{aligned}
$$

Here, to obtain the last equality, we use (4.1) for the bounds of v_j^A, v_j^B and Proposition 4.1.6 for the bounds of λ_A^*, λ_B^*.

Similarly, we can obtain $\mathbb{P}(B_i^n = X^B) \leq \left(1 - \frac{\underline{\lambda}}{\overline{\lambda}} \frac{\underline{w}^2}{\overline{w}^2}\right)^{n-1}$.

A.4 Proof of Lemma A.4

Fix $\varepsilon \in (0,1]$ and assume that (16) is satisfied, we prove that (4.13) also holds by comparing the terms in (16) with the terms in (4.13). First, due to the fact that $\alpha \leq 1$, we can find a lower bound of the left-hand side of (16) as follows:

$$\begin{aligned}\sum_{i=1}^n v_i^A F_{B_i^n}(x_i^A) &= \sum_{i=1}^n v_i^A \mathbb{P}(B_i^n = x_i^A) + \sum_{i=1}^n v_i^A \mathbb{P}(B_i^n < x_i^A) \\ &\geq \alpha \sum_{i=1}^n v_i^A \mathbb{P}(B_i^n = x_i^A) + \sum_{i=1}^n v_i^A \mathbb{P}(B_i^n < x_i^A) \\ &= \Pi^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*})/W^A. \end{aligned} \tag{24}$$

Now, we turn our focus to the right-hand-side of (16), we can rewrite the involved term as follows.

$$\sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) = \sum_{i=1}^n \int_0^\infty v_i^A \mathbb{P}(B_i^n = x) \mathrm{d}F_{A_i^n}(x) + \sum_{i=1}^n \int_0^\infty v_i^A \mathbb{P}(B_i^n < x) \mathrm{d}F_{A_i^n}(x).$$

We observe that $\sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x)$ is very similar to the expression of $\Pi^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*})$ stated in (15). The main difference lies at the coefficient of the term related to the tie cases that is the tie-breaking parameter α. Therefore, we consider the following two cases of α:

Case 1: $\alpha = 1$. For any n, divide two sides of (15) (with $\alpha = 1$) by W^A and recall that $v_i^A := w_i^A / W^A, \forall i$, we trivially have $\sum_{i=1}^n \int_0^\infty v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) = \Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A$.

Case 2: $\alpha < 1$. From Results (ii) and (iii) of Lemma A.3, for any $x > 0$, we have $\mathbb{P}(B_i^n = x) \leq D^{n-1}$ where we define $D := \left(1 - \frac{\underline{\lambda}}{\overline{\lambda}} \frac{\underline{w}^2}{\overline{w}^2}\right) < 1$. We consider two cases of α as follows.

- If $2(1-\alpha) \leq 1$, define $\hat{C}_1 := \frac{1}{\ln(1/D)} + 1 > 0$, we have that[16] $\hat{C}_1 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq \log_D \varepsilon + 1$; therefore, for any $n \geq \hat{C}_1 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, we obtain $n - 1 \geq \log_D \varepsilon$ and

$$D^{n-1} \leq D^{\log_D \varepsilon} = \varepsilon \leq \frac{\varepsilon}{2(1-\alpha)} \qquad \text{(note that } D < 1 \text{ and in this case } 2(1-\alpha) \leq 1).$$

[16] If $\varepsilon < 1/e$, then $\ln(1/\varepsilon) > 1$ and $\hat{C}_1 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) = \frac{\ln(1/\varepsilon)}{\ln(1/D)} + \ln(1/\varepsilon) > \log_D \varepsilon + 1$; otherwise, if $\varepsilon \geq 1/e$, we have $\ln(1/\varepsilon) \leq 1$ and $\hat{C}_1 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) = \frac{1}{\ln(1/D)} + 1 \geq \frac{\ln(1/\varepsilon)}{\ln(1/D)} + 1 = \log_D \varepsilon + 1$ (note that $\ln(1/D) > 0$ since $D < 1$).

- If $2(1-\alpha)>1$, define $\hat{C}_2 := \frac{1}{\ln(1/D)} + \frac{\ln(2-2\alpha)}{\ln(1/D)} + 1 > 0$ and deduce that[17]

$$\hat{C}_2 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq \log_D \frac{\varepsilon}{2(1-\alpha)} + 1.$$

We conclude that for any $n \geq \hat{C}_2 \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we obtain $n-1 \geq \log_D\left(\frac{\varepsilon}{2(1-\alpha)}\right)$ and

$$D^{n-1} \leq D^{\log_D \frac{\varepsilon}{2(1-\alpha)}} = \frac{\varepsilon}{2(1-\alpha)}.$$

Let us define $C_0^* = \max\{\hat{C}_1, \hat{C}_2\} > 0$, we conclude that for any $\alpha < 1, n \geq C_0^* \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, $i \in [n]$ and $x > 0$, we have

$$\mathbb{P}(B_i^n = x) \leq D^{n-1} \leq \frac{\varepsilon}{2(1-\alpha)}. \tag{25}$$

Note also that $\mathbb{P}(A_i^n = B_i^n = 0) = \mathbb{P}(A_i^n = 0)\mathbb{P}(B_i^n = 0) = \mathbb{P}(A_i^* = 0)\mathbb{P}(B_i^* = 0) = 0, \forall i,$[18] we conclude that when $\alpha < 1$, for any $n \geq C_0^* \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^\infty v_i^A F_{B_i^n}(x)\mathrm{d}F_{A_i^n}(x)\\
=&\left[\sum_{i=1}^{n}\int_0^\infty v_i^A \mathbb{P}(B_i^n < x)\mathrm{d}F_{A_i^n}(x) + \alpha\sum_{i=1}^{n}\int_0^\infty v_i^A \mathbb{P}(B_i^n = x)\mathrm{d}F_{A_i^n}(x)\right]\\
&\qquad + (1-\alpha)\sum_{i=1}^{n}\int_0^\infty v_i^A \mathbb{P}(B_i^n = x)\mathrm{d}F_{A_i^n}(x)\\
=&\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A + (1-\alpha)\sum_{i=1}^{n} v_i^A \int_{(0,\infty)} \frac{\varepsilon}{2(1-\alpha)}\mathrm{d}F_{A_i^n}(x)\\
&\qquad + (1-\alpha)\sum_{i=1}^{n} v_i^A \mathbb{P}(A_i^n = B_i^n = 0)\\
=&\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A + (1-\alpha)\sum_{i=1}^{n} v_i^A \int_{(0,\infty)} \frac{\varepsilon}{2(1-\alpha)}\mathrm{d}F_{A_i^n}(x) + 0\\
\leq&\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A + (1-\alpha)\frac{\varepsilon}{2(1-\alpha)}\\
=&\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A + \varepsilon/2.
\end{aligned}$$

---

[17] If $\varepsilon < 1/e$, we have $\hat{C}_2 \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) = \log_D \varepsilon + \left(\log_{1/D}(2-2\alpha)+1\right)\ln\left(\frac{1}{\varepsilon}\right) > \log_D \varepsilon - \log_D(2-2\alpha) + 1$; otherwise, if $\varepsilon \geq 1/e$, we have $\hat{C}_2 \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) = \hat{C}_2 \geq \frac{\ln(1/\varepsilon)}{\ln(1/D)} + \frac{\ln(2-2\alpha)}{\ln(1/D)} + 1 \geq \log_D \varepsilon - \log_D(2-2\alpha) + 1$ (since $1 \geq \ln\left(\frac{1}{\varepsilon}\right)$).

[18] Note that if $i \in \Omega_A(\gamma^*)$ then $\mathbb{P}(A_i^* = 0) = 0$, if $i \notin \Omega_A(\gamma^*)$ then $\mathbb{P}(B_i^* = 0) = 0$ (see (4.8)-(4.12)); therefore, $\mathbb{P}(A_i^* = 0)\mathbb{P}(A_i^* = 0) = 0, \forall i$.

In conclusion, regardless of the value of α, for any $n \geq C_0^* \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, we have

$$\sum_{i=1}^{n} \int_0^{\infty} v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) \leq \Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})/W^A + \varepsilon/2. \tag{26}$$

Combine (24), (26) and the assumption that (16) holds, for any $n \geq C_0^* \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, we have

$$\frac{\Pi^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*})}{W^A} \overset{(24)}{\leq} \sum_{i=1}^{n} v_i^A F_{B_i^n}\left(x_i^A\right) \overset{(16)}{\leq} \sum_{i=1}^{n} \int_0^{\infty} v_i^A F_{B_i^n}(x) \mathrm{d}F_{A_i^n}(x) + \varepsilon/2 \overset{(26)}{\leq} \frac{\Pi^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*})}{W^A} + \varepsilon.$$

By multiplying both sides of this inequality by W^A, we obtain (4.13).

A.5 Proof of Lemma A.5

We compute the right-hand-side of (17) based on the definition of $F_{A_i^*}$ and $F_{B_i^*}$ (see Definition 4.1.4).

$$\begin{aligned}
&\sum_{i=1}^{n} \int_0^{\infty} v_i^A F_{B_i^*}(x) \mathrm{d}F_{A_i^*}(x) \\
&= \sum_{i \in \Omega_A(\gamma^*)} \int_0^{\infty} v_i^A F_{B^{W}_{\gamma^*,i}}(x) \mathrm{d}F_{A^{S}_{\gamma^*,i}}(x) + \sum_{i \notin \Omega_A(\gamma^*)} \int_0^{\infty} v_i^A F_{B^{S}_{\gamma^*,i}}(x) \mathrm{d}F_{A^{W}_{\gamma^*,i}}(x) \\
&= \sum_{i \in \Omega_A(\gamma^*)} \int_0^{\frac{v_i^B}{\lambda_B^*}} v_i^A \left(\frac{\frac{v_i^A}{\lambda_A^*} - \frac{v_i^B}{\lambda_B^*}}{\frac{v_i^A}{\lambda_A^*}} + \frac{x}{\frac{v_i^A}{\lambda_A^*}} \right) \frac{1}{\frac{v_i^B}{\lambda_B^*}} \mathrm{d}x + \sum_{i \notin \Omega_A(\gamma^*)} \int_0^{\frac{v_i^A}{\lambda_A^*}} v_i^A \frac{x}{\frac{v_i^A}{\lambda_A^*}} \frac{1}{\frac{v_i^B}{\lambda_B^*}} \mathrm{d}x \\
&= \sum_{i \in \Omega_A(\gamma^*)} v_i^A \left(1 - \frac{v_i^B \gamma^*}{2 v_i^A}\right) + \sum_{i \notin \Omega_A(\gamma^*)} (v_i^A)^2 \frac{1}{2\gamma^* v_i^B}.
\end{aligned} \tag{27}$$

On the other hand, for any pure strategy $\boldsymbol{x}^A$ of player A, we have:

$$\begin{aligned}
&\sum_{i=1}^{n} v_i^A F_{B_i^*}\left(x_i^A\right) \\
&= \sum_{i \in \Omega_A(\gamma^*)} v_i^A F_{B^{W}_{\gamma^*,i}}\left(x_i^A\right) + \sum_{i \notin \Omega_A(\gamma^*)} v_i^A F_{B^{S}_{\gamma^*,i}}\left(x_i^A\right) \\
&\leq \sum_{i \in \Omega_A(\gamma^*)} v_i^A \left(\frac{\frac{v_i^A}{\lambda_A^*} - \frac{v_i^B}{\lambda_B^*}}{\frac{v_i^A}{\lambda_A^*}} + \frac{x_i^A \lambda_A^*}{v_i^A} \right) + \sum_{i \notin \Omega_A(\gamma^*)} v_i^A \left(\frac{x_i^A \lambda_A^*}{v_i^A} \right) \\
&\leq \sum_{i \in \Omega_A(\gamma^*)} \left(\frac{v_i^A}{\lambda_A^*} - \frac{v_i^B}{\lambda_B^*} \right) \lambda_A^* + \lambda_A^* X^A \qquad \text{(since } \textstyle\sum_{i=1}^n x_i^A \leq X^A\text{)} \\
&= \sum_{i \in \Omega_A(\gamma^*)} v_i^A \left(1 - \frac{v_i^B \gamma^*}{2 v_i^A}\right) + \sum_{i \notin \Omega_A(\gamma^*)} (v_i^A)^2 \frac{1}{2\gamma^* v_i^B}.
\end{aligned} \tag{28}$$

Here, to obtain the last equality, we use (10) to rewrite X^A. Finally, from (27) and (28), we conclude that (17) holds for any $\boldsymbol{x}^A$ and γ^*.

A.6 Proof of Lemma A.6

Since the definition of $F_{A_i^n}$ involves $\mathbb{P}(A_{=0})$ (see (12)), we first look for an upper-bound of $\mathbb{P}(A_{=0})$. For any n and $\gamma^* \in \mathcal{S}_n^{(4.5)}$, if $\Omega_A(\gamma^*) \neq \emptyset$, i.e., there exists i such that $A_i^* = A_{\gamma^*,i}^S$, then $\mathbb{P}(A_i^* = 0) = 0$ due to the definition of $A_{\gamma^*,i}^S$ (see (4.8)); in this case, $\mathbb{P}(A_{=0}) = \prod_{j\in[n]} \mathbb{P}\left(A_j^* = 0\right) = 0$. On the other hand, if $\Omega_A(\gamma^*) = \emptyset$, then $A_j^* = A_{\gamma^*,j}^W$ for any $j \in [n]$; therefore,

$$\mathbb{P}(A_{=0}) = \prod_{j\in[n]} \mathbb{P}\left(A_j^* = 0\right) = \prod_{j\in[n]} \left[\left(\frac{v_j^B}{\lambda_B^*} - \frac{v_j^A}{\lambda_A^*}\right)\Big/\frac{v_j^B}{\lambda_B^*}\right] = \prod_{j\in[n]} \left(1 - \frac{v_j^A}{v_j^B}\frac{\lambda_B^*}{\lambda_A^*}\right) \le \left(1 - \frac{\underline{\lambda}}{\overline{\lambda}}\frac{\underline{w}^2}{\overline{w}^2}\right)^n. \tag{29}$$

Here, the last inequality comes directly from (4.1) and Proposition 4.1.6. Recall the notation $D := \left(1 - \frac{\underline{\lambda}}{\overline{\lambda}}\frac{\underline{w}^2}{\overline{w}^2}\right)$ and define $\tilde{C}_0 := \frac{\ln(4)+1}{\ln(1/D)} > 0$, we have $\tilde{C}_0 \ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right) \ge \log_D\left(\frac{\varepsilon_1}{4}\right)$.[19] Therefore, for any $n \ge \tilde{C}_0 \ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$ we have $n \ge \log_D(\varepsilon_1/4)$ and since $D < 1$ we have:

$$\mathbb{P}(A_{=0}) \le D^n \le D^{\log_D(\varepsilon_1/4)} = \varepsilon_1/4. \tag{30}$$

Now, we look for an upper-bound of $\mathbb{P}(A_{>0})$. For any n, define the constants $\epsilon_n := \frac{\varepsilon_1}{4}\frac{\underline{w}}{n\overline{w}\overline{\lambda}}$ and $\tau := \frac{1}{X^A}\left(\frac{\overline{w}}{n\underline{w}\underline{\lambda}}\frac{1}{\epsilon_n} + 1\right) = \frac{1}{X^A}\left[\frac{4\overline{\lambda}}{\varepsilon_1\underline{\lambda}}\left(\frac{\overline{w}}{\underline{w}}\right)^2 + 1\right]$, we consider the following term for any $i \in [n]$:

$$\begin{aligned}
&\mathbb{P}\left(\left\{A_i^* - \frac{A_i^*}{\sum_{j\in[n]} A_j^*}X^A > \epsilon_n\right\} \bigcap A_{>0}\right) \\
&\le \mathbb{P}\left(\left\{\left|A_i^* - \frac{A_i^*}{\sum_{j\in[n]} A_j^*}X^A\right| > \epsilon_n\right\} \bigcap A_{>0}\right) \\
&\le \mathbb{P}\left(A_i^*\left|\sum_{j\in[n]} A_j^* - X^A\right| > \epsilon_n \sum_{j\in[n]} A_j^*\right) \\
&= \mathbb{P}\left(A_i^*\left|\sum_{j\in[n]} A_j^* - X^A\right| > \epsilon_n X^A - \epsilon_n\left(X^A - \sum_{j\in[n]} A_j^*\right)\right) \\
&\le \mathbb{P}\left(A_i^*\left|\sum_{j\in[n]} A_j^* - X^A\right| > \epsilon_n X^A - \epsilon_n\left|\sum_{j\in[n]} A_j^* - X^A\right|\right) \\
&= \mathbb{P}\left(\left|\sum_{j\in[n]} A_j^* - X^A\right| > \frac{\epsilon_n X^A}{A_i^* + \epsilon_n}\right) \\
&\le \mathbb{P}\left(\left|\sum_{j\in[n]} A_j^* - X^A\right| > \frac{\epsilon_n X^A}{\frac{\overline{w}}{n\underline{w}\underline{\lambda}} + \epsilon_n}\right)
\end{aligned}$$

[19] This is due to the fact that $\tilde{C}_0 \cdot \ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right) = \frac{1}{\ln(1/D)}(\ln(4)+1)\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right) \ge \frac{\ln(4/\varepsilon_1)}{\ln(1/D)} = \log_D\left(\frac{\varepsilon_1}{4}\right)$; here, we have applied Lemma A.2 (see Appendix A.1) for $\hat{\varepsilon} := \varepsilon_1$ and $\hat{C} := 4$.

$$= \mathbb{P}\left(\left|\sum\nolimits_{j\in[n]} A_j^* - X^A\right| > \frac{1}{\tau}\right). \quad (31)$$

Here, the second-to-last inequality comes from the fact that for any $i \in [n]$, A_i^* is upper-bounded by either v_i^A/λ_A^* or v_i^B/λ_B^* (see (4.8) and (4.10)), thus, it is bounded by $\bar{w}/(n\underline{w}\underline{\lambda})$ (due to (4.1) and Proposition 4.1.6).

Recall that $X^A = \mathbb{E}\left[\sum_{j=1}^{n} A_j^*\right]$ (see Lemma A.1-(iii)), we use the Hoeffding's inequality (see e.g., Theorem 2, Hoeffding (1963)) on the random variables $A_i^*, i \in [n]$ (bounded in $[0, \bar{w}/(n\underline{w}\underline{\lambda})]$) to obtain

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{j\in[n]} A_j^* - X^A\right| > \frac{1}{\tau}\right) \leq & 2\exp\left(\frac{-2\frac{1}{\tau^2}}{\sum\limits_{j\in[n]}\left(\frac{\bar{w}}{n\underline{w}\underline{\lambda}}\right)^2}\right)\\ =&2\exp\left[\frac{-2n}{\tau^2}\left(\frac{\underline{\lambda}\underline{w}}{\bar{w}}\right)^2\right]. \end{aligned} \quad (32)$$

Now, we define $\tilde{C}_1 := \frac{1}{2}\left(\frac{4}{X^A}\frac{\bar{\lambda}}{\underline{\lambda}}\frac{\bar{w}^2}{\underline{w}^2} + \frac{1}{X^A}\right)^2 (\ln 8 + 1)\left(\frac{\bar{w}}{\underline{w}\underline{\lambda}}\right)^2$; due to the definition of τ, we have that[20] $\tilde{C}_1 \cdot \frac{1}{\varepsilon_1^2}\ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right) \geq \frac{\tau^2}{2}\ln\left(\frac{8}{\varepsilon_1}\right)\left(\frac{\bar{w}}{\underline{w}\underline{\lambda}}\right)^2$; therefore, $\forall n \geq \tilde{C}_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$, we can deduce that $\frac{2n}{\tau^2}\left(\frac{\underline{w}\underline{\lambda}}{\bar{w}}\right)^2 \geq \ln\left(\frac{8}{\varepsilon_1}\right)$ and thus,

$$2\exp\left[\frac{-2n}{\tau^2}\left(\frac{\underline{\lambda}\underline{w}}{\bar{w}}\right)^2\right] \leq 2\exp\left[-\ln\left(\frac{8}{\varepsilon_1}\right)\right] = \frac{\varepsilon_1}{4}. \quad (33)$$

Combining (31), (32) and (33), we deduce that

$$\mathbb{P}\left(\left\{A_i^* - \frac{A_i^*}{\sum_{j\in[n]} A_j^*}X^A > \epsilon_n\right\}\bigcap A_{>0}\right) \leq \frac{\varepsilon_1}{4}, \forall n \geq \tilde{C}_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right). \quad (34)$$

Finally, note that for any n, $i \in [n]$ and $x \geq 0$, we also have

$$\begin{aligned}&\mathbb{P}\left(\left\{\frac{A_i^* \cdot X^A}{\sum_{j\in[n]} A_j^*} \leq x\right\}\bigcap A_{>0}\right)\\ =&\mathbb{P}\left(\left\{\frac{A_i^* X^A}{\sum\limits_{j\in[n]} A_j^*} \leq x\right\}\bigcap\left\{A_i^* - \frac{A_i^* X^A}{\sum\limits_{j\in[n]} A_j^*} \leq \epsilon_n\right\}\bigcap A_{>0}\right)\end{aligned}$$

[20] This is due to $\tilde{C}_1 \cdot \frac{1}{\varepsilon_1^2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right) = \frac{1}{2}\left[\frac{1}{X^A}\left(\frac{4}{\varepsilon_1}\frac{\bar{\lambda}}{\underline{\lambda}}\frac{\bar{w}^2}{\underline{w}^2} + \frac{1}{\varepsilon_1}\right)\right]^2 \cdot (\ln(8)+1)\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right) \cdot \left(\frac{\bar{w}}{\underline{w}\underline{\lambda}}\right)^2 \geq \frac{\tau^2}{2} \cdot \ln\left(\frac{8}{\varepsilon_1}\right) \cdot \left(\frac{\bar{w}}{\underline{w}\underline{\lambda}}\right)^2$; here, we have used Lemma A.2 with $\hat{\varepsilon} := \varepsilon_1$ and $\hat{C} := 8$ and the fact that $1/\varepsilon \geq 1$.

$$+\mathbb{P}\left(\left\{\frac{A_i^*\cdot X^A}{\sum\limits_{j\in[n]} A_j^*}\leq x\right\}\bigcap\left\{A_i^*-\frac{A_i^*X^A}{\sum\limits_{j\in[n]} A_j^*}>\epsilon_n\right\}\bigcap A_{>0}\right)$$

$$\leq\mathbb{P}\left(\{A_i^*\leq x+\epsilon_n\}\right)+\mathbb{P}\left(\left\{A_i^*-\frac{A_i^*X^A}{\sum\limits_{j\in[n]} A_j^*}>\epsilon_n\right\}\bigcap A_{>0}\right). \tag{35}$$

Therefore, define $C_1 := \max\{\tilde{C_0}, \tilde{C_1}\}$, for any $n \geq C_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$, from (12), we have

$$\begin{aligned}
&F_{A_i^n}(x)-F_{A_i^*}(x)\\
=&\mathbb{P}(A_{=0})+\mathbb{P}\left(\left\{\frac{A_i^*\cdot X^A}{\sum_{j\in[n]}A_j^*}\leq x\right\}\bigcap A_{>0}\right)-F_{A_i^*}(x)\\
\leq&\frac{\varepsilon_1}{4}+\mathbb{P}\left(\{A_i^*\leq x+\epsilon_n\}\right)+\mathbb{P}\left(\left\{A_i^*-\frac{A_i^*X^A}{\sum\limits_{j\in[n]} A_j^*}>\epsilon_n\right\}\bigcap A_{>0}\right)-F_{A_i^*}(x) \quad \text{(from (30) and (35))}\\
\leq&\frac{\varepsilon_1}{4}+F_{A_i^*}(x+\epsilon_n)+\frac{\varepsilon_1}{4}-F_{A_i^*}(x) \qquad \text{(due to (34)).} \qquad (36)
\end{aligned}$$

The final step is to bound the term $F_{A_i^*}(x+\epsilon_n)-F_{A_i^*}(x)$; we present this as the following lemma.

Lemma A.8. *For any $\epsilon > 0, n > 0, i \in [n]$ and $x \in [0,\infty)$, we have $F_{A_i^*}(x+\epsilon)-F_{A_i^*}(x)\leq\frac{\epsilon\lambda_B^*}{v_i^B}$.*

Proof. If $i \in \Omega_A(\gamma^*)$, then $A_i^* = A_{\gamma^*,i}^S$ and

$$F_{A^S_{\gamma^*,i}}(x+\epsilon)-F_{A^S_{\gamma^*,i}}(x)=\begin{cases}\frac{(x+\epsilon)\lambda_B^*}{v_i^B}-\frac{x\lambda_B^*}{v_i^B}=\frac{\epsilon\lambda_B^*}{v_i^B}, & \text{if } 0\leq x<\frac{v_i^B}{\lambda_B^*}-\epsilon\\ 1-\frac{x\lambda_B^*}{v_i^B}\leq\frac{\epsilon\lambda_B^*}{v_i^B}, & \text{if } \frac{v_i^B}{\lambda_B^*}-\epsilon\leq x\leq\frac{v_i^B}{\lambda_B^*}\\ 1-1\leq\frac{\epsilon v_i^B}{\lambda_B^*}, & \text{if } x>\frac{v_i^B}{\lambda_B^*}\end{cases}. \tag{37}$$

On the other hand, if $i \notin \Omega_A(\gamma^*)$, then $A_i^* = A_{\gamma^*,i}^W$ and we have

$$F_{A^W_{\gamma^*,i}}(x+\epsilon)-F_{A^W_{\gamma^*,i}}(x)=\begin{cases}\frac{(x+\epsilon)\lambda_B^*}{v_i^B}-\frac{x\lambda_B^*}{v_i^B}=\frac{\epsilon\lambda_B^*}{v_i^B}, & \text{if } 0\leq x<\frac{v_i^A}{\lambda_A^*}-\epsilon\\ 1-\frac{\frac{v_i^B}{\lambda_B^*}-\frac{v_i^A}{\lambda_A^*}}{\frac{v_i^B}{\lambda_B^*}}-\frac{x\lambda_B^*}{v_i^B}\leq\frac{\epsilon\lambda_B^*}{v_i^B}, & \text{if } \frac{v_i^A}{\lambda_A^*}-\epsilon\leq x\leq\frac{v_i^A}{\lambda_A^*}\\ 1-1\leq\frac{\epsilon v_i^B}{\lambda_B^*}, & \text{if } x>\frac{v_i^A}{\lambda_A^*}\end{cases}. \tag{38}$$

□

Combine this lemma with (36) and recall the definition of ϵ_n (which induces that $\epsilon_n\lambda_B^*/v_i^B \leq \varepsilon_1/2$), we conclude that $F_{A_i^n}(x)-F_{A_i^*}(x)\leq\varepsilon_1$ for any $n\geq C_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$.

Similarly, for $n \geq C_1 \varepsilon_1^{-2} \ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right)$ and $i \in [n]$, we deduce that for any $x \in [0, \infty)$, $F_{A_i^n}(x) - F_{A_i^*}(x) \geq -\varepsilon_1$. Therefore,

$$n \geq C_1 \varepsilon_1^{-2} \ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right) \implies \sup_{x \in [0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right| \leq \varepsilon_1.$$

The inequality $\sup_{x \in [0,\infty)} \left| F_{B_i^n}(x) - F_{B_i^*}(x) \right| \leq \varepsilon_1$ can be proved in a similar way.

A.7 Proof of Lemma A.7

In this proof, we will use the notations

$$\mathbb{E} f(X) := \int_0^\infty f(z) \mathrm{d}F_Z(x) \quad \text{and} \quad \mathbb{E}_{\mathcal{I}} f(X) := \int_{\mathcal{I}} f(z) \mathrm{d}F_Z(x),$$

for any function f, random variable Z and interval $\mathcal{I}$. To simplify the notation, let us define $M := \frac{\bar{\lambda}}{\underline{\lambda}} \frac{\bar{w}^2}{\underline{w}^2}$ and we denote by $\mathcal{I}_i$ the interval $\left[0, v_i^B / \lambda_B^*\right]$. For any $\varepsilon_2 \in (0, 1]$, we define $\delta_2 := \frac{\varepsilon_2}{6+2M}$. We first consider the case where $i \in \Omega_A(\gamma^*)$, i.e., $B_i^* = B_{\gamma^*,i}^W$. Note that $F_{A_i^n}(x) = F_{A_i^*}(x) = 1$ for any $x \geq 2X^B$ (see Lemma A.1-(iv)); the left-hand-side of (22) can be rewritten as follows.

$$\begin{aligned}
&\left| \mathbb{E} F_{B_i^*}(A_i^n) - \mathbb{E} F_{B_i^*}(A_i^*) \right| \\
=& \left| \int_{[0,2X^B]} F_{B_i^*}(x) \mathrm{d}F_{A_i^n}(x) - \int_{[0,2X^B]} F_{B_i^*}(x) \mathrm{d}F_{A_i^*}(x) \right| \\
\leq& \left| \mathbb{E}_{[0, v_i^B/\lambda_B^*]} F_{B_{\gamma^*,i}^W}(A_i^n) - \mathbb{E}_{[0, v_i^B/\lambda_B^*]} F_{B_{\gamma^*,i}^W}(A_i^*) \right| + \left| \int_{(v_i^B/\lambda_B^*, 2X^B]} \mathrm{d}F_{A_i^n}(x) - \int_{(v_i^B/\lambda_B^*, 2X^B]} \mathrm{d}F_{A_i^*}(x) \right| \\
=& \left| \mathbb{E}_{\mathcal{I}_i} F_{B_{\gamma^*,i}^W}(A_i^n) - \mathbb{E}_{\mathcal{I}_i} F_{B_{\gamma^*,i}^W}(A_i^*) \right| + \left| F_{A_i^n}(2X^B) - F_{A_i^n}(v_i^B/\lambda_B^*) - F_{A_i^*}(2X^B) + F_{A_i^*}(v_i^B/\lambda_B^*) \right| \\
\leq& \left| \mathbb{E}_{\mathcal{I}_i} F_{B_{\gamma^*,i}^W}(A_i^n) - \mathbb{E}_{\mathcal{I}_i} F_{B_{\gamma^*,i}^W}(A_i^*) \right| + 2 \sup_{x \in [0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right|. \qquad (39)
\end{aligned}$$

We now focus on bounding the first term in (39). Let us define $K := \left\lceil \frac{M}{\delta_2} \right\rceil$ and $K+1$ points x_j such that $x_0 := 0$ and $x_j := x_{j-1} + \frac{v_i^B}{\lambda_B^* K}, \forall j \in [K]$. In other words, we have the partitions $\mathcal{I}_i = \bigcup_{j=1}^K P_j$ where we denote by P_1 the interval $[x_0, x_1]$ and by P_j the interval $(x_{j-1}, x_j]$ for $j = 2, \ldots, K$. For any $x, x' \in P_j, \forall j \in [K]$, from the definition of $B_{\gamma^*,i}^W$, we have

$$|F_{B_{\gamma^*,i}^W}(x) - F_{B_{\gamma^*,i}^W}(x')| = \frac{\lambda_A^*}{v_i^A} |x - x'| \leq \frac{\lambda_A^*}{v_i^A} \cdot \frac{v_i^B}{\lambda_B^*} \cdot \frac{1}{K} \leq \frac{\bar{\lambda} n \bar{w}}{\underline{w}} \cdot \frac{\bar{w}}{n \underline{w} \underline{\lambda}} \cdot \frac{1}{K} = \frac{M}{K} \leq \delta_2. \quad (40)$$

Now, we define the function $g(x) := \sum_{j=1}^K F_{B_{\gamma^*,i}^W}(x_j) \mathbf{1}_{P_j}(x)$. Here, $\mathbf{1}_{P_j}$ is the indicator function of the set P_j. From this definition and Inequality (40), we trivially have

$|F_{B^W_{\gamma^*,i}}(x) - g(x)| \le \delta_2, \forall x \in \mathcal{I}_i$. Therefore,

$$\left|\mathbb{E}_{\mathcal{I}_i} F_{B^W_{\gamma^*,i}}(A_i^n) - \mathbb{E}_{\mathcal{I}_i} g(A_i^n)\right| \le \int_{I_i} \left|F_{B^W_{\gamma^*,i}}(x) - g(x)\right| \mathrm{d}F_{A_i^n}(x) \le \int_{\mathcal{I}_i} \delta_2 \mathrm{d}F_{A_i^n}(x) \le \delta_2, \quad (41)$$

$$\left|\mathbb{E}_{\mathcal{I}_i} F_{B^W_{\gamma^*,i}}(A_i^*) - \mathbb{E}_{\mathcal{I}_i} g(A_i^*)\right| \le \int_{I_i} \left|F_{B^W_{\gamma^*,i}}(x) - g(x)\right| \mathrm{d}F_{A_i^*}(x) \le \int_{\mathcal{I}_i} \delta_2 \mathrm{d}F_{A_i^*}(x) \le \delta_2. \quad (42)$$

Now, we note that for any $j \in [K]$, $F_{B^W_{\gamma^*,i}}(x_j) = \sum_{m=0}^{j} \left[F_{B^W_{\gamma^*,i}}(x_m) - F_{B^W_{\gamma^*,i}}(x_{m-1})\right]$; here, for the sake of notation, we denote by x_{-1} an arbitrary negative number (that is $F_{B^W_{\gamma^*,i}}(x_{-1}) = 0$). Using this, we have:

$$\begin{aligned}
&\left|\mathbb{E}_{\mathcal{I}_i} g\left(A_i^n\right) - \mathbb{E}_{\mathcal{I}_i} g\left(A_i^*\right)\right| \\
&= \left|\sum_{j=1}^{K} F_{B^W_{\gamma^*,i}}(x_j) \left[\mathbb{E}_{\mathcal{I}_i} \mathbf{1}_{P_j}\left(A_i^n\right) - \mathbb{E}_{\mathcal{I}_i} \mathbf{1}_{P_j}\left(A_i^*\right)\right]\right| \\
&= \left|\sum_{j=1}^{K} F_{B^W_{\gamma^*,i}}(x_j) \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&= \left|\sum_{j=1}^{K} \left(\sum_{m=0}^{j} \left[F_{B^W_{\gamma^*,i}}(x_m) - F_{B^W_{\gamma^*,i}}(x_{m-1})\right] \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right)\right| \\
&\le \left|\left[F_{B^W_{\gamma^*,i}}(x_0) - F_{B^W_{\gamma^*,i}}(x_{-1})\right] \sum_{j=1}^{K} \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&\qquad + \left|\sum_{m=1}^{K} \left(\left[F_{B^W_{\gamma^*,i}}(x_m) - F_{B^W_{\gamma^*,i}}(x_{m-1})\right] \sum_{j=m}^{K} \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right)\right|. \quad (43)
\end{aligned}$$

Note that $\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right) = F_{A_i^n}(x_j) - F_{A_i^n}(x_{j-1}) - F_{A_i^*}(x_j) + F_{A_i^*}(x_{j-1})$.[21] Moreover, due to the fact that $x_0 = 0$ and $F_{B^W_{\gamma^*,i}}(x_{-1}) = 0$, we can rewrite the first term in (43) as follows:

$$\begin{aligned}
&\left|\left[F_{B^W_{\gamma^*,i}}(x_0) - F_{B^W_{\gamma^*,i}}(x_{-1})\right] \sum_{j=1}^{K} \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&= \left|F_{B^W_{\gamma^*,i}}(0) \cdot \left[\sum_{j=1}^{K} \left(F_{A_i^n}(x_j) - F_{A_i^n}(x_{j-1}) - F_{A_i^*}(x_j) + F_{A_i^*}(x_{j-1})\right)\right]\right| \\
&= \left|F_{B^W_{\gamma^*,i}}(0) \cdot \left[F_{A_i^n}(x_K) - F_{A_i^n}(x_0) - F_{A_i^*}(x_K) + F_{A_i^*}(x_0)\right]\right|
\end{aligned}$$

[21] For any $j \ge 2$, this is trivially since $P_j := (x_{j-1}, x_j]$. For $P_1 = [0, x_1]$, we have that $\mathbb{P}\left(A_i^n \in P_1\right) - \mathbb{P}\left(A_i^* \in P_1\right) = \mathbb{P}(A_i^n \in (0, x_1]) - \mathbb{P}(A_i^* \in (0, x_1]) + \mathbb{P}(A_i^n = 0) - \mathbb{P}(A_i^* = 0)$; moreover, due to Lemma A.3-(i), we also note that $\mathbb{P}(A_i^n = 0) = \mathbb{P}(A_i^* = 0)$.

$$
\begin{aligned}
&\leq F_{B^{W}_{\gamma^*,i}}(0)\cdot 2 \sup_{x\in[0,\infty)} \left|F_{A_i^n}(x)-F_{A_i^*}(x)\right| \\
&\leq 2 \sup_{x\in[0,\infty)} \left|F_{A_i^n}(x)-F_{A_i^*}(x)\right|. \qquad (44)
\end{aligned}
$$

Now, recall that $x_m = x_{m-1} + v_i^B/(\lambda_B^* \cdot K), \forall m \in [K]$, by the definition of $F_{B^{W}_{\gamma^*,i}}$, we deduce that $F_{B^{W}_{\gamma^*,i}}(x_m) - F_{B^{W}_{\gamma^*,i}}(x_{m-1}) = \frac{v_i^B}{\lambda_B^* K}\frac{\lambda_A^*}{v_i^A} \leq \frac{\bar{\lambda}}{\underline{\lambda}}\frac{\bar{w}^2}{\underline{w}^2}\frac{1}{K} = \frac{M}{K}, \forall m \in [K]$. Therefore, the second term in (43) is

$$
\begin{aligned}
&\left|\sum_{m=1}^{K}\left(\left[F_{B^{W}_{\gamma^*,i}}(x_m)-F_{B^{W}_{\gamma^*,i}}(x_{m-1})\right]\sum_{j=m}^{K}\left[\mathbb{P}\left(A_i^n\in P_j\right)-\mathbb{P}\left(A_i^*\in P_j\right)\right]\right)\right| \\
=&\left|\sum_{m=1}^{K}\left(\left[F_{B^{W}_{\gamma^*,i}}(x_m)-F_{B^{W}_{\gamma^*,i}}(x_{m-1})\right]\sum_{j=m}^{K}\left[F_{A_i^n}(x_j)-F_{A_i^n}(x_{j-1})-F_{A_i^*}(x_j)+F_{A_i^*}(x_{j-1})\right]\right)\right| \\
=&\left|\sum_{m=1}^{K}\left(\left[F_{B^{W}_{\gamma^*,i}}(x_m)-F_{B^{W}_{\gamma^*,i}}(x_{m-1})\right]\left[F_{A_i^n}(x_K)-F_{A_i^n}(x_{m-1})-F_{A_i^*}(x_K)+F_{A_i^*}(x_{m-1})\right]\right)\right| \\
\leq&\sum_{m=1}^{K}\left(F_{B^{W}_{\gamma^*,i}}(x_m)-F_{B^{W}_{\gamma^*,i}}(x_{m-1})\right)\cdot 2\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right| \\
\leq&\sum_{m=1}^{K}\frac{M}{K}\cdot 2\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right| \\
=&2M\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right|. \qquad (45)
\end{aligned}
$$

Inject (44) and (45) into (43), we obtain that

$$
\left|\mathbb{E}_{\mathcal{I}_i} g\left(A_i^n\right)-\mathbb{E}_{\mathcal{I}_i} g\left(A_i^*\right)\right| \leq (2+2M)\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right|. \qquad (46)
$$

Apply the triangle inequality and combine (41), (42), (46), we have that:

$$
\left|\mathbb{E}_{\mathcal{I}_i} F_{B^{W}_{\gamma^*,i}}(A_i^n)-\mathbb{E}_{\mathcal{I}_i} F_{B^{W}_{\gamma^*,i}}(A_i^*)\right| \leq 2\delta_2+(2+2M)\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right|.
$$

From this and (39), we obtain that

$$
|\mathbb{E}F_{B_i^*}(A_i^n)-\mathbb{E}F_{B_i^*}(A_i^*)| \leq 2\delta_2+(4+2M)\sup_{x\in[0,\infty)}\left|F_{A_i^n}(x)-F_{A_i^*}(x)\right|. \qquad (47)
$$

Recall the constant C_1 indicated in Lemma A.6, we define $C_2 \coloneqq C_1(6+2M)^2\left[\ln(6+2M)+1\right]$ (note that C_2 does not depend on n nor ε_2) and deduce that $C_2\varepsilon_2^{-2}\ln\left(\frac{1}{\min\{\varepsilon_2,1/e\}}\right) \geq C_1\delta_2^{-2}\ln\left(\frac{1}{\min\{\delta_2,1/e\}}\right)$.[22] Take $\varepsilon_1 := \delta_2$, for any $n \geq C_2\varepsilon_2^{-2}\ln\left(\frac{1}{\min\{\varepsilon_2,1/e\}}\right)$, we have

[22] Apply Lemma A.2, we have $C_2\varepsilon_2^{-2}\ln\left(\frac{1}{\min\{\varepsilon_2,1/e\}}\right) = C_1\left(\frac{6+2M}{\varepsilon_2}\right)^2\left[\ln(6+2M)+1\right]\ln\left(\frac{1}{\min\{\varepsilon_2,1/e\}}\right) \geq C_1\left(\frac{6+2M}{\varepsilon_2}\right)^2\ln\left(\frac{6+2M}{\varepsilon_2}\right)$. Moreover, since $\frac{\varepsilon_2}{6+2M} = \min\left\{\frac{\varepsilon_2}{6+2M},\frac{1}{e}\right\} = \min\left\{\delta_2,\frac{1}{e}\right\}$ (due to the fact that $\delta_2 = \varepsilon_2/(6+2M) < 1/e$). Therefore, we have $C_2\varepsilon_2^{-2}\ln\left(\frac{1}{\min\{\varepsilon_2,1/e\}}\right) \geq C_1\delta_2^{-2}\ln\left(\frac{1}{\min\{\delta_2,1/e\}}\right)$.

$n \geq C_1 \varepsilon_1^{-2} \ln \left(\frac{1}{\min\{\varepsilon_1, 1/e\}} \right)$ and by Lemma A.6, we have $\sup_{x \in [0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right| \leq \varepsilon_1 = \delta_2$ and thus by applying (47), we obtain:

$$|\mathbb{E} F_{B_i^*}(A_i^n) - \mathbb{E} F_{B_i^*}(A_i^*)| \leq 2\delta_2 + (4 + 2M)\,\delta_2 = (6 + 2M)\delta_2 = \varepsilon_2.$$

This is exactly (22). We can have a similar result in the case where $i \notin \Omega_A(\gamma^*)$ (its proof is omitted here) and we conclude the proof of this lemma.

Appendix B

Supplementary Materials for Chapter 5 on the $\mathcal{DCB}_n^{m,p}$ Game

To prove Lemma 5.4.1, we first need the uniform convergence of the continuous distributions $F_{\tilde{A}_i^n}$ and $F_{\tilde{B}_i^n}$ towards $F_{\tilde{A}_i^*}$ and $F_{\tilde{B}_i^*}$ (defined in (5.5) and (5.1), (5.2)). This is stated as follows.

Lemma B.1. *Fix $\psi \geq 1$, for any $\varepsilon_c \in (0,1]$, there exists an $N^* := \mathcal{O}\left(\varepsilon_c^{-2}\ln(\varepsilon_c^{-1})\right)$ such that for any $n \geq N^*$ and $i \in [n]$, we have*

$$\sup_{x\in[0,\infty)} \left|F_{\tilde{A}_i^n}(x) - F_{\tilde{A}_i^*}(x)\right| \leq \varepsilon_c \text{ and } \sup_{x\in[0,\infty)} \left|F_{\tilde{B}_i^n}(x) - F_{\tilde{B}_i^*}(x)\right| \leq \varepsilon_c.$$

Proof. Choosing $\delta := \frac{\varepsilon_c \underline{w} \psi^2}{n\bar{w}}$, for any $i \in [n]$, we have

$$\begin{aligned}
P\left(\left|\tilde{A}_i^n - \tilde{A}_i^*\right| > \delta\right) &= P\left(\left|\frac{\tilde{A}_i^*}{\sum_{j=1}^n \tilde{A}_j^*} - \tilde{A}_i^*\right| > \delta\right)\\
&= P\left(\left|\tilde{A}_i^*\left(1 - \sum_{j=1}^n \tilde{A}_j^*\right)\right| > \delta\left|\sum_{j=1}^n \tilde{A}_j^*\right|\right)\\
&\leq P\left(\left|\tilde{A}_i^*\left(1 - \sum_{j=1}^n \tilde{A}_j^*\right)\right| > \delta - \delta\left|1 - \sum_{j=1}^n \tilde{A}_j^*\right|\right)\\
&= P\left(\left|\left(\tilde{A}_i^* + \delta\right)\left(1 - \sum_{j=1}^n \tilde{A}_j^*\right)\right| > \delta\right)\\
&\leq P\left(\left|1 - \sum_{j=1}^n \tilde{A}_j^*\right| > \frac{\delta}{2\frac{\bar{w}\psi}{n\underline{w}} + \delta}\right).
\end{aligned} \tag{48}$$

The last equality comes from the fact that the random variable $\tilde{A}_i^*$ is upper bounded by $2\frac{w_i}{W}\psi \leq 2\frac{\bar{w}}{n\underline{w}}\psi, \forall i$. Let $\tau := \frac{\delta}{2\frac{\bar{w}\psi}{n\underline{w}} + \delta}$. Using the fact that $E\left[\sum_{j=1}^n \tilde{A}_j^*\right] = 1$ and applying the Hoeffding's inequality on bounded random variables $\{\tilde{A}_i^*\}_{i\in[n]}$ (in the range $\left[0, 2\frac{\bar{w}}{n\underline{w}}\psi\right]$), we have

$$P\left(\left|1 - \sum_{j=1}^n \tilde{A}_j^*\right| > \tau\right) \leq P\left(\left|E\left[\sum_{j=1}^n \tilde{A}_j^*\right] - \sum_{j=1}^n \tilde{A}_j^*\right| \geq \tau\right)$$

$$\le 2\exp\left(-\frac{2\tau^2}{\sum_{i=1}^{n}\left(\frac{2w_i}{W}\psi\right)^2}\right)$$
$$\le 2\exp\left(-\frac{2\tau^2}{\psi^2 n\frac{4}{n^2}\left(\frac{\bar{w}}{\underline{w}}\right)^2}\right)$$
$$= 2\exp\left(-\frac{\tau^2 n}{2\psi^2}\left(\frac{\bar{w}}{\underline{w}}\right)^2\right).$$

On the other hand,

$$\frac{1}{\tau^2} = \left(\frac{2\bar{w}\psi}{n\underline{w}\delta}+1\right)^2$$
$$= \left(\frac{2\bar{w}\psi}{n\underline{w}\frac{\varepsilon_c\underline{w}\psi^2}{n\bar{w}}}+1\right)^2$$
$$= \left(2\left(\frac{\bar{w}}{\underline{w}}\right)^2\frac{1}{\varepsilon_c\psi}+1\right)^2$$
$$\le U := \frac{1}{\varepsilon_c^2}\left[\frac{2}{\psi}\left(\frac{\bar{w}}{\underline{w}}\right)^2+1\right]^2,\ \text{since } \varepsilon_c^2 \le \varepsilon_c \le 1.$$

That means, $\tau^2 \ge \frac{1}{U}$. Therefore, by the fact that $E[\sum_{j=1}^{n}\tilde{A}_j^*] = 1$, we have:

$$P\left(\left|\tilde{A}_i^n - \tilde{A}_i^*\right| > \delta\right) \le 2\exp\left(-\frac{n}{2U\psi^2}\left(\frac{\bar{w}}{\underline{w}}\right)^2\right)$$
$$\le \frac{\varepsilon_c}{2}, \forall n \ge N^*$$

where

$$N^* := 2U\psi^2\ln\left(\frac{4}{\varepsilon_c}\right)\left(\frac{\underline{w}}{\bar{w}}\right)^2 = \frac{2}{\varepsilon_c^2}\left[2\left(\frac{\bar{w}}{\underline{w}}\right)^2+\psi\right]^2\ln\left(\frac{4}{\varepsilon_c}\right)\left(\frac{\underline{w}}{\bar{w}}\right)^2 = O(\varepsilon_c^{-2}\ln(\varepsilon_c^{-1})).$$

Now, for any $i \in [n]$ and $x \in [0,\infty)$,

$$F_{\tilde{A}_i^n}(x) - F_{\tilde{A}_i^*}(x) = P(\{\tilde{A}_i^n \le x\}) - F_{\tilde{A}_i^*}(x)$$
$$\le P\left(\{\tilde{A}_i^n \le x\}\cap\{|\tilde{A}_i^n - \tilde{A}_i^*| \le \delta\}\right) + P\left(\left|\tilde{A}_i^n - \tilde{A}_i^*\right| > \delta\right) - F_{\tilde{A}_i^*}(x)$$
$$\le P\left(\{\tilde{A}_i^n \le x\}\cap\{\tilde{A}_i^* \le \tilde{A}_i^n + \delta\}\right) + P\left(\left|\tilde{A}_i^n - \tilde{A}_i^*\right| > \delta\right) - F_{\tilde{A}_i^*}(x)$$
$$\le P\left(\tilde{A}_i^* \le x + \delta\right) + \frac{\varepsilon_c}{2} - F_{\tilde{A}_i^*}(x), \forall n \ge N^*$$

$$= F_{\tilde{A}_i^*}(x+\delta) - F_{\tilde{A}_i^*}(x) + \frac{\varepsilon_c}{2}, \forall n \geq N^*$$
$$\leq \frac{\delta}{2\frac{w_i}{W}\psi^2} + \frac{\varepsilon_c}{2}, \forall n \geq N^*$$
$$\leq \varepsilon_c, \forall n \geq N^*.$$

Similarly, we can deduce the inequality $F_{\tilde{A}_i^n}(x) - F_{\tilde{A}_i^*}(x) \geq -\varepsilon_c, \forall x \in [0,\infty), \forall n \geq N^*$, $\forall i$ and we conclude that $\sup_{x\in[0,\infty)} \left| F_{\tilde{A}_i^n}(x) - F_{\tilde{A}_i^*}(x) \right| \leq \varepsilon_c$.

The inequality corresponding to player B $\sup_{x\in[0,\infty)} \left| F_{\tilde{B}_i^n}(x) - F_{\tilde{B}_i^*}(x) \right| \leq \varepsilon_c$ can be proved in a similar way (the bounds in the above proof are chosen for both random variables $\{\tilde{A}_i^*\}_i$ and $\{\tilde{B}_i^*\}_i$, thus, the precise definition of N^* given above also works to prove this inequality). □

Proof of Lemma 5.4.1

Lemma 5.4.1. *Fix $\psi \geq 1$, for any $\bar{\varepsilon}_1 \in (0,1]$, there exists $N^* := \mathcal{O}\left(\bar{\varepsilon}_1^{-2} \ln\left(\bar{\varepsilon}_1^{-1}\right)\right)$, such that for any $n \geq N^*$, there exists $M_0 := \mathcal{O}(n/\bar{\varepsilon}_1)$, such that for any $m \geq M_0$ and $i \in \{1,2,\dots,n\}$, we have*

$$\sup_{\hat{x}\in\mathbb{N}} \left| F_{A_i^D}(\hat{x}) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) \right| < \bar{\varepsilon}_1 \text{ and } \sup_{\hat{x}\in\mathbb{N}} \left| F_{B_i^D}(\hat{x}) - F_{\tilde{B}_i^*}\left(\frac{\hat{x}}{m}\right) \right| < \bar{\varepsilon}_1.$$

Proof. For any random variables U and V, from the definition of rounding function r^m, for any $m, \hat{x} \in \mathbb{N}$, we have[23]

$$U - V \leq \frac{\hat{x}}{m} \Rightarrow r^m(U) - r^m(V) \leq \frac{\hat{x}}{m} \Rightarrow U - V < \frac{\hat{x}+1}{m},$$

which induces

$$P\left(U - V \leq \frac{\hat{x}}{m}\right) \leq P\left(r^m(U) - r^m(V) \leq \frac{\hat{x}}{m}\right) \leq P\left(U - V \leq \frac{\hat{x}+1}{m}\right).$$

Therefore, for each $i \in [n]$, by replacing $U := \sum_{k=1}^{i} \tilde{A}_k^n$ and $V := \sum_{k=1}^{i-1} \tilde{A}_k^n$ together with definition of A_i^D given in (5.3), for any $\hat{x} \in \mathbb{N}$, we have

$$F_{\tilde{A}_i^n}\left(\frac{\hat{x}}{m}\right) \leq F_{A_i^D}(\hat{x}) \leq F_{\tilde{A}_i^n}\left(\frac{\hat{x}+1}{m}\right).$$

On the other hand, applying Lemma B.1 with $\varepsilon_c := \bar{\varepsilon}_1/2$, for any $n \geq N^*$ where $N^* := \mathcal{O}\left(\varepsilon_c^{-2}\ln(\varepsilon_c^{-1})\right) = \mathcal{O}\left(\bar{\varepsilon}_1^{-2}\ln(\bar{\varepsilon}_1^{-1})\right)$, for any $i \in [n]$, we have

$$\left| F_{\tilde{A}_i^n}\left(\frac{\hat{x}}{m}\right) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) \right| < \frac{\bar{\varepsilon}_1}{2} \text{ and } \left| F_{\tilde{A}_i^n}\left(\frac{\hat{x}+1}{m}\right) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}+1}{m}\right) \right| < \frac{\bar{\varepsilon}_1}{2}.$$

[23] If $U - V \leq \frac{\hat{x}}{m}$, then $r^m(U) \leq r^m(V + \hat{x}/m) = r^m(V) + \hat{x}/m$. If $r^m(U) - r^m(V) \leq \hat{x}/m$, then $U < r^m(U) + \frac{1}{2m} \leq r^m(V) + \frac{\hat{x}}{m} + \frac{1}{2m} \leq V + \frac{\hat{x}+1}{m}$.

Therefore, for any $n \geq N^*$, for any $\hat{x}, m \in \mathbb{N}$, we have that

$$
\begin{aligned}
&F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) - \frac{\bar{\varepsilon}_1}{2} < F_{A_i^D}(\hat{x}) < F_{\tilde{A}_i^*}\left(\frac{\hat{x}+1}{m}\right) + \frac{\bar{\varepsilon}_1}{2} \\
\Rightarrow &F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) - \frac{\bar{\varepsilon}_1}{2} < F_{A_i^D}(\hat{x}) < F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right) + \frac{1}{m}\frac{W}{2w_i\psi^2} + \frac{\bar{\varepsilon}_1}{2}.
\end{aligned} \tag{49}
$$

The last inequality in (49) is trivially deduced from the definition of $\tilde{A}_i^*$. Now, we can choose $M_0 := \frac{nv_{\max}}{\bar{\varepsilon}_1 v_{\min}\psi^2} = O(n/\bar{\varepsilon}_1)$; thus, $\forall m \geq M_0$, $\frac{1}{m}\frac{W}{2w_i\psi^2} \leq \frac{1}{M_0}\frac{W}{2w_i\psi^2} = \frac{\bar{\varepsilon}_1 v_{\min}\psi^2}{nv_{\max}}\frac{W}{2w_i\psi^2} \leq \frac{\bar{\varepsilon}_1}{2}$. Combining with (49), for any $n \geq N^*$, $m \geq M_0$, we have $\sup_{\hat{x}\in\mathbb{N}}\left|F_{A_i^D}(\hat{x}) - F_{\tilde{A}_i^*}\left(\frac{\hat{x}}{m}\right)\right| < \bar{\varepsilon}_1$.

The inequality with respect to $F_{B_i^D}$ and $F_{\tilde{B}_i^*}$ can be similarly proven. □

Proof of Lemma 5.4.2

Lemma 5.4.2. *Fix $\psi \geq 1$. For any $\varepsilon' \in (0,1]$ and $n \geq N^*$, there exists an $M_2 := O(n/\varepsilon')$, such that for any $m \geq M_2$, we have*

$$
\sum_{i=1}^{n}\left[w_i \sum_{\hat{y}=0}^{\lceil 2\frac{w_i}{W}p\rceil} F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right] \geq \frac{W}{2\psi} - \sum_{i=1}^{n}\frac{W}{2\psi m} - \varepsilon' W. \tag{5.19}
$$

Proof. Recalling that $\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) = F_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) - F_{\tilde{A}_i^*}\left(\frac{\hat{y}-1}{m}\right)$, to ease the notation, for any $\hat{y} = 1, 2, \ldots, \lceil 2\frac{w_i}{W}p\rceil - 1$, we denote

$$
F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right) = \frac{(\hat{y}-1)(W)^2}{\psi m^2 (2w_i\psi)^2} := g(\hat{y}),
$$

while $F_{\tilde{B}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil - 1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil}{m}\right)$ does not have the form $g\left(\lceil 2\frac{w_i}{W}p\rceil\right)$.

Then, we have

$$
\begin{aligned}
&\sum_{\hat{y}=0}^{\lceil \frac{2w_i}{W}p\rceil}\left[F_{\tilde{B}_i^*}\left(\frac{\hat{y}-1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\hat{y}}{m}\right)\right] \\
=&\sum_{\hat{y}=1}^{\lceil \frac{2w_i}{W}p\rceil - 1} g(\hat{y}) + F_{\tilde{B}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil - 1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil}{m}\right) \\
=&\sum_{\hat{y}=1}^{\lceil \frac{2w_i}{W}p\rceil} g(\hat{y}) + \left[F_{\tilde{B}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil - 1}{m}\right)\tilde{F}_{\tilde{A}_i^*}\left(\frac{\lceil 2\frac{w_i}{W}p\rceil}{m}\right) - g\left(\left\lceil 2\frac{w_i}{W}p\right\rceil\right)\right] \\
:=&E_1 + E_2.
\end{aligned} \tag{50}
$$

Here, the term E_1 will be bounded from below by an approximation of the upper bound given in (5.13), which is the objective of this lemma. Indeed, for any $m \geq M_{E_1} := \frac{n v_{\max}}{\sqrt{2\psi \varepsilon' v_{\min} \psi}} \leq \mathcal{O}\left(n/\varepsilon'\right)$, we have

$$E_1 := \sum_{\hat{y}=1}^{\left\lceil \frac{2w_i}{W} p \right\rceil} \mathrm{g}\left(\hat{y}\right) = \frac{(W)^2}{\psi m^2 \left(2 w_i \psi\right)^2} \sum_{\hat{y}=1}^{\left\lceil \frac{2w_i}{W} p \right\rceil} \left(\hat{y} - 1\right)$$
$$= \frac{(W)^2}{\psi m^2 \left(2 w_i \psi\right)^2} \sum_{\hat{y}=1}^{\left\lceil \frac{2w_i}{W} p \right\rceil} \hat{y} - \frac{(W)^2}{\psi m^2 \left(2 w_i \psi\right)^2} \sum_{\hat{y}=1}^{\left\lceil \frac{2w_i}{W} p \right\rceil} 1$$
$$= \frac{(W)^2}{\psi m^2 \left(2 w_i \psi\right)^2} \left(\left\lceil \frac{2w_i}{W} p \right\rceil + 1 \right) \left\lceil \frac{2w_i}{W} p \right\rceil - \frac{(W)^2}{\psi m^2 \left(2 w_i \psi\right)^2} \left\lceil \frac{2w_i}{W} p \right\rceil$$
$$\geq \frac{\left(\frac{2w_i}{W} \frac{p}{m} \right)^2}{\left(2 \frac{w_i}{W} \psi\right)^2} \frac{1}{2\psi} - \frac{2 \frac{w_i}{W} p + 1}{\psi m^2 \left(2 \frac{w_i}{W} \psi\right)^2} \geq \frac{1}{2\psi} - \frac{1}{m} \frac{W}{2 w_i \psi} - \frac{\varepsilon'}{2}. \tag{51}$$

Here, the last inequality in (51) comes from $\psi := \frac{p}{m} \geq 1$ and $m \geq M_{E_1}$.

Similarly, we can prove that for any $m \geq M_{E_2} := \frac{n v_{\max}}{\varepsilon' \psi^2 v_{\min}} = \mathcal{O}\left(n/\varepsilon'\right)$, since $p = \psi m$, we have

$$E_2 := \left[F_{\tilde{B}_i^*} \left(\frac{\left\lceil 2 \frac{w_i}{W} p \right\rceil - 1}{m} \right) \tilde{F}_{\tilde{A}_i^*} \left(\frac{\left\lceil 2 \frac{w_i}{W} p \right\rceil}{m} \right) - \mathrm{g}\left(\left\lceil 2 \frac{w_i}{W} p \right\rceil \right) \right]$$
$$= \left(\left\lceil \frac{2w_i}{W} p \right\rceil - 1 \right) \frac{W}{2 w_i m \psi} \left(\frac{1}{\psi} - \left\lceil \frac{2w_i}{W} p \right\rceil \frac{W}{2 w_i m \psi^2} \right)$$
$$\geq \left(\left\lceil \frac{2w_i}{W} p \right\rceil - 1 \right) \frac{W}{2 w_i m \psi} \left(\frac{1}{\psi} - \frac{2 w_i p}{W} \frac{W}{2 w_i m \psi^2} - \frac{W}{2 w_i m \psi^2} \right)$$
$$= \left(\left\lceil \frac{2w_i}{W} p \right\rceil - 1 \right) \frac{W}{2 w_i m \psi} \left(- \frac{W}{2 w_i m \psi^2} \right)$$
$$\geq - \left(\frac{2w_i}{W} p \right) \frac{W}{2 w_i m \psi} \left(\frac{W}{2 w_i m \psi^2} \right)$$
$$\geq - \frac{\varepsilon'}{2}.$$

Combining this with the inequalities (50) and (51), for any $n \geq N^*$ and $m \geq M_2$ where $M_2 := \max\left\{M_{E_1}, M_{E_2}\right\} = \mathcal{O}\left(\frac{n}{\varepsilon'}\right)$, we conclude that

$$\sum_{i=1}^{n} \left[w_i \sum_{\hat{y}=0}^{\left\lceil 2 \frac{w_i}{W} p \right\rceil} F_{\tilde{B}_i^*} \left(\frac{\hat{y} - 1}{m} \right) \tilde{F}_{\tilde{A}_i^*} \left(\frac{\hat{y}}{m} \right) \right] \geq \sum_{i=1}^{n} w_i \left(\frac{1}{2\psi} - \frac{1}{m} \frac{W}{2 w_i \psi} - \varepsilon' \right),$$

which is exactly (5.19). □

Appendix C

Supplementary Materials for Section 6.1 on the $\mathcal{LB}_n$ Game

C.1 Proof of Result in Section 6.1 on Lottery Blotto Games

Proof of Lemma 6.1.2

Lemma 6.1.2. *There exists $L_0 > 0$, such that for any $\varepsilon \in (0,1]$, any $n \geq L_0 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/\varepsilon\}}\right)$ and any game $\mathcal{LB}_n(\zeta)$, $y^* \in \mathcal{S}_n^{(4.5)}$, $\delta \in \Delta_{y^*}(\zeta, \varepsilon)$ and $i \in [n]$, we have:*

$$\max\left\{\sup_{y^* \in [0,2X^B]} \int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^n}(x), \sup_{x^* \in [0,2X^B]} \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^n}(y)\right\} \leq \delta + \varepsilon. \qquad (6.3)$$

Fix $y^* \in [0, 2X_B]$, we look for the condition on n such that $\int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^n}(x) \leq \delta + \varepsilon$ holds. The condition corresponding to the inequality $\int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^n}(x) \leq \varepsilon + \delta$ with $x^* \in [0, 2X_B]$ can be proved similarly and thus is omitted in this section.

First, we note that if $\mathcal{X}_\zeta(y^*, \varepsilon)$ is empty, $\int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^n}(x) = 0$ and the result trivially holds. Now, let us assume that $\mathcal{X}_\zeta(y^*, \varepsilon) \neq \emptyset$, we can write $\mathcal{X}_\zeta(y^*, \varepsilon) = I_1 \bigcup I_2 \bigcup I_3$ with[24]

$$\begin{aligned}
I_1 &:= \{x \in [0, 2X_B] : x = y^*, |\zeta_A(x, y^*) - \alpha| \geq \varepsilon\},\\
I_2 &:= \{x \in [0, 2X_B] : x < y^*, \zeta_A(x, y^*) \geq \varepsilon\},\\
I_3 &:= \{x \in [0, 2X_B] : x > y^*, 1 - \zeta_A(x, y^*) \geq \varepsilon\}.
\end{aligned}$$

It is trivial that I_1 is either an empty set or a singleton; on the other hand, due to the monotonicity of the CSF ζ_A (see $(C2)$, Definition 3.2.2), I_2 and I_3 are either empty sets or half intervals. Moreover, for any arbitrary distribution F, we have that

$$\int_{x \in I'} \mathrm{d}F(x) = \begin{cases} 0\text{ , if } I' = \emptyset, \\ F(a)\text{ , if } I' = \{a\}, \text{i.e., } I' \text{ is a singleton}, \\ F(b) - F(a)\text{ , if } I' = (a, b], \text{i.e., } I' \text{ is a half interval.} \end{cases}$$

Therefore, we can deduce that

$$\begin{aligned}
\int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^n}(x) - \int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^*}(x) &= \sum_{j=1}^{3} \left(\int_{I_j} \mathrm{d}F_{A_i^n}(x) - \int_{I_j} \mathrm{d}F_{A_i^*}(x)\right)\\
&\leq 5 \sup_{x \in [0,\infty)} |F_{A_i^n}(x) - F_{A_i^*}(x)|.
\end{aligned}$$

[24]Recall that by definition, $\beta_A(x, y^*) = \alpha$ if $x = y^*$, $\beta_A(x, y^*) = 0$ if $x < y^*$ and $\beta_A(x, y^*) = 1$ if $x > y^*$

Recall the constant C_1 indicated in Lemma A.6, we define $L_0 := C_1 5^2(\ln(5)+1)$. Note that L_0 does not depend on the choice of y^*. Take $\varepsilon_1 := \varepsilon/5$, we can deduce that $L_0\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1\varepsilon_1^2\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$.[25] Therefore, for any $n \geq L_0\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have $n \geq C_1\varepsilon_1^2\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$ and by Lemma A.6, $\sup_{x\in[0,\infty)} |F_{A_i^n}(x) - F_{A_i^*}(x)| \leq \varepsilon_1 = \varepsilon/5$.

Hence, for any $n \geq L_0\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $\delta \in \Delta_{\gamma^*}(\zeta,\varepsilon)$,

$$\int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^n}(x) \leq \int_{\mathcal{X}_\zeta(y^*,\varepsilon)} \mathrm{d}F_{A_i^*}(x) + 5\cdot\varepsilon/5 \leq \delta + \varepsilon.$$

Proof of Theorem 6.1.3

Theorem 6.1.3. ***(Approximate equilibria of the generalized Lottery Blotto game).***

(i) In any game $\mathcal{LB}_n(\zeta)$, there exists a positive number $\varepsilon \leq \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and $\delta \in \Delta_{\gamma^*}(\zeta,\varepsilon)$, the following inequalities hold for any pure strategy $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ of players A and B:[26]

$$\Pi_\zeta^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*}) \leq \Pi_\zeta^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (8\delta + 13\varepsilon)\,W^A, \tag{6.4}$$

$$\Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \boldsymbol{x}^B) \leq \Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (8\delta + 13\varepsilon)\,W^B. \tag{6.5}$$

(ii) There exists $L^* > 0$, such that for any $\varepsilon \in (0,1]$ and in any game $\mathcal{LB}_n(\zeta)$ where $n \geq L^*\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, (6.4) and (6.5) hold for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, $\delta \in \Delta_{\gamma^*}(\zeta,\varepsilon)$ and any pure strategy $\boldsymbol{x}^A, \boldsymbol{x}^B$ of players A and B.

Proof. We first give the proof of Result (ii). For the sake of brevity, we only focus on (6.4). The proof that (6.5) holds under the same condition can be done similarly and thus is omitted. Note that in this proof, we often use the Fubini's Theorem to exchange the order of the double integrals.

Recall that $\boldsymbol{x}^A = (x_i^A)_{i\in[n]}$, by the definition of the payoff functions in $\mathcal{LB}_n(\zeta)$, (6.4) can be rewritten as

$$\sum_{i=1}^n\left(v_i^A\int_0^\infty \zeta_A\left(x_i^A, y\right)\mathrm{d}F_{B_i^n}\left(y\right)\right) - \sum_{i=1}^n\left(v_i^A\int_0^\infty\int_0^\infty \zeta_A\left(x,y\right)\mathrm{d}F_{A_i^n}\left(x\right)\mathrm{d}F_{B_i^n}\left(y\right)\right) \leq 8\delta + 13\varepsilon. \tag{52}$$

We now prove that (52) holds under appropriate parameters values. To do this, we prepare two useful lemmas as follows.

[25] Note that $\varepsilon_1 = \frac{\varepsilon}{5}$ and apply Lemma A.2, $L_0\cdot\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) = C_1\left(\frac{5}{\varepsilon}\right)^2\cdot(\ln(5)+1)\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1\cdot\left(\frac{5}{\varepsilon}\right)^2\ln\left(\frac{5}{\varepsilon}\right)$; moreover, $\frac{\varepsilon}{5} = \min\{\frac{\varepsilon}{5},\frac{1}{e}\}$ since $\varepsilon \leq 1$; thus, we can rewrite $\ln\left(\frac{5}{\varepsilon}\right) = \ln\left(\frac{1}{\min\{\varepsilon/5,1/e\}}\right) = \ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$.

[26] Recall that Π_ζ^A and Π_ζ^B denote the payoffs functions of players A and B in the game $\mathcal{LB}_n(\zeta)$.

Lemma C.1. *For any pair of CSFs $\zeta = (\zeta_A, \zeta_B)$, any $\varepsilon \in (0,1]$ and $x^* \in [0, 2X_B]$, the following results hold:*

(i) *For any n, $i \in [n]$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$,*

$$\left| \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^*}\left(y\right) - \int_0^\infty \beta_A\left(x^*, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \right| \leq \delta + \varepsilon. \tag{53}$$

(ii) *There exists a constant $L_1 > 0$ such that for any $n \geq L_1 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, $i \in [n]$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$,*

$$\left| \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^n}\left(y\right) - \int_0^\infty \beta_A\left(x^*, y\right) \mathrm{d}F_{B_i^n}\left(y\right) \right| \leq \delta + 2\varepsilon. \tag{54}$$

Lemma C.2. *Given $\varepsilon \in (0,1]$, there exists $L_2 > 0$ such that for any $n \geq L_2 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, any game $\mathcal{LB}_n(\zeta)$, any $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and $i \in [n]$, we have:*

$$\left| \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^n}\left(y\right) - \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \right| \leq 2\delta + 4\varepsilon, \forall x \geq 0, \tag{55}$$

$$\left| \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{A_i^*}(x) \mathrm{d}F_{B_i^*}\left(y\right) - \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{A_i^*}(x) \mathrm{d}F_{B_i^n}\left(y\right) \right| \leq 2\delta + 3\varepsilon, \tag{56}$$

$$\left| \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^n}\left(y\right) \mathrm{d}F_{A_i^*}(x) - \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^n}\left(y\right) \mathrm{d}F_{A_i^n}(x) \right| \leq 2\delta + 4\varepsilon. \tag{57}$$

Lemma C.1 states the relation between the first term appearing in the left-hand-side of (52) and the corresponding terms when we replace the CSF ζ by the Blotto functions β and replace $F_{B_i^n}$ by $F_{B_i^*}$. A proof of Lemma C.1 is given in subsubsection C.1. On the other hand, Lemma C.2 indicates several useful inequalities involving the players' payoffs in the game $\mathcal{LB}_n$ (when they play according to the IU$^{\gamma^*}$ strategy or playing such that the marginals are $F_{A_i^*}, F_{B_i^*}$). Its proof is given in subsubsection C.1 that is based on Lemma C.1 and the convergence of the distributions $F_{A_i^n}, F_{B_i^n}$ toward $F_{A_i^*}, F_{B_i^*}$ (i.e., Lemma A.6).

We have another remark: for any n and $i \in [n]$,

$$\mathbb{P}(A_i^* = B_i^* = x) = 0, \forall x \geq 0. \tag{58}$$

This can be trivially proved as follows: first, $\mathbb{P}(A_i^* = B_i^* = x) = \mathbb{P}(A_i^* = x)\mathbb{P}(B_i^* = x)$ since they are independent; now, if $x > 0$, both $F_{A_i^*}$ and $F_{B_i^*}$ are continuous at x and thus $\mathbb{P}(A_i^* = x) = \mathbb{P}(B_i^* = x) = 0$; on the other hand, if $x = 0$, in the case where $i \in \Omega_A(\gamma^*)$, since $A_i^* = A_{\gamma^*,i}^S$, we have $\mathbb{P}(A_i^* = x) = 0$, in the case where $i \notin \Omega_A(\gamma^*)$, since $B_i^* = B_{\gamma^*,i}^S$, we have $\mathbb{P}(B_i^* = x) = 0$.

Finally, use Lemma C.1 and Lemma C.2 and take $L^* = \max\{L_1, L_2\}$, for any $n \geq L^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and any pure strategy $\boldsymbol{x}^A$ of player A, we have:

$$
\begin{aligned}
&\sum_{i=1}^{n}\left(v_i^A \int_0^\infty \zeta_A\left(x_i^A, y\right) \mathrm{d}F_{B_i^n}\left(y\right)\right)\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \zeta_A\left(x_i^A, y\right) \mathrm{d}F_{B_i^*}\left(y\right)\right) + \sum_{i=1}^{n} v_i^A(2\delta + 4\varepsilon) && \text{(due to (55))}\\
&= \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \zeta_A\left(x_i^A, y\right) \mathrm{d}F_{B_i^*}\left(y\right)\right) + 2\delta + 4\varepsilon && \text{(note that } \textstyle\sum_{i=1}^{n} v_i^A = 1)\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \beta_A\left(x_i^A, y\right) \mathrm{d}F_{B_i^*}\left(y\right)\right) + 3\delta + 5\varepsilon && \text{(due to (53))}\\
&= \sum_{i=1}^{n}\left[v_i^A\left(\alpha\mathbb{P}(B_i^* = x_i^A) + \mathbb{P}(B_i^* < x_i^A)\right)\right] + 3\delta + 5\varepsilon\\
&\leq \sum_{i=1}^{n} v_i^A F_{B_i^*}\left(x_i^A\right) + 3\delta + 5\varepsilon && \text{(since } \alpha \leq 1)\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty F_{B_i^*}(x) \mathrm{d}F_{A_i^*}(x)\right) + 3\delta + 5\varepsilon && \text{(due to Lemma A.5)}\\
&= \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \mathbb{P}(B_i^* < x) \mathrm{d}F_{A_i^*}(x)\right) + 3\delta + 5\varepsilon && \text{(due to (58))}\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \int_0^\infty \beta_A\left(x, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \mathrm{d}F_{A_i^*}\left(x\right)\right) + 3\delta + 5\varepsilon\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \mathrm{d}F_{A_i^*}\left(x\right)\right) + 4\delta + 6\varepsilon && \text{(due to (53))}\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{A_i^*}\left(x\right) \mathrm{d}F_{B_i^n}\left(y\right)\right) + 6\delta + 9\varepsilon && \text{(due to (56))}\\
&\leq \sum_{i=1}^{n}\left(v_i^A \int_0^\infty \int_0^\infty \zeta_A\left(x, y\right) \mathrm{d}F_{B_i^n}\left(y\right) \mathrm{d}F_{A_i^n}\left(x\right)\right) + 8\delta + 13\varepsilon && \text{(due to (57)).}
\end{aligned}
$$

Hence, we conclude that for $n \geq L^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$ (52) holds and thus, (6.4) also holds.

To prove that Result (ii) implies Result (i), we can proceed similarly to the proof that Theorem 4.2.3-(ii) implies Theorem 4.2.3-(i) (see Section A.2). We conclude this proof.

□

Proof of Lemma C.1

First, we prove (53). Note that $F_{B_i^*}(y) = 1, \forall y > 2X_B$ (see Lemma A.1-(iv)), for any n, $i \in [n]$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, we have

$$\begin{aligned}
&\left| \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^*}\left(y\right) - \int_0^\infty \beta_A\left(x^*, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \right| \\
\le& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|\zeta_A(x^*, y) - \beta_A(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) + \int_{[0,\infty)\backslash\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|\zeta_A(x^*, y) - \beta_A(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) \\
=& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|1 - \zeta_B(x^*, y) - 1 + \beta_B(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) \\
&\qquad + \int_{[0,2X_B]\backslash\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|1 - \zeta_B(x^*, y) - 1 + \beta_B(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) \\
=& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|\zeta_B(x^*, y) - \beta_B(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) + \int_{[0,2X_B]\backslash\mathcal{Y}_\zeta(x^*,\varepsilon)} \left|\zeta_B(x^*, y) - \beta_B(x^*, y)\right| \mathrm{d}F_{B_i^*}\left(y\right) \\
\le& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^*}\left(y\right) + \int_{[0,2X_B]\backslash\mathcal{Y}_\zeta(x^*,\varepsilon)} \varepsilon \mathrm{d}F_{B_i^*}\left(y\right) \\
\le& \delta + \varepsilon.
\end{aligned} \tag{59}$$

Here, the second-to-last inequality comes from the fact that $0 \le \zeta_B(x, y), \beta_B(x, y) \le 1$ for any x, y and the definition of $\mathcal{Y}_\zeta(x^*, \varepsilon)$ while the last inequality is due to the definition of $\Delta_{\gamma^*}(\zeta, \varepsilon)$.

Now, in order to prove (54), we proceed similarly as in (59) to show that

$$\begin{aligned}
&\left| \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^n}\left(y\right) - \int_0^\infty \beta_A\left(x^*, y\right) \mathrm{d}F_{B_i^n}\left(y\right) \right| \\
\le& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^n}\left(y\right) + \int_{[0,2X_B]\backslash\mathcal{Y}_\zeta(x^*,\varepsilon)} \varepsilon \mathrm{d}F_{B_i^n}\left(y\right) \\
\le& \int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^n}\left(y\right) + \varepsilon.
\end{aligned} \tag{60}$$

Finally, by Lemma 6.1.2, for any $n \ge L_0\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, we have $\int_{\mathcal{Y}_\zeta(x^*,\varepsilon)} \mathrm{d}F_{B_i^n}\left(y\right) \le \varepsilon + \delta$. Combine this with (60), we conclude that (54) holds for any $n \ge L_0\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$. Take $L_1 := L_0$, we conclude the proof.

Proof of Lemma C.2

In this proof, we use the notations

$$\mathbb{E}h(X, y) := \int_0^\infty h(x, y)\mathrm{d}F_X(x) \;\text{ and }\; \mathbb{E}h(x, Y) := \int_0^\infty h(x, y)\mathrm{d}F_Y(y),$$

where X, Y are arbitrary non-negative random variables and h is any function.

Proof of (55): For any $i \in [n]$ and $x \geq 0$, we have

$$\left|\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^n}(y) - \int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^*}(y)\right|$$
$$\leq \left|\mathbb{E}\zeta_A(x,B_i^n) - \mathbb{E}\beta_A(x,B_i^n)\right| + \left|\mathbb{E}\beta_A(x,B_i^n) - \mathbb{E}\beta_A(x,B_i^*)\right| + \left|\mathbb{E}\beta_A(x,B_i^*) - \mathbb{E}\zeta_A(x,B_i^*)\right|. \quad (61)$$

We notice that upper-bounds of the first and third terms in the right-hand-side of (61) are given by (54) and (53) from Lemma C.1. We focus on finding an upper-bound of the second term of (61); to do this, we rewrite this term as follows.

$$\mathbb{E}\beta_A(x,B_i^n) = \int_{y<x} \mathrm{d}F_{B_i^n}(y) + \alpha\mathbb{P}(B_i^n = x) = F_{B_i^n}(x) - (1-\alpha)\mathbb{P}(B_i^n = x), \quad (62)$$

$$\text{and } \mathbb{E}\beta_A(x,B_i^*) = \int_{y<x} \mathrm{d}F_{B_i^*}(y) + \alpha\mathbb{P}(B_i^* = x) = F_{B_i^*}(x) - (1-\alpha)\mathbb{P}(B_i^* = x). \quad (63)$$

If $\alpha = 1$, we trivially have $\left|\mathbb{E}\beta_A(x,B_i^n) - \mathbb{E}\beta_A(x,B_i^*)\right| = \left|F_{B_i^n}(x) - F_{B_i^*}(x)\right|$. In the following, we assume that $\alpha < 1$ and consider three cases:

Case 1: If $x = 0$. From Lemma A.3-(i), we have $\mathbb{P}(B_i^n = 0) = \mathbb{P}(B_i^* = 0)$ and thus

$$\left|\mathbb{E}\beta_A(0,B_i^n) - \mathbb{E}\beta_A(0,B_i^*)\right| = \left|\int_{y<0} \mathrm{d}F_{B_i^n}(y) - \int_{y<0} \mathrm{d}F_{B_i^*}(y) + \alpha\mathbb{P}(B_i^n = 0) - \alpha\mathbb{P}(B_i^* = 0)\right| = 0.$$

Case 2: If $x > 0$, $\mathbb{P}(B_i^* = x) = 0$ by definition. On the other hand, from Results (ii) and (iii) of Lemma A.3, we have $\mathbb{P}(B_i^n = x) \leq D^{n-1}$ where we define $D := \left(1 - \frac{\underline{\lambda}}{\overline{\lambda}}\frac{\underline{w}^2}{\overline{w}^2}\right)$. Following (25), for any $n \geq C_0 \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$ (here, C_0 is defined as in Section A.4), we have $D^{n-1} \leq \frac{\varepsilon}{2(1-\alpha)}$. Therefore, for any $n \geq C_0 \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have

$$\begin{aligned}
&\left|\mathbb{E}\beta_A(x,B_i^n) - \mathbb{E}\beta_A(x,B_i^*)\right| \\
\leq &|F_{B_i^n}(x) - F_{B_i^*}(x)| + (1-\alpha)\left|\mathbb{P}(B_i^n = x)\right| \qquad \text{(due to (62) – (63))} \\
\leq &\sup_{x\in[0,\infty)} |F_{B_i^n}(x) - F_{B_i^*}(x)| + (1-\alpha)\frac{\varepsilon}{2(1-\alpha)} \\
= &\sup_{x\in[0,\infty)} |F_{B_i^n}(x) - F_{B_i^*}(x)| + \frac{\varepsilon}{2}.
\end{aligned}$$

In conclusion, $\left|\mathbb{E}\beta_A(x,B_i^n) - \mathbb{E}\beta_A(x,B_i^*)\right| \leq \sup_{x\in[0,\infty)} |F_{B_i^n}(x) - F_{B_i^*}(x)| + \varepsilon/2$ for any $x \geq 0$, $\alpha \in [0,1]$ and $n \geq C_0 \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$. Now, define $C_1' = C_1 \cdot 4(\ln(2)+1)$ (where C_1 is indicated in Lemma A.6); take $\varepsilon_1 := \varepsilon/2$, we have $C_1'\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right) \geq C_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$. Therefore, for any $n \geq C_1'\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$, we have $n \geq C_1\varepsilon_1^{-2}\ln\left(\frac{1}{\min\{\varepsilon_1,1/e\}}\right)$ and apply Lemma A.6, we have $\sup_{x\in[0,\infty)} |F_{B_i^n}(x) - F_{B_i^*}(x)| \leq \varepsilon_1 = \varepsilon/2$.

We deduce that for any $x \geq 0$, for any $n \geq \max\{C_0, C_1'\}\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $i \in [n]$, we have:

$$\left|\mathbb{E}\beta_A(x, B_i^n) - \mathbb{E}\beta_A(x, B_i^*)\right| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{64}$$

Finally, apply Lemma C.1 to (61) to bounds the first and third term of its right-hand-side, use (64) to bound its second-term and take $L_{(55)} = \max\{L_1, C_0, C_1'\}$, we deduce that for any $n \geq L_{(55)}\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$,

$$\left|\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^n}(y) - \int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^*}(y)\right| \leq (\delta + 2\varepsilon) + \varepsilon + (\delta + \varepsilon) = 2\delta + 4\varepsilon.$$

Proof of (56): To prove (56), we note that similar to the proof of (53) in Lemma C.1 (by replacing $F_{B_i^*}$ by $F_{A_i^*}$ and replacing $\zeta_A(x^*, y), \beta_A(x^*, y)$ by $\zeta_A(x, y^*), \beta_A(x, y^*)$), we can prove that for any n, $i \in [n]$, $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and $y^* \in [0, 2X_B]$, the following inequality holds

$$\left|\int_0^\infty \zeta_A(x,y^*)\,\mathrm{d}F_{A_i^*}(x) - \int_0^\infty \beta_A(x,y^*)\,\mathrm{d}F_{A_i^*}(x)\right| \leq \delta + \varepsilon. \tag{65}$$

Using this, we have

$$\begin{aligned}
&\left|\int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{A_i^*}(x)\mathrm{d}F_{B_i^*}(y) - \int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{A_i^*}(x)\mathrm{d}F_{B_i^n}(y)\right| \\
\leq &\int_0^\infty \left|\int_0^\infty \zeta_A(x,y)\mathrm{d}F_{A_i^*}(x) - \int_0^\infty \beta_A(x,y)\mathrm{d}F_{A_i^*}(x)\right| \mathrm{d}F_{B_i^*}(y) \\
&+ \left|\int_0^\infty\int_0^\infty \beta_A(x,y)\mathrm{d}F_{A_i^*}(x)\mathrm{d}F_{B_i^*}(y) - \int_0^\infty\int_0^\infty \beta_A(x,y)\mathrm{d}F_{A_i^*}(x)\mathrm{d}F_{B_i^n}(y)\right| \\
&+ \int_0^\infty \left|\int_0^\infty \beta_A(x,y)\mathrm{d}F_{A_i^*}(x) - \int_0^\infty \zeta_A(x,y)\mathrm{d}F_{A_i^*}(x)\right| \mathrm{d}F_{B_i^n}(y) \\
\leq &\int_0^\infty (\delta+\varepsilon)\mathrm{d}F_{B_i^*}(y) + \left|\int_0^\infty \mathbb{E}\beta_A(x, B_i^*)\mathrm{d}F_{A_i^*}(x) - \int_0^\infty \mathbb{E}\beta_A(x, B_i^n)\mathrm{d}F_{A_i^*}(x)\right| \\
&+ \int_0^\infty (\delta+\varepsilon)\mathrm{d}F_{B_i^n}(y) \\
\leq &2\delta + 2\varepsilon + \int_0^\infty \left|\mathbb{E}\beta_A(x, B_i^n) - \mathbb{E}\beta_A(x, B_i^*)\right| \mathrm{d}F_{A_i^*}(x).
\end{aligned}$$

Finally, take $L_{(56)} = \max\{C_0, C_1'\}$ and apply (64), we deduce that (56) holds for any $n \geq L_{(56)}\varepsilon^{-2}\ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$.

Proof of (57) Note that similar to the proof of (54) in Lemma C.1 (by replacing $F_{B_i^n}$ by $F_{A_i^n}$ and replacing $\zeta_A(x^*, y), \beta_A(x^*, y)$ by $\zeta_A(x, y^*), \beta_A(x, y^*)$), we can prove that for $n \geq L_1 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, $i \in [n]$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$,

$$\left| \int_0^\infty \zeta_A(x, y^*) \, \mathrm{d}F_{A_i^n}(x) - \int_0^\infty \beta_A(x, y^*) \, \mathrm{d}F_{A_i^n}(x) \right| \leq \delta + 2\varepsilon. \tag{66}$$

Now, as in the proof leading to (64), we can prove that the following inequality holds for any $n \geq \max\{C_0, C_1'\} \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, $i \in [n]$ and $y \geq 0$

$$\left| \mathbb{E}\beta_A(A_i^*, y) - \mathbb{E}\beta_A(A_i^n, y) \right| \leq \varepsilon. \tag{67}$$

Finally, take $L_{(57)} = \max\{L_1, C_0, C_1'\}$, for any $n \geq L_{(57)} \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, $i \in [n]$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, we have

$$\begin{aligned}
&\left| \int_0^\infty \int_0^\infty \zeta_A(x, y) \, \mathrm{d}F_{B_i^n}(y) \mathrm{d}F_{A_i^*}(x) - \int_0^\infty \int_0^\infty \zeta_A(x, y) \, \mathrm{d}F_{B_i^n}(y) \mathrm{d}F_{A_i^n}(x) \right| \\
\leq & \int_0^\infty \left| \int_0^\infty \zeta_A(x, y) \mathrm{d}F_{A_i^*}(x) - \int_0^\infty \beta_A(x, y) \mathrm{d}F_{A_i^*}(x) \right| \mathrm{d}F_{B_i^n}(y) \\
& \quad + \left| \int_0^\infty \int_0^\infty \beta_A(x, y) \mathrm{d}F_{A_i^*}(x) \mathrm{d}F_{B_i^n}(y) - \int_0^\infty \int_0^\infty \beta_A(x, y) \mathrm{d}F_{A_i^n}(x) \mathrm{d}F_{B_i^n}(y) \right| \\
& \quad + \int_0^\infty \left| \int_0^\infty \beta_A(x, y) \mathrm{d}F_{A_i^n}(x) - \int_0^\infty \zeta_A(x, y) \mathrm{d}F_{A_i^n}(x) \right| \mathrm{d}F_{B_i^n}(y) \\
\leq & \int_0^\infty (\delta + \varepsilon) \mathrm{d}F_{A_i^*}(x) + \int_0^\infty \left| \mathbb{E}\beta_A(A_i^*, y) - \mathbb{E}\beta_A(A_i^n, y) \right| \mathrm{d}F_{B_i^n}(y) + \int_0^\infty (\delta + 2\varepsilon) \mathrm{d}F_{A_i^n}(x) \\
\leq & 2\delta + 4\varepsilon.
\end{aligned}$$

Here, the second-to-last inequality comes from (65) and (66); and the last inequality is due to (67). In conclusion, take $L_2 := \max\{L_{(55)}, L_{(56)}, L_{(57)}\}$, we conclude the proof of this lemma.

Remark on the Lottery Blotto games with continuous CSFs

In this section, we present and prove the remark stating that under the additional assumption that the CSFs ζ_A and ζ_B are Lipschitz continuous on $[0, 2X_B] \times [0, 2X_B]$, the statements in Theorem 6.1.3 also hold with (68) and (69) (see below) in places of (6.4) and (6.5). For the sake of completeness, we formally state this result as follows.

Remark C.3. *For any CSF ζ_A and ζ_B that are Lipschitz continuous on $[0, 2X_B] \times [0, 2X_B]$, the following results hold (here, we denote $\zeta := (\zeta_A, \zeta_B)$):*

(i) In any game $\mathcal{LB}_n(\zeta)$, there exists a positive number $\varepsilon \leq \tilde{O}(n^{-1/2})$ such that for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$, the following inequalities hold for any pure strategy $\boldsymbol{x}^A$ and $\boldsymbol{x}^B$ of players A and B:

$$\Pi_\zeta^A(\boldsymbol{x}^A, \mathrm{IU}_B^{\gamma^*}) \leq \Pi_\zeta^A(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (2\delta + 5\varepsilon)\, W^A, \tag{68}$$

$$\Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \boldsymbol{x}^B) \leq \Pi_\zeta^B(\mathrm{IU}_A^{\gamma^*}, \mathrm{IU}_B^{\gamma^*}) + (2\delta + 5\varepsilon)\, W^B. \tag{69}$$

(ii) For any $\varepsilon \in (0,1]$, there exists a constant $L_\zeta > 0$ (that depends on ζ but does not depend on ε) such that in any game $\mathcal{LB}_n(\zeta)$ where $n \geq L_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, (68) and (69) hold for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$, $\delta \in \Delta_{\gamma^*}(\zeta, \varepsilon)$ and any pure strategy $\boldsymbol{x}^A, \boldsymbol{x}^B$ of players A and B.

Proof. We define the Lipschitz constant of ζ_A, ζ_B respectively by $\mathcal{L}_{\zeta_A}, \mathcal{L}_{\zeta_B}$ and let $\mathcal{L}_\zeta := \max\{\mathcal{L}_{\zeta_A}, \mathcal{L}_{\zeta_B}\}$. We focus on proving Result (ii) of this Remark; Result (i) can be deduced from Result (ii) and thus is omitted.

Step 1: We prove that for any $x^*, y^* \in [0, 2X_B]$, there exists a constant C_ζ (that does not depend on ε nor x^*, y^*) such that for any $n \geq C_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, the following inequalities hold:

$$\left| \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^n}\left(x\right) - \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^*}\left(x\right) \right| \leq \varepsilon, \tag{70}$$

$$\left| \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^n}\left(y\right) - \int_0^\infty \zeta_A\left(x^*, y\right) \mathrm{d}F_{B_i^*}\left(y\right) \right| \leq \varepsilon. \tag{71}$$

The proof of this statement is quite similar to the proof of Lemma A.7 (see Section A.7). We present here the proof of (70); the proof of (71) can be done similarly.

Fix $y^* \in [0, 2X_B]$; we define $f(x) := \zeta_A(x, y^*)$ and $\tilde{\varepsilon}_1 := \frac{\varepsilon}{4 + 4X_B\mathcal{L}_\zeta}$. From Lemma A.1, $F_{A_i^n}(x) = F_{A_i^*}(x) = 1, \forall x > 2X_B$; therefore, the left-hand-side of (70) can be rewritten as follows.

$$\left| \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^n}\left(x\right) - \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^*}\left(x\right) \right| = \left| \int_0^{2X_B} f(x) \mathrm{d}F_{A_i^n}\left(x\right) - \int_0^{2X_B} f(x) \mathrm{d}F_{A_i^*}\left(x\right) \right|. \tag{72}$$

Let us define $K := \left\lceil \frac{2X_B\mathcal{L}_\zeta}{\tilde{\varepsilon}_1} \right\rceil$ and $K+1$ points x_j such that $x_0 := 0$ and $x_j := x_{j-1} + \frac{2X_B}{K}$, $\forall j \in [K]$. In other words, we have the partitions $[0, 2X_B] = \bigcup_{j=1}^K P_j$ where we denote by P_1 the interval $[x_0, x_1]$ and by P_j the interval $(x_{j-1}, x_j]$ for $j = 2, \ldots, K$. For any $x, x' \in P_j, \forall j \in [K]$, since f is Lipschitz continuous, we have

$$|f(x) - f(x')| \leq \mathcal{L}_\zeta |x - x'| \leq \mathcal{L}_\zeta \frac{2X_B}{K} \leq \tilde{\varepsilon}_1. \tag{73}$$

Now, we define the function $g(x) := \sum_{j=1}^{K} f(x_j)\mathbf{1}_{P_j}(x)$. Here, $\mathbf{1}_{P_j}$ is the indicator function of the set P_j. From this definition and Inequality (73), we have $|f(x) - g(x)| \leq \tilde{\varepsilon}_1$, $\forall x \in [0, 2X_B]$. Therefore,

$$\left|\int_0^{2X_B} f(x)\mathrm{d}F_{A_i^n}(x) - \int_0^{2X_B} g(x)\mathrm{d}F_{A_i^n}(x)\right| \leq \int_0^{2X_B} \tilde{\varepsilon}_1 \mathrm{d}F_{A_i^n}(x) \leq \tilde{\varepsilon}_1, \tag{74}$$

$$\left|\int_0^{2X_B} f(x)\mathrm{d}F_{A_i^*}(x) - \int_0^{2X_B} g(x)\mathrm{d}F_{A_i^*}(x)\right| \leq \int_0^{2X_B} \tilde{\varepsilon}_1 \mathrm{d}F_{A_i^*}(x) \leq \tilde{\varepsilon}_1. \tag{75}$$

Now, we note that for any $j \in [K]$, $f(x_j) = \sum_{m=0}^{j} \left[f(x_m) - f(x_{m-1})\right]$; here, by convention, we denote by x_{-1} an arbitrary negative number and set $f(x_{-1}) = 0$. Using this, we have:

$$\begin{aligned}
&\left|\int_0^{2X_B} g(x)\mathrm{d}F_{A_i^n}(x) - \int_0^{2X_B} g(x)\mathrm{d}F_{A_i^*}(x)\right| \\
&= \left|\sum_{j=1}^{K} f(x_j)\left[\int_0^{2X_B} \mathbf{1}_{P_j}(x)\mathrm{d}F_{A_i^n}(x) - \int_0^{2X_B} \mathbf{1}_{P_j}(x)\,\mathrm{d}F_{A_i^*}(x)\right]\right| \\
&= \left|\sum_{j=1}^{K} f(x_j)\left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&= \left|\sum_{j=1}^{K}\left(\sum_{m=0}^{j}\left[f(x_m) - f(x_{m-1})\right]\left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right)\right| \\
&\leq \left|\left[f(x_0) - f(x_{-1})\right]\sum_{j=1}^{K}\left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&\qquad + \left|\sum_{m=1}^{K}\left(\left[f(x_m) - f(x_{m-1})\right]\sum_{j=m}^{K}\left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right)\right|.
\end{aligned} \tag{76}$$

Note that $\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right) = F_{A_i^n}(x_j) - F_{A_i^n}(x_{j-1}) - F_{A_i^*}(x_j) + F_{A_i^*}(x_{j-1})$.[27] Now, we can rewrite the first term in (76) as follows.

$$\begin{aligned}
&\left|\left[f(x_0) - f(x_{-1})\right]\sum_{j=1}^{K}\left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right)\right]\right| \\
&= \left|f(0) \cdot \left[\sum_{j=1}^{K}\left(F_{A_i^n}(x_j) - F_{A_i^n}(x_{j-1}) - F_{A_i^*}(x_j) + F_{A_i^*}(x_{j-1})\right)\right]\right| \\
&= \left|f(0) \cdot \left[F_{A_i^n}(x_K) - F_{A_i^n}(x_0) - F_{A_i^*}(x_K) + F_{A_i^*}(x_0)\right]\right|
\end{aligned}$$

[27] For any $j \geq 2$, this is trivially since $P_j := (x_{j-1}, x_j]$. For $P_1 = [0, x_1]$, we have that $\mathbb{P}\left(A_i^n \in P_1\right) - \mathbb{P}\left(A_i^* \in P_1\right) = \mathbb{P}(A_i^n \in (0, x_1]) - \mathbb{P}(A_i^* \in (0, x_1]) + \mathbb{P}(A_i^n = 0) - \mathbb{P}(A_i^* = 0)$; moreover, due to Lemma A.3-(i), we also note that $\mathbb{P}(A_i^n = 0) = \mathbb{P}(A_i^* = 0)$.

$$\leq 2 \sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right|. \tag{77}$$

Here, the last inequality comes from the fact that $f(x) \leq 1, \forall x \in [0, 2X_B]$ (since it is a CSF).

Now, we recall that for any $m \in [K]$, $f(x_m) - f(x_{m-1}) \leq \frac{2X_B \mathcal{L}_\zeta}{K}$. Therefore, the second term in (76) is

$$\begin{aligned}
&\left| \sum_{m=1}^{K} \left([f(x_m) - f(x_{m-1})] \sum_{j=m}^{K} \left[\mathbb{P}\left(A_i^n \in P_j\right) - \mathbb{P}\left(A_i^* \in P_j\right) \right] \right) \right| \\
=& \left| \sum_{m=1}^{K} \left([f(x_m) - f(x_{m-1})] \sum_{j=m}^{K} \left[F_{A_i^n}(x_j) - F_{A_i^n}(x_{j-1}) - F_{A_i^*}(x_j) + F_{A_i^*}(x_{j-1}) \right] \right) \right| \\
=& \left| \sum_{m=1}^{K} \left([f(x_m) - f(x_{m-1})] \left[F_{A_i^n}(x_K) - F_{A_i^n}(x_{m-1}) - F_{A_i^*}(x_K) + F_{A_i^*}(x_{m-1}) \right] \right) \right| \\
\leq& \sum_{m=1}^{K} \frac{2X_B \mathcal{L}_\zeta}{K} \cdot 2 \sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right| \\
=& 4X_B \mathcal{L}_\zeta \sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right|.
\end{aligned} \tag{78}$$

Inject (77) and (78) into (76), we obtain that

$$\left| \int_0^{2X_B} g(x) \mathrm{d}F_{A_i^n}(x) - \int_0^{2X_B} g(x) \mathrm{d}F_{A_i^*}(x) \right| \leq (2 + 4X_B \mathcal{L}_\zeta) \sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right|. \tag{79}$$

Apply the triangle inequality and combine (74), (75), (79), we have that:

$$\left| \int_0^{2X_B} f(x) \mathrm{d}F_{A_i^n}(x) - \int_0^{2X_B} f(x) \mathrm{d}F_{A_i^*}(x) \right| \leq 2\tilde{\varepsilon}_1 + (2 + 4X_B \mathcal{L}_\zeta) \sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right|. \tag{80}$$

Now, let us define $C_\zeta := C_1 \cdot (4 + 4X_B \mathcal{L}_\zeta)^2 \left[\ln(4 + 4X_B \mathcal{L}_\zeta) + 1\right]$ (note that C_1 is defined in Lemma A.6) and we deduce that $C_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq C_1 \tilde{\varepsilon}_1^{-2} \ln\left(\frac{1}{\min\{\tilde{\varepsilon}_1, 1/e\}}\right)$.[28] Take $\varepsilon_1 := \tilde{\varepsilon}_1$, for any $n \geq C_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, we have $n \geq C_1 \varepsilon_1^{-2} \ln\left(\frac{1}{\min\{\varepsilon_1, 1/e\}}\right)$ and by applying Lemma A.6, we obtain that $\sup_{x\in[0,\infty)} \left| F_{A_i^n}(x) - F_{A_i^*}(x) \right| \leq \varepsilon_1 = \tilde{\varepsilon}_1$ and thus by (72) and (80), we have:

$$\left| \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^n}(x) - \int_0^\infty \zeta_A\left(x, y^*\right) \mathrm{d}F_{A_i^*}(x) \right|$$

[28] Apply Lemma A.2, $C_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) = C_1 \left(\frac{4+4X_B \mathcal{L}_\zeta}{\varepsilon}\right)^2 \left[\ln(4+4X_B \mathcal{L}_\zeta) + 1\right] \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right) \geq C_1 \left(\frac{4+4X_B \mathcal{L}_\zeta}{\varepsilon}\right)^2 \ln\left(\frac{4+4X_B \mathcal{L}_\zeta}{\varepsilon}\right)$. Moreover, since $\frac{\varepsilon}{4+4X_B \mathcal{L}_\zeta} = \min\left\{\frac{\varepsilon}{4+4X_B \mathcal{L}_\zeta}, \frac{1}{e}\right\} = \min\left\{\tilde{\varepsilon}_1, \frac{1}{e}\right\}$ (due to the fact that $\tilde{\varepsilon}_1 = \frac{\varepsilon}{4+4X_B \mathcal{L}_\zeta} < \frac{1}{e}$).

$$\leq 2\tilde{\varepsilon}_1 + (2 + 4X_B\mathcal{L}_\zeta)\,\tilde{\varepsilon}_1$$
$$=(4 + 4X_B\mathcal{L}_\zeta)\tilde{\varepsilon}_1 = \varepsilon.$$

This is exactly (70).

Step 2: Based on (70) and (71), we can trivially deduce that the following inequalities hold for any $n \geq C_\zeta \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $i \in [n]$:

$$\left|\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^n}(y) - \int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^*}(y)\right| \leq \varepsilon, \forall x \geq 0, \tag{81}$$

$$\left|\int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^*}(y)\,\mathrm{d}F_{A_i^*}(x) - \int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{B_i^n}(y)\,\mathrm{d}F_{A_i^*}(x)\right| \leq \varepsilon, \tag{82}$$

$$\left|\int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{A_i^*}(x)\,\mathrm{d}F_{B_i^n}(y) - \int_0^\infty\int_0^\infty \zeta_A(x,y)\,\mathrm{d}F_{A_i^n}(x)\,\mathrm{d}F_{B_i^n}(y)\right| \leq \varepsilon. \tag{83}$$

We notice that the left-hand-sides of these inequalities are exactly the terms considered in Lemma C.2; moreover, the upper-bounds given in(81), (82) and (83) are smaller than that in (55), (56) and (57) of Lemma C.2.

Step 3: To complete the proof of Remark C.3, we follow the proof of Theorem 6.1.3 where we use (81), (82) and (83) instead of (55), (56) and (57). By doing this, we obtain (68) and (69). □

Proof of Lemma 6.1.7

(i) We first consider the game $\mathcal{LB}_n(\mu^R)$.

Step 1: We prove that there exists $\delta_0 = O(\varepsilon^{-1/R}-1)$ such that $\mathcal{X}_{\mu^R}(y^*,\varepsilon) \subset [y^* - \delta_0, y^* + \delta_0]$ for any $y^* \in [0, 2X_B]$. Note that this is trivial if $\mathcal{X}_{\mu^R}(y^*,\varepsilon) = \emptyset$. In the following, we consider the case where $\mathcal{X}_{\mu^R}(y^*,\varepsilon) \neq \emptyset$. We denote by $f : [0,2X_B] \times [0,2X_B] \to [0,1]$ the function:

$$f(x,y^*) := |\mu_A^R(x,y^*) - \beta_A(x,y^*)| = \begin{cases} \frac{\alpha x^R}{\alpha x^R + (1-\alpha)(y^*)^R}, \text{ if } x < y^* \\ 0, \text{ if } x = y^* \\ 1 - \frac{\alpha x^R}{\alpha x^R + (1-\alpha)(y^*)^R}, \text{ if } x > y^* \end{cases}.$$

Trivially, $y^* \notin \mathcal{X}_{\mu^R}(y^*,\varepsilon)$. Take an arbitrary $x \in \mathcal{X}_{\mu^R}(y^*,\varepsilon)$. If $x < y^*$, we have

$$f(x,y^*) \geq \varepsilon \Rightarrow \frac{\alpha x^R}{\alpha x^R + (1-\alpha)(y^*)^R} \geq \varepsilon \Rightarrow \frac{x}{y^*} \geq \left(\frac{\varepsilon}{1-\varepsilon}\frac{1-\alpha}{\alpha}\right)^{1/R}.$$

Therefore, $0 < y^* - x \leq y^*\left[1 - \left(\frac{\varepsilon}{1-\varepsilon}\frac{1-\alpha}{\alpha}\right)^{1/R}\right]$. Here, we note that the right-hand side is positive (due to the condition $\varepsilon < \alpha$); moreover, it is upper-bounded by $O(1-\varepsilon^{1/R})$, thus bounded by $O(\varepsilon^{-1/R}-1)$.

On the other hand, if $x > y^*$, we have:

$$f(x, y^*) \geq \varepsilon \Rightarrow 1 - \frac{\alpha x^R}{\alpha x^R + (1-\alpha)(y^*)^R} \geq \varepsilon \Rightarrow \frac{x}{y^*} \leq \left(\frac{1-\varepsilon}{\varepsilon}\frac{1-\alpha}{\alpha}\right)^{1/R}.$$

Therefore we have $0 < x - y^* \leq y^* \left[\left(\frac{1-\varepsilon}{\varepsilon}\frac{1-\alpha}{\alpha}\right)^{1/R} - 1\right]$. Here the right-hand side is positive (due to the condition $\alpha + \varepsilon < 1$) and is upper-bounded by $O(\varepsilon^{-1/R} - 1)$.

In conclusion, for any $\varepsilon < \min\{\alpha, 1-\alpha\}$, there exists $\delta_0 = O(\varepsilon^{-1/R} - 1)$ such that $\mathcal{X}_{\mu^R}(y^*, \varepsilon) \subset [y^* - \delta_0, y^* + \delta_0]$. Note that a similar proof can be done to prove that there exists $\hat{\delta}_0 = O(\varepsilon^{-1/R} - 1)$ such that for any $x^* \in [0, 2X_B]$, $\mathcal{Y}_{\mu^R}(x^*, \varepsilon) \subset [x^* - \hat{\delta}_0, x^* + \hat{\delta}_0]$.

Step 2: For any $y^* \in [0, 2X_B]$ and $\delta_0 \geq 0$, let us define the set

$$I_0(y^*) := [y^* - \delta_0, y^* + \delta_0] \bigcap [0, 2X_B];$$

we want to show that $\int_{x \in I_0(y^*)} \mathrm{d}F_{A_i^*}(x) \leq \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}, \forall i \in [n]$.

Case 1: For $i \in \Omega_A(\gamma^*)$, then $A_i^* A_{\gamma^*,i}^S$, we have that

$$\begin{aligned}
\int_{x \in I_0(y^*)} \mathrm{d}F_{A_i^*}(x) \leq & F_{A_{\gamma^*,i}^S}\left(y^* + \delta_0\right) - F_{A_{\gamma^*,i}^S}\left(y^* - \delta_0\right) \\
= & \begin{cases}
\frac{(y^*+\delta_0)\lambda_B^*}{v_i^B} \leq \frac{2\delta_0\lambda_B^*}{v_i^B}, \text{ if } 0 \leq y^* \leq \delta_0 \\
\frac{(y^*+\delta_0)\lambda_B^*}{v_i^B} - \frac{(y^*-\delta_0)\lambda_B^*}{v_i^B} = \frac{2\delta_0\lambda_B^*}{v_i^B}, \text{ if } \delta_0 \leq y^* < \frac{v_i^B}{\lambda_B^*} - \delta_0 \\
1 - \frac{(y^*-\delta_0)\lambda_B^*}{v_i^B} = \frac{v_i^B - y^*\lambda_B^* + \delta_0\lambda_B^*}{v_i^B} \leq \frac{2\delta_0\lambda_B^*}{v_i^B}, \text{ if } \frac{v_i^B}{\lambda_B^*} - \delta_0 \leq y^* \leq \frac{v_i^B}{\lambda_B^*} + \delta_0 \\
1 - 1 = 0, \text{ otherwise}
\end{cases} \\
\leq & \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}.
\end{aligned}$$

Case 2: For $i \notin \Omega_A(\gamma^*)$, then $A_i^* = A_{\gamma^*,i}^W$. We have

$$\begin{aligned}
& \int_{x \in I_0(y^*)} \mathrm{d}F_{A_i^*}(x) \\
\leq & F_{A_{\gamma^*,i}^W}\left(y^* + \delta_0\right) - F_{A_{\gamma^*,i}^W}\left(y^* - \delta_0\right) \\
= & \begin{cases}
\frac{(y^*+\delta_0)\lambda_B^*}{v_i^B} \leq \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}, \text{ if } 0 \leq y^* \leq \delta_0 \\
\frac{(y^*+\delta_0)\lambda_B^*}{v_i^B} - \frac{(y^*-\delta_0)\lambda_B^*}{v_i^B} = \frac{2\delta_0\lambda_B^*}{v_i^B}, \text{ if } \delta_0 < y^* < \frac{v_i^A}{\lambda_A^*} - \delta_0 \\
1 - \frac{\frac{v_i^B}{\lambda_B^*} - \frac{v_i^A}{\lambda_A^*}}{\frac{v_i^B}{\lambda_B^*}} - \frac{(y^*-\delta_0)\lambda_B^*}{v_i^B} = \frac{v_i^A\frac{\lambda_B^*}{\lambda_A^*} - y^*\lambda_B^* + \delta_0\lambda_B^*}{v_i^B} \leq \frac{2\delta_0\lambda_B^*}{v_i^B}, \text{ if } \frac{v_i^A}{\lambda_A^*} - \delta_0 \leq y^* \leq \frac{v_i^A}{\lambda_A^*} + \delta_0 \\
1 - 1 = 0, \text{ otherwise}
\end{cases} \\
\leq & \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}.
\end{aligned}$$

Note that we can similarly prove that for any $x^* \in [0, 2X_B]$ and $\delta_0 \geq 0$ and $i \in [n]$, we also have $\int_{y \in I_0(x^*)} \mathrm{d}F_{B_i^*}(y) \leq \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}$.

Step 3: Conclusion. We note that all random variable $A_i^*, B_i^*, i \in [n]$ are bounded in $[0, 2X_B]$; therefore, for any $x^*, y^* \in [0, 2X_B]$ and $\delta_0 \geq 0$, we have:

$$\int_{x \in [y^*-\delta, y^*+\delta_0]} \mathrm{d}F_{A_i^*}(x) = \int_{x \in I_0(y^*)} \mathrm{d}F_{A_i^*}(x) \text{ and } \int_{y \in [x^*-\delta, x^*+\delta_0]} \mathrm{d}F_{B_i^*}(y) = \int_{y \in I_0(x^*)} \mathrm{d}F_{B_i^*}(x).$$

Let us define $\delta_\mu := min\{1, \frac{2n\bar{\lambda}\delta_0\bar{w}}{\underline{w}}\} = O\left(n(\varepsilon^{-1/R} - 1)\right)$ and we conclude that:

$$\max\left\{\max_{y^* \in [0,2X_B]} \int_{\mathcal{X}_{\mu^R}(y^*,\varepsilon)} \mathrm{d}F_{A_i^*}(x), \max_{x^* \in [0,2X_B]} \int_{\mathcal{Y}_{\mu^R}(x^*,\varepsilon)} \mathrm{d}F_{B_i^*}(y)\right\} \leq \delta_\mu.$$

This implies that $\delta_\mu \in \Delta_{\gamma^*}(\mu^R, \varepsilon)$.

(ii) **We now turn our focus on the game $\mathcal{LB}_n(\nu^R)$.** We first prove the existence of $\delta_1 > 0$ such that $\mathcal{X}_{\nu^R}(y^*, \varepsilon) \subset [y^* - \delta_1, y^* + \delta_1]$ for any $y^* \in [0, 2X_B]$. As in step 1 in the above analysis of the game $\mathcal{LB}_n(\mu^R)$, we denote by $g : [0, 2X_B] \times [0, 2X_B] \to [0, 1]$ the function:

$$g(x, y^*) := |\nu_A^R(x, y^*) - \beta_A(x, y^*)| = \begin{cases} \frac{\alpha e^{xR}}{\alpha e^{xR} + (1-\alpha)e^{y^*R}}, & \text{if } x < y^*, \\ 0, & \text{if } x = y^*, \\ 1 - \frac{\alpha e^{xR}}{\alpha e^{xR} + (1-\alpha)e^{y^*R}} & , \text{if } x > y^*. \end{cases}$$

Trivially, $y^* \notin \mathcal{X}_{\nu^R}(y^*, \varepsilon)$. Take an arbitrary $x \in \mathcal{X}_{\mu^R}(y^*, \varepsilon)$. If $x < y^*$, we have

$$g(x, y^*) \geq \varepsilon \Rightarrow \frac{\alpha e^{xR}}{\alpha e^{xR} + (1-\alpha)e^{y^*R}} \geq \varepsilon.$$

Therefore, $0 < y^* - x \leq \frac{1}{R} \ln\left(\frac{1-\varepsilon}{\varepsilon} \frac{\alpha}{1-\alpha}\right)$. Here, we note that the right-hand side is positive (due to the condition $\varepsilon < \alpha$).

On the other hand, if $x > y^*$, we have:

$$g(x, y^*) \geq \varepsilon \Rightarrow 1 - \frac{\alpha e^{xR}}{\alpha e^{xR} + (1-\alpha)e^{y^*R}} \geq \varepsilon.$$

Therefore, $0 < x - y^* \leq \frac{1}{R} \ln\left(\frac{1-\varepsilon}{\varepsilon} \frac{1-\alpha}{\alpha}\right)$. Here, the right-hand side is positive (due to the condition $\alpha + \varepsilon < 1$).

In conclusion, we have proved that $\mathcal{X}_{\nu^R}(y^*, \varepsilon) \subset [y^* - \delta_1, y^* + \delta_1]$ for any $y^* \in [0, 2X_B]$ and $\delta_1 = O(R^{-1}\ln(\varepsilon^{-1}))$. Now, we define $I_1(y^*) := [y^* - \delta_1, y^* + \delta_1] \bigcap [0, 2X_B]$. Similar to step 2 of the above analysis regarding the game $\mathcal{LB}_n(\mu^R)$, we can prove that $\int_{I_1(y^*)} \mathrm{d}F_{A_i^*}(x) \leq 2n\bar{\lambda}\delta_1\bar{w}/\underline{w}$ for any $y^* \in [0, 2X_B]$. Therefore,

$$\max\left\{\max_{y^* \in [0,2X_B]} \int_{\mathcal{X}_{\nu^R}(y^*,\varepsilon)} \mathrm{d}F_{A_i^*}(x), \max_{x^* \in [0,2X_B]} \int_{\mathcal{Y}_{\nu^R}(x^*,\varepsilon)} \mathrm{d}F_{B_i^*}(y)\right\} \leq \delta_\nu,$$

where $\delta_\nu := \min\{1, \frac{2n\bar{\lambda}\delta_1\bar{w}}{\underline{w}}\} = O\left(nR^{-1}\ln(\varepsilon^{-1})\right)$ and $\delta_\nu \in \Delta_{\gamma^*}(\nu^R, \varepsilon)$.

Proof of Theorem 6.1.8

Theorem 6.1.8. ***(Approximate equilibria of the ratio-form Lottery Blotto games)*** *For any $\bar{\varepsilon} > 0$ and $\alpha \in (0,1)$ such that $\bar{\varepsilon} < \min\{\alpha, 1-\alpha\}$, there exists $\tilde{L} > 0$ such that for any $n \geq \tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right)$, $R \geq O\left(\frac{n}{\bar{\varepsilon}} \ln\left(\frac{1}{\bar{\varepsilon}}\right)\right)$ and $\gamma^* \in \mathcal{S}_n^{(4.5)}$, the IU^{γ^*} strategy is an $\bar{\varepsilon}W$-equilibrium of any game $\mathcal{LB}_n(\mu^R)$ and $\mathcal{LB}_n(\nu^R)$ having α as the tie-breaking-rule parameter.*

Proof. Take $\varepsilon = \bar{\varepsilon}/21$ and $\tilde{L} = L^* 21^2(\ln(21) + 1)$ (where L^* is indicated in Theorem 6.1.3). We note that $\tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right) \geq L^* \varepsilon \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$;[29] therefore, for any $n \geq \tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right)$, we have

$$n \geq L^* \varepsilon \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right).$$

Therefore, by applying Result (ii)-Theorem 6.1.3, for any $R > 0$, we condlude that the IU^{γ^*} strategy is an $(8\delta_\mu + 13\varepsilon)W$-equilibrium of the game $\mathcal{LB}_n(\mu^R)$. Similarly, the IU^{γ^*} strategy is an $(8\delta_\nu + 13\varepsilon)W$-equilibrium of the game $\mathcal{LB}_n(\nu^R)$).

We first consider the game $\mathcal{LB}_n(\mu^R)$. Apply Lemma 6.1.7, we have $\delta_\mu \leq \varepsilon$, for any $R \geq O\left(\ln\left(\frac{1}{\varepsilon}\right) \ln\left(\frac{\varepsilon}{n} + 1\right)\right) = O\left(\ln\left(\frac{1}{\varepsilon}\right) \frac{n}{\varepsilon}\right)$ and $\gamma^* \in \mathcal{S}_n^{(4.5)}$. Therefore, for any $n \geq \tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right)$, $R \geq O\left(\frac{n}{\bar{\varepsilon}} \ln\left(\frac{1}{\bar{\varepsilon}}\right)\right)$, the IU^{γ^*} strategy is an $21\varepsilon W$-equilibrium (i.e., $\bar{\varepsilon}W$-equilibrium) of the game $LB(\mu^R)$.

Similarly, apply Lemma 6.1.7, for any $\gamma^* \in \mathcal{S}_n^{(4.5)}$ and $R \geq O\left(\frac{n}{\bar{\varepsilon}} \ln\left(\frac{1}{\bar{\varepsilon}}\right)\right)$, we have $\delta_\nu \leq \varepsilon$. Therefore, for any $n \geq \tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right)$, $R \geq O\left(\frac{n}{\bar{\varepsilon}} \ln\left(\frac{1}{\bar{\varepsilon}}\right)\right)$, the IU^{γ^*} strategy is an $21\varepsilon W$-equilibrium (i.e., $\bar{\varepsilon}W$-equilibrium) of the game $LB(\nu^R)$.

□

[29] Note that $\varepsilon = \bar{\varepsilon}/21$ and apply Lemma A.2 to have that $\tilde{L}\bar{\varepsilon}^{-2} \ln\left(\frac{1}{\min\{\bar{\varepsilon},1/e\}}\right) \geq L^* \left(\frac{21}{\bar{\varepsilon}}\right)^2 \ln\left(\frac{21}{\bar{\varepsilon}}\right)$; moreover, we recall that $\frac{\bar{\varepsilon}}{21} < \frac{1}{e}$; therefore, $\ln\left(\frac{21}{\bar{\varepsilon}}\right) = \ln\left(\frac{1}{\min\{\bar{\varepsilon}/21,1/e\}}\right) = \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$.

Appendix D

Supplementary Materials for Section 6.2 on the $\mathcal{GR}-\mathcal{CB}_n^C$ Game

D.1 Proof of Results in Section 6.2.2

Proof of Theorem 6.2.5

Theorem 6.2.5. *In the F-APA game where $p \geq 0$, we have the following results:*

(i) If $qu^B - p \leq 0$, there exists a unique pure equilibrium where players' bids are $x^A = x^B = 0$ and their equilibrium payoffs are $\Pi^A_{\text{F-APA}} = u^A$ and $\Pi^B_{\text{F-APA}} = 0$ respectively.

(ii) If $0 < qu^B - p \leq u^A$, there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution $G_{A_2^+}$ (resp. $G_{B_2^+}$) defined as follows:

$$G_{A_2^+}(x) = \begin{cases} \frac{p}{qu^B} + \frac{x}{qu^B}, \forall x \in \left[0, qu^B - p\right], \\ 1 \quad , \forall x > qu^B - p, \end{cases} \tag{6.8}$$

$$\text{and } G_{B_2^+}(x) = \begin{cases} 1 - \frac{qu^B}{u^A} + \frac{p}{u^A}, \forall x \in \left[0, \frac{p}{q}\right), \\ 1 - \frac{qu^B}{u^A} + \frac{q \cdot x}{u^A}, \forall x \in \left[\frac{p}{q}, u^B\right], \\ 1 \quad , \forall x > u^B. \end{cases} \tag{6.9}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = u^A - qu^B + p$ and $\Pi^B_{\text{F-APA}} = 0$.

(iii) If $qu^B - p > u^A$, there exists no pure equilibrium; there is a unique mixed equilibrium where player A (resp. player B) draws her bid from the distribution $G_{A_3^+}$ (resp. $G_{B_3^+}$) defined as follows:

$$G_{A_3^+}(x) = \begin{cases} 1 - \frac{u^A}{qu^B} + \frac{x}{qu^B}, \forall x \in \left[0, u^A\right], \\ 1 \quad , \forall x > u^A, \end{cases} \tag{6.10}$$

$$\text{and } G_{B_3^+}(x) = \begin{cases} 0 \quad , \forall x \in \left[0, \frac{p}{q}\right), \\ -\frac{p}{u^A} + \frac{q \cdot x}{u^A}, \forall x \in \left[\frac{p}{q}, \frac{u^A + p}{q}\right], \\ 1 \quad , \forall x > \frac{u^A + p}{q}. \end{cases} \tag{6.11}$$

In this mixed equilibrium, players' payoffs are $\Pi^A_{\text{F-APA}} = 0$ and $\Pi^B_{\text{F-APA}} = u^B - (u^A + p)/q$.

Proof. **Proof of Result (i):** For any $x^B \geq u^B$ and any x^A, we have $\Pi^B_{\text{F-APA}}\left(x^A, x^B\right) < 0$. Moreover, due to the condition $qu^B - p \leq 0$, we have $x^A > qx^B - p$ for any $x^A \geq 0$ and

$0 \leq x^B < u^B$; that is, player B always loses if she bids strictly lower than u^B. Trivially, $x^B = 0$ is the unique dominant strategy of player B. Player A's best response against $x^B = 0$ is $x^A = 0$. In conclusion, we have:

$$\begin{aligned}
&\Pi^A_{\text{F-APA}}(0,0) = u^A \text{ and } \Pi^A_{\text{F-APA}}\left(x^A, 0\right) = u^A - x^A < u^A, \forall x^A > 0,\\
&\Pi^B_{\text{F-APA}}(0,0) = 0 \text{ and } \Pi^B_{\text{F-APA}}\left(0, x^B\right) < 0, \forall x^B > 0,
\end{aligned}$$

Proof of Result (ii) First, from $0 < qu^B - p \leq u^A$, we have $0 \leq p/q < u^B$. We prove (by contradiction) that there exists no pure equilibrium under this condition. Assume that the profile x^A, x^B is a pure equilibrium of the F-APA game. We consider two cases:

- Case 1: If $x^A = 0$, then player B's best response is to choose $x^B = p/q + \varepsilon$ with an infinitesimal $\varepsilon > 0$ since by doing it, she can guarantee to win (since $q(p/q + \varepsilon) - p = q\varepsilon > 0$) and gets the payoff $u^B - p/q - \varepsilon > 0$.[30] However, player A's best response against $x^B = p/q + \varepsilon$ *is not* $x^A = 0$.[31]
- Case 2: If $x^A > 0$, then player B's best response is either $x^B = (x^A + p)/q + \varepsilon$ if there exists $\varepsilon > 0$ small enough such that $qu^B - p - x^A - \varepsilon > 0$ or $x^B = 0$ if there is no such ε. However, $x^A > 0$ is not the best response of player A against neither $x^B = (x^A + p)/q + \varepsilon$ nor against $x^B = 0$.[32]

We conclude that x^A, x^B cannot the best response against each other; thus, there exists no pure equilibrium in this case.

Now, we prove that if player B plays according to $G_{B_2^+}$, player A has no incentive to deviate from playing according to $G_{A_2^+}$. Denote by A_2^+ and B_2^+ the random variables that correspond to $G_{A_2^+}$ and $G_{B_2^+}$, since $G_{A_2^+}$ is a continuous distribution on $\left(0, qu^B - p\right]$, we have:

$$\begin{aligned}
\Pi^A_{\text{F-APA}}\left(G_{A_2^+}, G_{B_2^+}\right) &= \left[u^A \mathbb{P}\left(B_2^+ < \frac{p}{q}\right) - 0\right] \mathbb{P}\left(A_2^+ = 0\right) + \left[\alpha u^A \mathbb{P}\left(B_2^+ = \frac{p}{q}\right) - 0\right] \mathbb{P}\left(A_2^+ = 0\right)\\
&\quad + \int_0^{qu^B - p} \left[u^A \mathbb{P}\left(B_2^+ < \frac{x+p}{q}\right) - x\right] \mathrm{d}G_{A_2^+}(x)\\
&= u^A G_{B_2^+}\left(\frac{p}{q}\right) \frac{p}{qu^B} + 0 + \int_0^{qu^B - p} \left[u^A G_{B_2^+}\left(\frac{x+p}{q}\right) - x\right] \mathrm{d}G_{A_2^+}(x)\\
&= \left(u^A - qu^B + p\right) \frac{p}{qu^B} + \int_0^{qu^B - p} \left(u^A - qu^B + p\right) \frac{1}{qu^B} \mathrm{d}x
\end{aligned} \tag{84}$$

[30] Note that if player B choose $x^B = 0$, she loses and hes payoff is only 0.

[31] Player A's best response against $x^B = p/q + \varepsilon$ is $x^A = q\varepsilon + \delta$ (with an infinitesimal $\delta > 0$ such that $u^A - q\varepsilon - \delta > 0$).

[32] The best response of player A against $x^B = (x^A + p)/q + \varepsilon$ is $x^A + q\varepsilon + \delta$ where $0 < \delta < u^A - x^A - q\varepsilon$ (δ exists thanks to the condition on ε and that $qu^B - p \leq u^A$) and her best response against $x^B = 0$ is $x^A = 0$.

$$= u^A - qu^B + p.$$

Here, (84) comes from the fact that $\mathbb{P}\left(B_2^+ = p/q\right) = 0$, due to definition. Now, if player A plays a pure strategy $x^A > qu^B - p$ while player B plays $G_{B_2^+}$, her payoff is:

$$\Pi^A_{\text{F-APA}}\left(x^A, G_{B_2^+}\right) \leq u^A - x^A < u^A - qu^B + p = \Pi^A_{\text{F-APA}}\left(G_{A_2^+}, G_{B_2^+}\right).$$

Moreover, for any pure strategy $x^A \in [0, qu^B - p]$, we have:

$$\begin{aligned}\Pi^A_{\text{F-APA}}\left(x^A, G_{B_2^+}\right) &= u^A \mathbb{P}\left(B_2^+ < \frac{x^A+p}{q}\right) + \alpha u^A \mathbb{P}\left(B_2^+ = \frac{x^A+p}{q}\right) - x^A \\ &\leq u^A G_{B_2^+}\left(\frac{x^A+p}{q}\right) - x^A = u^A\left[1 - \frac{qu^B}{u^A} + \frac{q}{u^A}\frac{(x^A+p)}{q}\right] - x^A \\ &= u^A - qu^B + p \\ &= \Pi^A_{\text{F-APA}}\left(G_{A_2^+}, G_{B_2^+}\right).\end{aligned}$$

In conclusion, $\Pi^A\left(G_{A_2^+}, G_{B_2^+}\right) \geq \Pi^A\left(x^A, G_{B_2^+}\right)$ for any $x^A \geq 0$.

Similarly, we prove that when player A plays $G_{A_2^+}$, player B has no incentive to deviate from $G_{B_2^+}$. Indeed, since $G_{B_2^+}$ is a continuous distribution on $[p/q, u^B]$, we have

$$\begin{aligned}\Pi^B_{\text{F-APA}}\left(G_{A_2^+}, G_{B_2^+}\right) &= \left[u^B \mathbb{P}\left(A_2^+ < 0\right) - \frac{p}{q}\right] \mathbb{P}\left(B_2^+ = \frac{p}{q}\right) \\ &\quad + \left[(1-\alpha)u^B \mathbb{P}\left(A_2^+ = 0\right) - \frac{p}{q}\right] \mathbb{P}\left(B_2^+ = \frac{p}{q}\right) \\ &\quad + \int_{p/q}^{u^B} \left[u^B \mathbb{P}\left(A_2^+ < qx - p\right) - x\right] \mathrm{d}G_{B_2^+}(x) \\ &= 0 + 0 + \int_{p/q}^{u^B} \left[u^B G_{A_2^+}\left(qx - p\right) - x\right] \mathrm{d}G_{B_2^+}(x) \qquad (85) \\ &= \int_{p/q}^{u^B} \left[u^B\left(\frac{p}{qu^B} + \frac{qx-p}{qu^B}\right) - x\right] \frac{q}{u^A}\mathrm{d}x \\ &= 0.\end{aligned}$$

Here, (85) comes from the fact $\mathbb{P}\left(B_2^+ = p/q\right) = 0$ and that $\mathbb{P}(A_2^+ = z) = 0$ for any $z \in (0, qu^B - p]$ due to definition. Now, as stated above, for any pure strategy $x^B > u^B$, trivially $\Pi^B_{\text{F-APA}}(G_{A_2^+}, x^B) < 0$. Moreover, $\Pi^B_{\text{F-APA}}\left(G_{A_2^+}, x^B\right) \leq u^B G_{A_2^+}\left(qx^B - p\right) - x^B = 0$ for any $x^B \in [0, u^B]$. Therefore, we conclude that $\Pi^B_{\text{F-APA}}\left(G_{A_2^+}, G_{B_2^+}\right) \geq \Pi^B_{\text{F-APA}}\left(G_{A_2^+}, x^B\right)$ for any $x^B \geq 0$.

Proof of Result (iii) Similarly to the proof of Result (ii), we can prove that there exists no pure equilibrium if $qu^B - p > u^A > 0$. Now, let us denote by A_3^+ and B_3^+

the random variables that correspond to $G_{A_3^+}$ and $G_{B_3^+}$; we prove that if player B plays according to $G_{B_3^+}$, player A has no incentive to deviate from playing according to $G_{A_3^+}$.

$$\begin{aligned}\Pi^A_{\text{F-APA}}\left(G_{A_3^+}, G_{B_3^+}\right) &= \left[u^A \mathbb{P}\left(B_3^+ < \frac{p}{q}\right) - 0\right] \mathbb{P}\left(A_3^+ = 0\right) + \left[\alpha u^A \mathbb{P}\left(B_3^+ = \frac{p}{q}\right) - 0\right] \mathbb{P}\left(A_3^+ = 0\right) \\ &\quad + \int_0^{u^A} \left[u^A \mathbb{P}\left(B_3^+ < \frac{x+p}{q}\right) - x\right] \mathrm{d}G_{A_3^+}(x) \\ &= 0 + 0 + \int_0^{u^A} \left[u^A G_{B_2^+}\left(\frac{x+p}{q}\right) - x\right] \mathrm{d}G_{A_3^+}(x) \\ &= \int_0^{u^A} \left[u^A \left(\frac{-p}{u^A} + \frac{q}{u^A}\frac{(x+p)}{q}\right) - x\right] \mathrm{d}G_{A_3^+}(x) \\ &= 0.\end{aligned}$$

Moreover, trivially, for any $x^A > u^B$, we have $\Pi^A_{\text{F-APA}}\left(x^A, G_{B_3^+}\right) < 0$ and for any $x^A \in [0, u^B]$, we have

$$\begin{aligned}\Pi^A_{\text{F-APA}}\left(x^A, G_{B_3^+}\right) \leq& u^A G_{B_3^+}\left(\frac{x^A+p}{q}\right) - x^A \\ =& u^A \left[\frac{-p}{u^A} + \frac{q}{u^A}\frac{(x^A+p)}{q}\right] - x^A \\ =& 0 = \Pi^A_{\text{F-APA}}\left(G_{A_3^+}, G_{B_3^+}\right).\end{aligned}$$

Therefore, $\Pi^A_{\text{F-APA}}\left(G_{A_3^+}, G_{B_3^+}\right) \geq \Pi^A_{\text{F-APA}}\left(x^A, G_{B_3^+}\right)$ for any $x^A \geq 0$.

On the other hand, since $G_{B_3^+}$ is a continuous distribution on $\left[\frac{p}{q}, \frac{u^A+p}{q}\right]$, we do not need to consider the tie cases and we can deduce that:

$$\begin{aligned}\Pi^B_{\text{F-APA}}\left(G_{A_3^+}, G_{B_3^+}\right) &= \int_{p/q}^{\frac{u^A+p}{q}} \left[u^B G_{A_3^+}\left(qx - p\right) - x\right] \mathrm{d}G_{B_3^+}(x) \\ &= \int_{p/q}^{\frac{u^A+p}{q}} \left[u^B \left(1 - \frac{u^A}{qu^B} + \frac{qx-p}{qu^B}\right) - x\right] \frac{q}{u^A} \mathrm{d}x \\ &= u^B - \frac{u^A+p}{q}.\end{aligned}$$

Moreover, trivially, for any $x^B > u^B$, we have $\Pi^B\left(G_{A_3^+}, x^B\right) < 0 < u^B - \frac{u^A+p}{q}$; and for any $x^B \in [0, u^B]$, we have:

$$\Pi^B_{\text{F-APA}}\left(G_{A_3^+}, x^B\right) \leq u^B G_{A_3^+}\left(qx^B - p\right) - x^B = u^B \left(1 - \frac{u^A}{qu^B} + \frac{qx^B-p}{qu^B}\right) - x^B = u^B - \frac{u^A+p}{q}.$$

Therefore, we can conclude that $\Pi^B_{\text{F-APA}}\left(G_{A_3^+}, G_{B_3^+}\right) \geq \Pi^B\left(G_{A_3^+}, x^B\right)$ for any $x^B \geq 0$.

Finally, for a proof of uniqueness of the mixed equilibrium in Result (ii) and (iii), we can follow the scheme presented by Baye, Kovenock, and Vries (1996) and check through a series of lemmas. This is a standard approach in the literature of all-pay auction and we omit the detailed proof here. □

Proof of Theorem 6.2.13

For a proof of this theorem, we prove the following statement: There exists a constant $C^* > 0$ such that for any $\varepsilon \in (0,1]$ and in any game $\mathcal{GR}-\mathcal{CB}_n^C$ with $n \geq C^* \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, the following inequalities hold for any $(\tilde{\kappa}^A, \tilde{\kappa}^B) \in \mathbb{R}^2$ satisfying (6.18) (with $\tilde{\delta} = \varepsilon$) and any pure strategy $\boldsymbol{x}^A$, $\boldsymbol{x}^B$ of players A and B.

$$\Pi^A_{GR}(\boldsymbol{x}^A, \mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}) \leq \Pi^A_{GR}(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}, \mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}) + 2\varepsilon W, \tag{86}$$

$$\Pi^B_{GR}(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}, \boldsymbol{x}^B) \leq \Pi^B_{GR}(\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}, \mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}) + 2\varepsilon W. \tag{87}$$

This is very similar to Result (ii)-Theorem 4.2.3 that states the result for the IU^{γ^*} strategy in the game $\mathcal{CB}_n$. We can prove (86)-(87) by following the same scheme as in the proof of Theorem 4.2.3 presented in Section A. This proof is long and in this section, we will not repeat all the technical details. Instead, we only point out here the high-level ideas and show the main difference between the two proofs.

At the high-level, we need to prove two statements:

Statement 1: $\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\}$ is optimal against $\{G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\}$. This can be deduced from the fact that $\left(G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right)$ is the equilibrium of the all-pay auction with favoritism F-APA with $u^A = \tilde{\kappa}^A w_i$, $u^B = \tilde{\kappa}^B w_i$, $p := p_i$ and $q := q_i$.

Statement 2: $G_{A^{GR}_{n,i}}$ (resp. $G_{B^{GR}_{n,i}}$) uniformly converges toward $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$ (resp. $G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$) as n increases. Here, $G_{A^{GR}_{n,i}}$ (resp. $G_{B^{GR}_{n,i}}$) denotes the marginal of the $\mathrm{IU}_A^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ strategy (resp. $\mathrm{IU}_B^{\tilde{\kappa}^A,\tilde{\kappa}^B}$) corresponding to battlefield i. Formally, we want to prove that:

Proposition D.1. *There exists $C > 0$, such that for any $\varepsilon \in (0,1]$, any $n \geq C_1 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon,1/e\}}\right)$ and $i \in [n]$,*

$$\sup_{x \in [0,\infty)} \left| G_{A_n^{GR}}(x) - G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}(x) \right| \leq \varepsilon \quad \textit{and} \quad \sup_{x \in [0,\infty)} \left| G_{A_n^{GR}}(x) - G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}(x) \right| \leq \varepsilon. \tag{88}$$

To prove this, we can simply copy the proof of Lemma A.6 in Appendix A.6 (that is based on the Hoeffding's theorem) and replace $F_{A_i^n}, F_{B_i^n}$ by $G_{A^{GR}_{n,i}}, G_{B^{GR}_{n,i}}$ and replace $F_{A_i^*}, F_{B_i^*}$ by $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$. The only difference is that in Lemma A.6, in proving Inequality (32) leading to Inequality (33), we use the fact that in expectation, the budget constraints hold when players draw their allocations from $\{F_{A_i^*}\}_{i\in[n]}$ and $\{F_{B_i^*}\}_{i\in[n]}$. On the other hand, in the $\mathcal{GR}-\mathcal{CB}_n^C$ game, in expectation, $\left\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$ and $\left\{G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$

may violate the budget constraints; however, this violation only involves a small error $\tilde{\delta} = \varepsilon$. We only need to check that this violation does not affect the proof: denote $A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}$ the random variable corresponding to the distribution $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$ ($i \in [n]$), we need to find the condition on n such that when[33]

$$\left|\sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right] - X^A\right| \leq \varepsilon, \tag{89}$$

the following inequality holds with a number $\tau = O(\varepsilon^{-1})$ (see the proof of Lemma A.6 for a precise definition of τ):

$$\mathbb{P}\left(\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A\right| > \frac{1}{\tau}\right) \leq O(\varepsilon). \tag{90}$$

Indeed, we can prove this statement by first observing that under the assumption (89),

$$\mathbb{P}\left(\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A\right| > \frac{1}{\tau}\right) \leq \mathbb{P}\left(\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right]\right| > \frac{1}{\tau} - \varepsilon\right). \tag{91}$$

This can be proved by considering two cases:

- Case 1: If $\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A > 1/\tau$, then we have:

$$\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right] > \sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A - \varepsilon > \frac{1}{\tau} - \varepsilon.$$

- Case 2: If $\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A < -1/\tau$, then we have:

$$\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right] < \sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A + \varepsilon < -\frac{1}{\tau} + \varepsilon.$$

Therefore, we have

$$\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - X^A\right| > \frac{1}{\tau} \implies \left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right]\right| > \frac{1}{\tau} - \varepsilon,$$

thus, (91) holds.

Now, we apply the Hoeffding's theorem on the random variables $\left\{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right\}_{i\in[n]}$ that are bounded in $[0, \bar{w}\tilde{\kappa}^A]$ (see the definition of $G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}$ in Definition 6.2.7 and Table 6.2),

$$\mathbb{P}\left(\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right]\right| > \frac{1}{\tau} - \varepsilon\right) \leq 2\exp\left[\frac{2\left(\frac{1}{\tau} - \varepsilon\right)^2}{\sum_{i\in[n]} w_i \tilde{\kappa}^A}\right].$$

[33]This is a corollary of (6.18).

Note that $\tau = \mathcal{O}(\varepsilon)$, thus, $\frac{1}{\tau} - \varepsilon = \mathcal{O}(\varepsilon)$. Therefore, by choosing $n \geq \tilde{C}_1 \varepsilon^{-2} \ln\left(\frac{1}{\min\{\varepsilon, 1/e\}}\right)$, where C_1 is the constant chosen similarly as in the proof of Lemma A.6 (see Appendix A.6), we have that

$$\mathbb{P}\left(\left|\sum_{i\in[n]} A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B} - \sum_{i\in[n]} \mathbb{E}\left[A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}\right]\right|\right) \leq \mathcal{O}(\varepsilon).$$

Combining this with (91), we conclude that (90) holds. At a high-level, the condition on n mentioned above is the same as the one indicated in the proof of Lemma A.6; therefore, the fact that $\left\{G_{A_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}, G_{B_i^{\tilde{\kappa}^A,\tilde{\kappa}^B}}\right\}_{i\in[n]}$ are only "almost" the optimal univariate distributions does not affect the proof and Statement 2 can be proved.

The remaining proof of (86) and (87) can be done by copying the proof of Theorem 4.2.3 (see Section A) and replace the elements of the $\mathcal{CB}_n$ game by the corresponding elements of the $\mathcal{GR}-\mathcal{CB}_n^C$ game. We omit the details here.

Appendix E

Supplementary Materials for Chapter 8 on SOOSP and the Online Semi-Bandit CB Game

E.1 Proof of Algorithm 11's Output

Proof. Fixing an edge $e \in \mathcal{E}$, we prove that when Algorithm 11 takes the edges weights $\{w_t(e), e \in \mathcal{E}\}$ as the input, it outputs exactly $q_t = \sum_{\boldsymbol{p} \in \mathbb{O}_t(e)} x_t(\boldsymbol{p})$. We note that if $e' \in \mathfrak{R}_t(e) := \{e' : e' \to e\}$, then $\{\boldsymbol{p} \in \mathcal{P} : \boldsymbol{p} \ni e'\} \subset \mathbb{O}_t(e)$.

We denote $|\mathfrak{R}_t(e)| = \rho_e$ and label the edges in the set $\mathfrak{R}_t(e)$ by $\{e_1, e_2, \dots, e_{\rho_e}\}$. The for-loop in lines 4-8 of Algorithm 11 consecutively run with the edges in $R_t(e)$ as follows:

(i) After the for-loop runs for e_1, we have $K(e_1) := \sum_{\boldsymbol{p} \ni e_1} \prod_{\bar{e} \in \boldsymbol{p}} \bar{w}(\bar{e}) = \sum_{\boldsymbol{p} \ni e_1} w_t(\boldsymbol{p})$; therefore, $q_t(e) = \sum_{\boldsymbol{p} \ni e_1} x_t(\boldsymbol{p})$ since $H^*(s,d) = \sum_{\boldsymbol{p} \in \mathcal{P}} w_t(\boldsymbol{p})$ computed from the original weights $w_t(\bar{e}), \bar{e} \in \mathcal{E}$. From line 8 that sets $\bar{w}(e_1) := 0$, henceforth in Algorithm 11, the weight $\bar{w}(\boldsymbol{p}) := \prod_{e \in \boldsymbol{p}} \bar{w}(e)$ of any path $\boldsymbol{p}$ that contains e_1 is set to 0.

(ii) Let the for-loop run for e_2, we have $K(e_2) := \sum_{\boldsymbol{p} \ni e_2} \bar{w}(\boldsymbol{p}) = \sum_{\{\boldsymbol{p} \ni e_2\} \setminus \{\boldsymbol{p} \ni e_1\}} w_t(\boldsymbol{p})$ because any path $\boldsymbol{p} \ni e_1$ has the weight $\bar{w}(\boldsymbol{p}) = 0$. Therefore,

$$q_t(e) = \sum\nolimits_{\boldsymbol{p} \ni e_1} x_t(\boldsymbol{p}) + \sum\nolimits_{\{\boldsymbol{p} \ni e_2\} \setminus \{\boldsymbol{p} \ni e_1\}} x_t(\boldsymbol{p}).$$

(iii) Similarly, after the for-loop runs for e_i (where $i \in \{3, \dots, \rho_e\}$), we have:

$$q_t(e) = \sum_{k=1}^{i} \left(\sum_{\{\boldsymbol{p} \ni e_k\} \setminus \bigcup\limits_{j<k} \{\boldsymbol{p} \ni e_j\}} x_t(\boldsymbol{p}) \right).$$

(iv) Therefore, after the for-loop finishes running for every edge in $\mathfrak{R}_t(e)$; we have $q_t := \sum_{\boldsymbol{p} \in \mathbb{O}_t(e)} x_t(\boldsymbol{p})$ where each term $x_t(\boldsymbol{p})$ was only counted once even if $\boldsymbol{p}$ contains more than one edge that reveals the edge e.

□

E.2 Proof of Theorem 8.2.1

Theorem 8.2.1. *The expected regret of the* Exp3-OE *algorithm in the* SOOSP *satisfies:*

$$R_T \leq \log(P)/\eta + \left[\beta + (n \cdot \eta)/2\right] \cdot \sum\nolimits_{t \in [T]} Q_t. \tag{8.3}$$

Proof. We first denote[34] $W_t := \sum_{\boldsymbol{p}\in\mathcal{P}} w_t(\boldsymbol{p}), \forall t \in [T]$. From line 9 of Algorithm 10, we trivially have:

$$w_{t+1}(\boldsymbol{p}) = w_t(\boldsymbol{p}) \cdot \exp(-\eta \hat{L}_t(\boldsymbol{p})), \forall \boldsymbol{p} \in \mathcal{P}, \forall t \in [T-1]. \tag{92}$$

We recall that $\hat{L}_t(\boldsymbol{p}) := \sum_{e\in\boldsymbol{p}} \hat{\boldsymbol{\ell}}_t(e)$ and the notation $\mathbb{E}_t$ denoting the expectation w.r.t. to the randomness in choosing $\tilde{\boldsymbol{p}}_t$ in Algorithm 10 (i.e., w.r.t. the information up to time $t-1$). From (8.2), we have:

$$\mathbb{E}_t\left[\hat{L}_t(\boldsymbol{p})\right] \leq L_t(\boldsymbol{p}) := \sum_{e\in\boldsymbol{p}} \boldsymbol{\ell}_t(e), \forall \boldsymbol{p} \in \mathcal{P}. \tag{93}$$

Under the condition that $0 < \eta$, we obtain:

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{\boldsymbol{p}\in\mathcal{P}} \frac{w_{t+1}(\boldsymbol{p})}{W_t} \\
&= \sum_{\boldsymbol{p}\in\mathcal{P}} \frac{w_t(\boldsymbol{p}) \cdot \exp(-\eta \hat{L}_t(\boldsymbol{p}))}{W_t} \\
&= \sum_{\boldsymbol{p}\in\mathcal{P}} x_t(\boldsymbol{p}) \cdot \exp(-\eta \hat{L}_t(\boldsymbol{p}))) \\
&\leq \sum_{\boldsymbol{p}\in\mathcal{P}} \left[x_t(\boldsymbol{p})\left(1 - \eta \hat{L}_t(\boldsymbol{p}) + \frac{\eta^2}{2}(\hat{L}_t(\boldsymbol{p}))^2\right)\right] \\
&= 1 - \sum_{\boldsymbol{p}\in\mathcal{P}} \left[x_t(\boldsymbol{p})\left(\eta \hat{L}_t(\boldsymbol{p}) - \frac{\eta^2}{2}(\hat{L}_t(\boldsymbol{p}))^2\right)\right].
\end{aligned} \tag{94}$$

Here, the second equality comes from (92) and the inequality comes from the fact that $\exp(-a) \leq 1 - a + a^2/2$ for $a := \eta \hat{L}_t(\boldsymbol{p}) \geq 0$. Now, we use the inequality $\ln(1-y) \leq -y$, $\forall y < 1$ for $y := \sum_{\boldsymbol{p}\in\mathcal{P}} \left[x_t(\boldsymbol{p})\left(\eta \hat{L}_t(\boldsymbol{p}) - \frac{\eta^2}{2}(\hat{L}_t(\boldsymbol{p}))^2\right)\right]$,[35] then from (94), we obtain

$$\begin{aligned}
&\ln\left(\frac{W_{T+1}}{W_1}\right) \\
&= \sum_{t=1}^{T} \ln\left(\frac{W_{t+1}}{W_t}\right) \\
&\leq \sum_{t=1}^{T} \left(-\eta \sum_{\boldsymbol{p}\in\mathcal{P}} x_t(\boldsymbol{p}) \hat{L}_t(\boldsymbol{p}) + \frac{\eta^2}{2} \sum_{\boldsymbol{p}\in\mathcal{P}} x_t(\boldsymbol{p})(\hat{L}_t(\boldsymbol{p}))^2\right).
\end{aligned} \tag{95}$$

On the other hand, let us fix a path $\boldsymbol{p}^* \in \mathcal{P}$, then

$$\begin{aligned}
&\ln\left(\frac{W_{T+1}}{W_1}\right) \\
&\geq \ln\left(\frac{w_{T+1}(\boldsymbol{p}^*)}{W_1}\right)
\end{aligned}$$

[34] We recall that $w_t(\boldsymbol{p}) := \prod_{e\in\boldsymbol{p}} w_t(e)$.

[35] We can easily check that $\eta \hat{L}_t(\boldsymbol{p}) - \eta^2 \hat{L}_t(\boldsymbol{p})^2/2 < 1$ for any $\eta > 0$ and thus, $\sum_{\boldsymbol{p}\in\mathcal{P}} \left[x_t(\boldsymbol{p})\left(\eta \hat{L}_t(\boldsymbol{p}) - \frac{\eta^2}{2}(\hat{L}_t(\boldsymbol{p}))^2\right)\right] < 1$.

$$= \ln \frac{w_T(\boldsymbol{p}^*) \exp(-\eta \hat{L}_T(\boldsymbol{p}^*))}{P}$$
$$= \ln \frac{w_{T-1}(\boldsymbol{p}^*) \exp(-\eta \hat{L}_T(\boldsymbol{p}^*) - \eta \hat{L}_{T-1}(\boldsymbol{p}^*))}{P}$$
$$= -\eta \sum_{t=1}^{T} \hat{L}_t(\boldsymbol{p}^*) - \ln(P). \tag{96}$$

In the arguments leading to (96), we again use (92) and the fact that $w_1(\boldsymbol{p}) = 1, \forall \boldsymbol{p} \in \mathcal{P}$, including $w_1(\boldsymbol{p}^*)$. Therefore, combining (95) and (96) then dividing both sides by η, we have:

$$\sum_{t=1}^{T} \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \hat{L}_t(\boldsymbol{p})$$
$$\leq \frac{\ln(P)}{\eta} + \sum_{t=1}^{T} \hat{L}_t(\boldsymbol{p}^*) + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) (\hat{L}_t(\boldsymbol{p}))^2. \tag{97}$$

Now, we take $\mathbb{E}_t$ on both sides of (97), then we apply (93) to obtain:

$$\sum_{t=1}^{T} \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t[\hat{L}_t(\boldsymbol{p})]$$
$$\leq \frac{\ln(P)}{\eta} + \sum_{t=1}^{T} L_t(\boldsymbol{p}^*) + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t[\hat{L}_t(\boldsymbol{p})^2]. \tag{98}$$

Now, we look for a lower bound of $\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p}) \right]$. For any fixed $\boldsymbol{p} \in \mathcal{P}$, we consider:

$$\mathbb{E}_t \left[\sum_{e \in \boldsymbol{p}} \hat{\boldsymbol{\ell}}_t(e) \right] = \sum_{\tilde{\boldsymbol{p}} \in \mathcal{P}} \left[x_t(\tilde{\boldsymbol{p}}) \sum_{e \in \boldsymbol{p}} \left(\frac{\boldsymbol{\ell}_t(e)}{q_t(e) + \beta} \mathbb{I}_{\{e \in \mathbb{O}_t(\tilde{\boldsymbol{p}})\}} \right) \right]$$
$$= \sum_{e \in \boldsymbol{p}} \sum_{\tilde{\boldsymbol{p}} \in \mathbb{O}(e)} x_t(\tilde{\boldsymbol{p}}) \frac{\boldsymbol{\ell}_t(e)}{q_t(e) + \beta}$$
$$= \sum_{e \in \boldsymbol{p}} \frac{q_t(e) \boldsymbol{\ell}_t(e)}{q_t(e) + \beta}. \tag{99}$$

Using (99) and recalling that $\boldsymbol{\ell}_t(e) \leq 1, \forall e \in \mathcal{E}$, we have:

$$\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p}) \right] - \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) L_t(\boldsymbol{p})$$
$$= \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \sum_{e \in \boldsymbol{p}} \frac{q_t(e) \boldsymbol{\ell}_t(e)}{q_t(e) + \beta} - \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \sum_{e \in \boldsymbol{p}} \boldsymbol{\ell}_t(e)$$
$$= \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \sum_{e \in \boldsymbol{p}} \boldsymbol{\ell}_t(e) \left(\frac{q_t(e)}{q_t(e) + \beta} - 1 \right)$$

$$
\begin{aligned}
&\geq - \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \sum_{e \in \boldsymbol{p}} \frac{\beta}{q_t(e) + \beta} \\
&= - \beta \sum_{e \in \mathcal{E}} \frac{\sum_{\boldsymbol{p} \ni e} x_t(\boldsymbol{p})}{q_t(e) + \beta} \\
&= - \beta Q_t. \qquad (100)
\end{aligned}
$$

Therefore, a lower bound of $\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p}) \right]$ is $\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) L_t(\boldsymbol{p}) - \beta Q_t$.

Now, we look for an upper bound of $\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p})^2 \right]$. To do this, fix $\boldsymbol{p} \in \mathcal{P}$, we consider

$$
\begin{aligned}
&\mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p})^2 \right] \\
=&\mathbb{E}_t \left[\left(\sum\nolimits_{e \in \boldsymbol{p}} \hat{\boldsymbol{\ell}}_t(e) \right)^2 \right] \\
\leq& n \cdot \mathbb{E}_t \left[\sum\nolimits_{e \in \boldsymbol{p}} \hat{\boldsymbol{\ell}}_t(e)^2 \right] \\
=& n \cdot \sum_{\tilde{\boldsymbol{p}} \in \mathcal{P}} \left[x_t(\tilde{\boldsymbol{p}}) \sum_{e \in \boldsymbol{p}} \left(\frac{\boldsymbol{\ell}_t(e)}{q_t(e) + \beta} \mathbb{I}_{\{e \in \mathbb{O}_t(\tilde{\boldsymbol{p}})\}} \right)^2 \right] \\
\leq& n \cdot \sum_{e \in \boldsymbol{p}} \sum_{\tilde{\boldsymbol{p}} \in \mathbb{O}_t(e)} x_t(\tilde{\boldsymbol{p}}) \frac{1}{(q_t(e) + \beta)^2} \\
=& n \cdot \sum_{e \in \boldsymbol{p}} q_t(e) \frac{1}{(q_t(e) + \beta)^2} \\
\leq& n \cdot \sum_{e \in \boldsymbol{p}} \frac{1}{q_t(e) + \beta}. \qquad (101)
\end{aligned}
$$

The first inequality comes from applying Cauchy–Schwarz inequality. The second inequality comes from the fact that $\boldsymbol{\ell}_t(e) \leq 1$ and the last inequality comes from $q_t(e) \leq q_t(e) + \beta$ since $\beta > 0$.

Now, applying (101), we can bound

$$
\begin{aligned}
\sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \mathbb{E}_t \left[\hat{L}_t(\boldsymbol{p})^2 \right] \leq& n \cdot \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) \sum_{e \in \boldsymbol{p}} \frac{1}{q_t(e) + \beta} \\
=& n \cdot \sum_{e \in \mathcal{E}} \sum_{\boldsymbol{p} \ni e} x_t(\boldsymbol{p}) \frac{1}{q_t(e) + \beta} \\
=& n \cdot \sum_{e \in \mathcal{E}} \frac{r_t(e)}{q_t(e) + \beta} = n \cdot Q_t. \qquad (102)
\end{aligned}
$$

Here, we recall the notation $r_t(e)$ and Q_t defined in Section 8.2.2. Replacing (100) and (102) into (98), we have that the following inequality holds for any $\boldsymbol{p}^* \in \mathcal{P}$.

$$
\sum_{t=1}^{T} \sum_{\boldsymbol{p} \in \mathcal{P}} x_t(\boldsymbol{p}) L_t(\boldsymbol{p}) - \sum_{t=1}^{T} \beta Q_t - \sum_{t=1}^{T} L_t(\boldsymbol{p}^*)
$$

$$\leq \frac{\ln(P)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} n Q_t.$$

Therefore, we conclude that

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \sum_{p \in \mathcal{P}} x_t(\boldsymbol{p}) L_t(\boldsymbol{p}) - \sum_{t=1}^{T} L_t(\boldsymbol{p}^*) \\ &\leq \frac{\ln(P)}{\eta} + \sum_{t=1}^{T} Q_t \left(n \frac{\eta}{2} + \beta \right). \end{aligned}$$

□

E.3 Lemmas on Graphs' Independence Numbers

In this section, we present some lemmas in graph theory that will be used in the next section to prove Theorem 8.2.2. Consider a graph $\tilde{G}$ whose vertices set and edges set are respectively denoted by $\tilde{\mathcal{V}}$ and $\tilde{\mathcal{E}}$. Let $\tilde{\alpha}$ be its independence number.

Lemma E.1. *Let $\tilde{G}$ be an directed graph and I_v be the in-degree of the vertex $v \in \tilde{\mathcal{V}}$, then*

$$\sum_{v \in \tilde{\mathcal{V}}} [1/(1 + I_v)] \leq 2\tilde{\alpha} \ln\left(1 + |\tilde{\mathcal{V}}|/\tilde{\alpha}\right).$$

A proof of this lemma can be found in Lemma 10 of Alon, Cesa-Bianchi, Gentile, and Mansour (2013).

Lemma E.2. *Let $\tilde{G}$ be a directed graph with self-loops and consider the numbers $k(v) \in [0,1]$, $\forall v \in \tilde{\mathcal{V}}$ such that there exists $\gamma > 0$ and $\sum_{v \in \tilde{\mathcal{V}}} k(v) \leq \gamma$. For any $c > 0$, we have*

$$\sum_{v \in \tilde{\mathcal{V}}} \frac{k(v)}{\frac{1}{\gamma} \sum\limits_{v' \to v} k(v') + c} \leq 2\gamma \tilde{\alpha} \ln\left(1 + \frac{\gamma \lceil |\tilde{\mathcal{V}}|^2/c \rceil + |\tilde{\mathcal{V}}|}{\tilde{\alpha}}\right) + 2\gamma.$$

A proof of this lemma can be found in Lemma 1 of Kocák et al. (2014).

Lemma E.3. *Let $\tilde{G}$ be an undirected graph with self-loops and consider the numbers $k(v) \geq 0$, $v \in \tilde{\mathcal{V}}$. We have*

$$\sum_{v \in \tilde{\mathcal{V}}} \left[k(v) / \sum_{v' \to v} k(v') \right] \leq \tilde{\alpha}.$$

This lemma is extracted from Lemma 3 of Mannor and Shamir (2011).

E.4 Proof of Theorem 8.2.2

Theorem 8.2.2. *Let us define $M := \lceil 2E^2/\beta \rceil$, $N_t := \log\left(1 + \frac{M+E}{\alpha_t}\right)$ and $K_t := \log\left(1 + \frac{nM+E}{\alpha_t}\right)$. Upper-bounds of Q_t in different cases of G_t^O are given in the following table:*

	Satisfies $(A1)$	Does not satisfy $(A1)$
Symmetric	α_t	$n\alpha_t$
Non-Symmetric	$1+2\alpha_t N_t$	$2n\,(1+\alpha_t K_t)$

Case 1: G_t^O *does not satisfy Assumption* $(A1)$. Fixing an edge e, due to the fact that n is the length of the longest paths in $\mathcal{P}$, we have

$$\begin{aligned} n q_t(e) = n \sum_{\boldsymbol{p} \in \mathbb{O}_t(e)} x_t(\boldsymbol{p}) &\geq \sum_{e' \to e} \sum_{\boldsymbol{p} \ni e'} x_t(\boldsymbol{p}) = \sum_{e' \to e} r_t(e') \\ \Rightarrow Q_t = \sum_{e \in \mathcal{E}} \frac{r_t(e)}{q_t(e) + \beta} &\leq \sum_{e \in \mathcal{E}} \frac{r_t(e)}{\frac{1}{n} \sum\limits_{e' \to e} r_t(e') + \beta}. \end{aligned} \tag{103}$$

Case 1.1: If G_t^O is a non-symmetric (i.e., directed) graph, we apply Lemma E.2 with $\gamma = n, c = \beta$ on the graph $\tilde{G} = G_t^O$ (whose vertices set $\tilde{\mathcal{V}}$ corresponds to the edges set $\mathcal{E}$ of G) and the numbers[36] $k(v_e) = r_t(e), \forall v_e \in \tilde{\mathcal{V}}$ (i.e., $\forall e \in \mathcal{E}$). We obtain the following inequality:

$$\sum_{e \in \mathcal{E}} \frac{r_t(e)}{\frac{1}{n} \sum\limits_{e' \to e} r_t(e') + \beta} \leq 2 n \alpha_t \ln \left(1 + \frac{n \lceil E^2 / \beta \rceil + E}{\alpha_t} \right) + 2n.$$

Case 1.2: If G_t^O is a symmetric (i.e. undirected) graph, we apply Lemma E.3 with the graph $\tilde{G} = G_t^O$ (whose vertices set $\tilde{\mathcal{V}}$ corresponds to the edges set $\mathcal{E}$ of the graph G) and the numbers $k(v_e) = r_t(e), \forall v_e \in \tilde{V}$ (i.e., $\forall e \in \mathcal{E}$) to obtain:

$$\sum_{e \in \mathcal{E}} \frac{r_t(e)}{\frac{1}{n} \sum\limits_{e' \to e} r_t(e') + \beta} \leq n \sum_{e \in \mathcal{E}} \frac{r_t(e)}{\sum\limits_{e' \to e} r_t(e')} \leq n \alpha_t.$$

Case 2: G_t^O *satisfies Assumption* $(A1)$. Under this Assumption, $q_t(e) = \sum_{e' \to e} r_t(e')$ due to the definition of $\mathbb{O}_t(e)$. Therefore, $Q_t = \sum_{e \in \mathcal{E}} \left[r_t(e) / \left(\sum_{e' \to e} r_t(e') + \beta \right) \right]$.

Case 2.1: If G_t^O is a non-symmetric (i.e., directed) graph. We consider a discretized version of $x_t(\boldsymbol{p})$ for any path $\boldsymbol{p} \in \mathcal{P}$ that is $\tilde{x}_t(\boldsymbol{p}) := k/M$ where k is the unique integer such that $(k-1)/M \leq x_t(\boldsymbol{p}) \leq k/M$; thus, $\tilde{x}_t(\boldsymbol{p}) - 1/M \leq x_t(\boldsymbol{p}) \leq \tilde{x}_t(\boldsymbol{p})$.

Let us denote the discretized version of $r_t(e)$ by $\tilde{r}_t(e) := \sum_{\boldsymbol{p} \ni e} \tilde{x}_t(\boldsymbol{p})$. We deduce that $r_t(e) \leq \tilde{r}_t(e)$ and

$$\sum_{e' \to e} r_t(e) \geq \sum_{e' \to e} \left(\tilde{r}_t(e') - \frac{1}{M} \right) \geq \sum_{e' \to e} \tilde{r}_t(e') - \frac{E}{M}.$$

We obtain the bound:

$$Q_t = \sum_{e \in \mathcal{E}} \frac{r_t(e)}{\left(\sum\limits_{e' \to e} r_t(e') + \beta \right)} \leq \sum_{e \in \mathcal{E}} \frac{\tilde{r}_t(e)}{\sum\limits_{e' \to e} \tilde{r}_t(e') - E/M + \beta}. \tag{104}$$

We now consider the following inequality: If $a, b \geq 0$ and $a + b \geq B > A > 0$, then

$$\frac{a}{a + b - A} \leq \frac{a}{a + b} + \frac{A}{B - A}. \tag{105}$$

[36] We verify that these numbers satisfy

$$\sum_{e \in \mathcal{E}} r_t(e) = \sum_{e \in \mathcal{E}} \sum_{\boldsymbol{p} \ni e} x_t(\boldsymbol{p}) = \sum_{\boldsymbol{p} \in \mathcal{P}} \sum_{e \in \boldsymbol{p}} x_t(\boldsymbol{p}) \leq \sum_{\boldsymbol{p} \in \mathcal{P}} n x_t(\boldsymbol{p}) = n.$$

A proof of this inequality can be found in Lemma 12 of Alon, Cesa-Bianchi, Gentile, and Mansour (2013). Applying (105)[37] with $a = \tilde{r}_t(e)$, $b = \sum_{e' \to e, e' \neq e} \tilde{r}_t(e') + \beta$, $A = \frac{E}{M}$, and $B = \beta$ to (104),

$$Q_t \leq \sum_{e \in \mathcal{E}} \left(\frac{\tilde{r}_t(e)}{\sum_{e' \to e} \tilde{r}_t(e') + \beta} + \frac{E/M}{\beta - E/M} \right)$$

$$\leq \sum_{e \in \mathcal{E}} \frac{\tilde{r}_t(e)}{\sum_{e' \to e} \tilde{r}_t(e')} + 1. \tag{106}$$

The last inequality comes from the fact that $\frac{E}{M\beta - E} \leq \frac{E}{2E^2 - E} \leq \frac{1}{2E-1} \leq \frac{1}{E}, \forall E \geq 1$.

Finally, we create an auxiliary graph G^*_t such that:

(i) Corresponding to each edge e in G (i.e., each vertex v_e in G^O_t), there is a clique, called $\mathbb{C}(e)$, in the auxiliary graph G^*_t with $M\tilde{r}_t(e) \in \mathbb{N}$ vertices.

(ii) In each clique $\mathbb{C}(e)$ of G^*_t, all vertices are pairwise connected with length-two cycles. That is, for any $k, k' \in \mathbb{C}(e)$, there is an edge from k to k' and there is an edge from k' to k in G^*_t.

(iii) If $e \to e'$, i.e., there is an edge in G^O_t connecting v_e and $v_{e'}$; then in G^*_t, all vertices in the clique $\mathbb{C}(e)$ are connected to all vertices in $\mathbb{C}(e')$.

We observe that the independence number α_t of G^O_t is equal to the independence number of G^*_t. Moreover, the in-degree of each vertex $k \in (e)$ in the graph G^*_t is:

$$I^*_k = M\tilde{r}_t(e) - 1 + \sum_{e' \to e, e' \neq e} M\tilde{r}_t(e') = \sum_{e' \to e} M\tilde{r}_t(e') - 1. \tag{107}$$

Let us denote V^*_t the set of all vertices in G^*_t, then we have:

$$\sum_{e \in \mathcal{E}} \frac{\tilde{r}_t(e)}{\sum_{e' \to e} \tilde{r}_t(e')} = \sum_{e \in \mathcal{E}} \frac{M\tilde{r}_t(e)}{\sum_{e' \to e} M\tilde{r}_t(e')} = \sum_{e \in \mathcal{E}} \sum_{k \in \mathbb{C}(e)} \frac{1}{I^*_k + 1}$$

$$= \sum_{k \in V^*_t} \frac{1}{\tilde{I}_k + 1} \leq 2\alpha_t \ln\left(1 + \frac{M + E}{\alpha_t}\right). \tag{108}$$

Here, the second equality comes from the fact that $|\mathbb{C}(e)| = M\tilde{r}_t(e)$ and (107). The inequality is obtained by applying Lemma E.1 to the graph G^*_t and the fact that $|V^*_t| = \sum_{e \in \mathcal{E}} M\tilde{r}_t(e) \leq M \sum_{e \in \mathcal{E}} (r_t(e) + 1/M) \leq E + M$.

In conclusion, combining (106) and (108), we obtain the regret-upper bound as given in Theorem 8.2.2 for this case of the observation graph.

Case 2.2: Finally, if G^O_t is a symmetric (i.e., undirected) graph, we again apply Lemma E.3 to the graph $\tilde{G} = G^O_t$ and the numbers $k(v_e) = r_t(e)$ to obtain that $Q_t \leq \sum_{e \in \mathcal{E}} \left[r_t(e) / \sum_{e' \to e} r_t(e') \right] \leq \alpha_t$. □

[37] Trivially, we can verify that $a + b \geq B$ and $B > A$ comes from the fact that $\beta \geq \beta \frac{1}{E} > \frac{E}{\lceil 2E^2/\beta \rceil}$.

E.5 Parameters Tuning for Exp3-OE: Proof of Corollary 8.2.3

In this section, we suggest a choice of β and η that guarantees the expected regret given in Corollary 8.2.3.

Corollary 8.2.3. *In* SOOSP, *let α be an upper bound of $\alpha_t, \forall t \in [T]$; with appropriate choices of the parameters η and β, the expected regret of the* Exp3-OE *algorithm is:*[38]

(i) $R_T \leq \tilde{O}(n\sqrt{T\alpha\log(P)})$ *in the general cases.*

(ii) $R_T \leq \tilde{O}(\sqrt{nT\alpha\log(P)})$ *if Assumption (A1) is satisfied by the observation graphs G_t^O, $\forall t \in [T]$.*

Case 1: Non-symmetric (i.e. directed) observation graphs that do not satisfy Assumption $(A1)$. We find the parameters β and η such that $R_t \leq \tilde{O}\left(n\sqrt{T\alpha}\right)$. We note that $\alpha_t \geq 1$, $\forall t \in [T]$; therefore, recalling that α is an upper bound of α_t, from Theorem 8.2.1 and Theorem 8.2.2, we have:

$$\begin{aligned} R_T \leq & \frac{\ln(P)}{\eta} + \sum_{t=1}^{T}\left(n\frac{\eta}{2}+\beta\right)2n\left[1+\alpha_t\ln\left(1+\frac{nM+E}{\alpha_t}\right)\right] \\ \leq & \frac{\ln(P)}{\eta} + T\left(n\frac{\eta}{2}+\beta\right)2n\left[1+\alpha\ln\left(\alpha+nM+E\right)\right] \\ = & \frac{\ln(P)}{\eta} + \eta T n^2\left[1+\alpha\ln\left(\alpha+nM+E\right)\right] \\ & + 2\beta T n\left[1+\alpha\ln\left(\alpha+nM+E\right)\right]. \end{aligned} \tag{109}$$

Recalling that $M := \lceil 2E^2/\beta \rceil$, by choosing any

$$\beta \leq 1/\sqrt{Tn[1+\alpha\ln(\alpha+n\lceil E^2/\beta\rceil+E)]}, \tag{110}$$

$$\text{and } \eta = \sqrt{\ln(P)}/\sqrt{n^2T\left[1+\alpha\ln\left(\alpha+n\lceil E^2/\beta\rceil+E\right)\right]},$$

we obtain the bound:

$$\begin{aligned} R_T \leq & 2n\sqrt{T\ln(P)\cdot[1+\alpha\ln(\alpha+nM+E)]} \\ & + 2\sqrt{Tn[\alpha+\alpha\ln(\alpha+nM+E)]} \\ \leq & \tilde{O}\left(n\sqrt{T\alpha\ln(P)}\right). \end{aligned} \tag{111}$$

In practice, as long as it satisfies (110), the larger β is, the better upper-bounds that Exp3-OE gives. As an example that (110) always has at least one solution, we now prove that it holds with

$$\beta^* = \frac{-Tn^2E^2+\sqrt{(Tn^2E^2)^2+4Tn(1+\alpha\ln\alpha+E+n)}}{2Tn(1+\alpha\ln\alpha+E+n)}. \tag{112}$$

[38] Recall that $\tilde{O}$ is the variant of the asymptotic notation O that ignores the logarithmic factors (in terms of n and T).

Indeed, $\beta^* > 0$ and it satisfies:

$$\begin{aligned}
&\beta^{*2} \cdot Tn(1+\alpha\ln\alpha+E+n)+\beta^* Tn^2E^2 = 1.\\
\Rightarrow&\beta^{*2} \cdot Tn(1+\alpha\ln\alpha+E)+\beta^{*2}Tn^2\left(\frac{E^2}{\beta^*}+1\right) = 1\\
\Rightarrow&\beta^{*2} \cdot Tn(1+\alpha\ln\alpha+E)+\beta^{*2}Tn^2\lceil\frac{E^2}{\beta^*}\rceil \le 1\\
\Rightarrow&\beta^* \le \frac{1}{\sqrt{Tn\,(1+\alpha\ln\alpha+E+nM)}}.
\end{aligned}$$

On the other hand, applying the inequality $\ln(1+x) \le x, \forall x \ge 0$, we have:

$$\begin{aligned}
&\frac{nM+E}{\alpha} \ge \ln\left(1+\frac{nM+E}{\alpha}\right)\\
\Rightarrow&\frac{nM+E}{\alpha}+\ln\alpha \ge \ln(\alpha+nM+E)\\
\Rightarrow&nM+E+\alpha\ln\alpha+1 \ge \alpha\ln(\alpha+nM+E)+1\\
\Rightarrow&\frac{1}{\sqrt{Tn\,(1+\alpha\ln\alpha+nM+E)}} \le \frac{1}{\sqrt{Tn\,(\alpha\ln(\alpha+nM+E)+1)}}.
\end{aligned}$$

Therefore, β^* satisfies (110). Finally, choosing $\beta = \beta^* = \Omega\left(nE^2/[1+\alpha\ln\alpha+E+n]\right)$ as in (112), we have

$$M = \lceil 2E^2/\beta\rceil \le O([1+\alpha\ln\alpha+E+n]/n).$$

Combining this with (111), we obtain the regret bound indicated in Section 8.2.2.

Case 2: symmetric observation graphs that do not satisfy $(A1)$. Trivially, we have that if $\beta := 1/\sqrt{n\alpha T}$ and $\eta = 2\sqrt{\ln(P)}/\sqrt{n^2\alpha T}$, then

$$\begin{aligned}
R_T &\le \frac{\ln(P)}{\eta}+\left(n\frac{\eta}{2}+\beta\right)n\alpha T\\
&= \frac{1}{2}n\sqrt{\alpha T\ln(P)}+n\sqrt{\alpha T\ln(P)}+\sqrt{n\alpha T} \qquad (113)\\
&\le \tilde{O}\left(n\sqrt{\alpha T\ln(P)}\right).
\end{aligned}$$

Case 3: non-symmetric observation graphs G_t^O satisfying Assumption $(A1)$**,** $\forall t$. We will prove that $R_T \le \tilde{O}\left(\sqrt{nT\alpha\ln(P)}\right)$ for any

$$\beta \le 1/\sqrt{T\alpha[1+2\ln\left(1+\lceil E^2/\beta\rceil+E\right)]}, \qquad (114)$$

$$\eta = 2\sqrt{\ln(P)}/\sqrt{Tn\alpha\,[1+2\ln(\alpha+M+E)]}. \qquad (115)$$

Indeed, from Theorem 8.2.1 and Theorem 8.2.2, we have:

$$R_T \le \frac{\ln(P)}{\eta}+\sum_{t=1}^{T}\left(n\frac{\eta}{2}+\beta\right)\left[1+2\alpha_t\ln\left(1+\frac{M+E}{\alpha_t}\right)\right]$$

$$\leq \frac{\ln(P)}{\eta} + \sum_{t=1}^{T} \left(n\frac{\eta}{2}+\beta\right) [\alpha + 2\alpha \ln(1 + M + E)]$$
$$= \frac{\ln(P)}{\eta} + \eta T \alpha \frac{n}{2} [1 + 2\ln(1 + M + E)]$$
$$+ \beta T \alpha [1 + 2\ln(1 + M + E)]. \tag{116}$$

We replace (114) and (115) into (116) and obtain:

$$R_T \leq \frac{3}{2}\sqrt{Tn\alpha [1 + 2\ln(1 + M + E)] \cdot \ln(P)}$$
$$+ \sqrt{T\alpha [1 + 2\ln(1 + M + E)]}. \tag{117}$$
$$\leq \tilde{O}\left(\sqrt{n\alpha T \ln(P)}\right).$$

A choice for β that satisfies (114) is

$$\beta^* := \frac{-T\alpha E^2 + \sqrt{(T\alpha E^2)^2 + T\alpha(3 + 2E)}}{T\alpha(3 + 2E)}. \tag{118}$$

Moreover, with this choice of $\beta^* = \Omega(E^2/(3+2E))$, we deduce $M := \lceil 2E^2/\beta^* \rceil \leq O(3 + 2E)$. Combining this with (117), we obtain the regret bound indicated in Section 8.2.2.

Case 4: all observation graphs are symmetric and satisfy $(A1)$. From Theorem 8.2.1 and Theorem 8.2.2, we trivially have that if $\beta := 1/\sqrt{\alpha T}$ and $\eta = 2\sqrt{\ln(P)}/\sqrt{n\alpha T}$, then $R_T \leq 2\sqrt{n\alpha T \ln(P)} + \sqrt{\alpha T} \leq \tilde{O}\left(\sqrt{n\alpha T \ln(P)}\right)$.

E.6 The Actions Set of the Hide-and-Seek game

We give a description of the graph corresponding to the actions set of the learner in the HS games with the n-search among k locations and coherence constraints $|z_t(i) - z_t(i + 1)| \leq \kappa, \forall i \in [n]$ for a fixed $\kappa \in [0, k - 1]$.

Definition E.4 (HS Graph)**.** *The graph $G_{k,n,\kappa}$ is a DAG that contains:*

(i) $N := 2 + kn$ vertices arranged into $n + 2$ layers. Layer 0 and Layer $(n + 1)$, each contains only one vertex, respectively labeled s–the source vertex and d–the destination vertex. Each Layer $i \in \{1, \ldots, n\}$ contains k vertices whose labels are ordered from left to right by $(i, 1), (i, 2), \ldots, (i, k)$.

(ii) There are directed edges from vertex s to every vertex in Layer 1 and edges from every vertex in Layer n to vertex d. For $i \in \{1, 2, \ldots, n - 1\}$, there exists an edge connecting vertex (i, j_1) to vertex $(i + 1, j_2)$ if $|j_1 - j_2| \leq \kappa$.

The graph $G_{k,n,\kappa}$ has $E = 2k + (n-1)[k+\kappa(2k-\kappa-1)] = O(nk^2)$ edges and at least $\Omega(\kappa^{n-1})$ paths from s to d. The edges ending at vertex d are the auxiliary edges that are added just to guarantee that all paths end at d; these edges do not represent any intuitive quantity related to the game. For the remaining edges, any edge that ends at the vertex (i, j) represents choosing the location j as the i-th move. In other words, a path starting from s, passing by vertices $(1, j_1), (2, j_2), \dots, (n, j_n)$ and ending at d represents the n-search that chooses location j_1, then moves to location j_2, then moves to location j_3, and so on.

Proposition E.5. *Given k, κ and n, there is a one-to-one mapping between the action set $S_{k,n,\kappa}$ of the learner in the HS game (with n-search among k locations and coherence constraints with parameter κ) and the set of all paths from vertex s to vertex d of the graph $G_{k,n,\kappa}$.*

E.7 Exp3-OE Algorithm and OSMD Algorithm in the CB and HS Games

(i) As stated in Section 8.3.1, the observation graphs in the CB games are non-symmetric and they satisfy Assumption $(A1)$. If we choose $\beta = \beta^*$ as in (118), then β satisfies (114). Moreover, $\beta = O(1/\sqrt{TnE})$; thus, $M = O(E^2\sqrt{TnE})$. From (117), the expected regret of Exp3-OE in this case is bounded by $O\sqrt{Tn(\alpha_{CB}) \ln M \ln(P)}$ (recall that $\alpha_{CB} = kn$ is an upper bound of independence numbers of the observation graphs in the CB games). Therefore, to guarantee that this bound is better than the bound of the OSMD algorithm (that is $\sqrt{2TnE}$), the following inequality needs to hold:

$$
\begin{aligned}
&O\left(\alpha_{CB} \cdot \ln M \ln(P)\right) \leq E \\
\Rightarrow &O\left(nk \cdot \ln\left(E^2\sqrt{TnE}\right) \ln(2^n)\right) \leq nk^2 \\
\Rightarrow &O\left(\ln\left(E^2\sqrt{TnE}\right) \ln(2^n)\right) \leq k \\
\Rightarrow &O\left(n \ln\left(n^3 k^5 \sqrt{T}\right)\right) \leq k.
\end{aligned}
$$

(ii) As stated in Section 8.3.2, the observation graphs in the HS games with condition $(C1)$ are symmetric and do not satisfy Assumption $(A1)$. If we choose $\beta = 1/\sqrt{n\alpha T}$ then by (113), we have that R_T is bounded by $O\left(n\sqrt{\alpha_{HS} T \ln(P)}\right)$ (recall that $\alpha_{HS} = k$ is an upper bound of the independence numbers of the observation graphs in the HS games). Therefore, to guarantee that this bound is better than the bound of the OSMD algorithm in HS games, the following inequality needs to hold:

$$
\begin{aligned}
&O\left(\alpha_{HS} \cdot n \ln(P)\right) \leq E \\
\Rightarrow &O\left(k \cdot n \ln(P)\right) \leq nk^2 \\
\Rightarrow &O\left(\ln(P)\right) \leq k \\
\Rightarrow &O\left(n \ln \kappa\right) \leq k.
\end{aligned}
$$

(iii) Finally, the observation graphs in the HS games with condition $(C2)$ are non-symmetric and do not satisfy Assumption $(A1)$. Therefore, if we choose $\beta = \beta^*$ as in (112), then β satisfies (110). In this case, $\beta = O(1/\sqrt{TnE})$ and $M = O(E^2\sqrt{TnE})$. Therefore, from (111), in this case, R_T is bounded by $O(n\sqrt{T\alpha_{HS} \ln \alpha_{HS} \ln(nM)}$. Therefore, to guarantee that this bound is better than the bound of OSMD (that is, $\sqrt{2TnE}$), the following inequality needs to hold:

$$
O\left(\alpha_{HS} \cdot n \ln nM \ln(P)\right) \leq E
$$

$$\Rightarrow \mathcal{O}\left(nk\ln\left(\kappa^{n}\right)\ln(nE^{2}\sqrt{TnE})\right) \leq nk^{2}$$
$$\Rightarrow \mathcal{O}\left(n\ln\kappa\ln\left(n^{4}k^{5}\sqrt{T}\right)\right) \leq k.$$

Appendix F

Supplementary Materials for Chapter 9 on OSPBand and the Online Bandit CB Game

F.1 Optimization Exploration Distributions of Edge by SDPs

To formulate the problem (9.5)-(9.6) into a SDP, we first observe that for any distribution μ such that the paths set $\mathcal{P}$ is spanned by the support of μ, the matrix $M(\mu)$ always has a fixed number of zero eigenvalues (denoted by K) and this number can be easily computed.[39] Therefore, the problem of maximizing $\lambda^*[M(\mu)]$ is equivalent to maximizing the sum of $K+1$ smallest eigenvalues of $M(\mu)$ which is formulated as:

$$\text{minimize} \quad (K+1)s + \text{Tr}(Z) \tag{119}$$

$$\text{subject to} \quad Z \succeq 0 \tag{120}$$

$$Z + \sum_{i=1}^{P} x_i \cdot \boldsymbol{p}_i \boldsymbol{p}_i^\top + sI_E \succeq 0. \tag{121}$$

Here, $\boldsymbol{x} \in [0,1]^P$ and $r, s \in \mathbb{R}$, $Z \in \mathbb{M}_{E\times E}$ are the variables. I_E is the identity matrix and the notation $X \succeq 0$ indicates that the matrix X is positive semi-definite. This is trivially deduced from the Linear Matrix Inequalities representation of the sum of $K+1$ largest eigenvalues of the matrix (see e.g., Nesterov and Nemirovsky (1994) and Vandenberghe and Boyd (1996)).

F.2 A Regret Lower-bound of OSPBand

Proposition 9.5.1. *There exists an instance of* OSPBand *on a graph, where the number of edges is E and the length of the longest paths is n, such that*

$$\inf_{\textit{strategies}} \sup_{\textit{adversaries}} R_T = \Omega\left(n\sqrt{ET}\right).$$

Proof. This proof follows Audibert, Bubeck, and Lugosi (2014) (Theorem 4.1) regarding the lower-bound of an online combinatorial bandit problem (i.e., the learner's action set is a subset of $\{0,1\}^E$) whose main argument follows the standard lower bounds of the bandit problems (see Cesa-Bianchi and Lugosi (2006)).[40] Intuitively, the instance used in Audibert, Bubeck, and Lugosi (2014) is the problem where players play in

[39] $Rank(M(\mu)) < E$ is the size of the largest linear independent subset of $\mathcal{P}$, which is fixed and only depends on the structure of the layered graph $G_{k,n}$. *Rank-nullity* theorem implies that K is also fixed. We can compute K by computing rank of any particular matrix, say $M(\mu_{\text{uni}})$.

[40] Note also that a similar proof can be found in Audibert and Bubeck (2010) (Theorem 30) but they only focus on the case corresponding to the problem where $n = 1$ and it involves only the oblivious adversaries.

parallel n finite games with $k := E/n$ actions in each game; the learner plays against a $\boldsymbol{p}$-adversary which is parameterized by an action $\boldsymbol{p} \in S$ of the learner ($\boldsymbol{p}$ is chosen randomly and hidden) and defined in such a way that $\boldsymbol{p}$ is the optimal action of the learner against the $\boldsymbol{p}$-adversary.

Given an $n > 0$, $E = k \cdot n$ ($k \geq 2$) and $T \geq E$. Let us consider a OSPBAND on a DAG (multi-graph) G as follows: G contains $n+1$ nodes labeled from 1 (the source node) to $n+1$ (the destination node); for each $i \in [n]$, there are k edges from a node i to node $i+1$ labeled from $[i,1]$ to $[i,k]$ (each edge from i to $i+1$ represents an action in the i-th game). An illustration of the graph with $n = 2, k = 4$ is given in Figure 5. Note that we can always construct a simple graph corresponding to G (each edge in G corresponds to a pair of edges in the simple graph) that have $2E$ edges and the length of the longest path is $2n$ (an illustration is given in Figure 6).

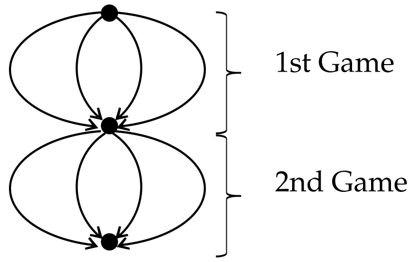


Figure 5: Multigraph with $n = 2, k = 4$.

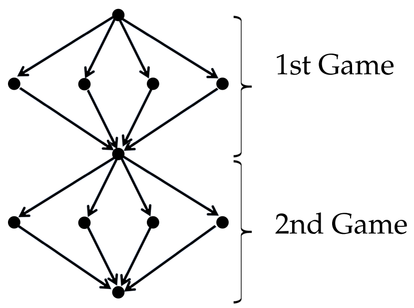


Figure 6: The simple graph corresponding to G with $n = 2, k = 4$.

The action set of the learner in the OSPBAND on G is the set of paths from the source to the destination, that is $S = \left\{\boldsymbol{p} \in \{0,1\}^E : \sum_{j=1}^{k} \boldsymbol{p}[i,j] = 1, \forall i \in [n]\right\}$. Here, $\boldsymbol{p}[i,j] = 1$ if and only if the edge $[i,j]$ (j-th edge from node i to $i+1$) belongs to the path $\boldsymbol{p}$.

Given a path $\boldsymbol{p} \in S$, a $\boldsymbol{p}$-adversary is the adversary who at stage t, samples a loss $\boldsymbol{\ell}_t(e)$ from a Bernoulli distribution with parameter $1/2$ if $e \notin \boldsymbol{p}$ and from a Bernoulli distribution with parameter $1/2-\varepsilon$ if $e \in \boldsymbol{p}$. That is, in expectation, the edges belonging to $\boldsymbol{p}$ has a slightly smaller loss than other edges. We can easily see that this adversary corresponds to the one defined in Theorem 4.1 of Audibert, Bubeck, and Lugosi (2014). Therefore, following Audibert, Bubeck, and Lugosi (2014), we can conclude this proposition. □

F.3 A Regret Lower-bound of Online Bandit CB Games

Proposition 9.5.2. *There exists an instance of the online bandit CB game with* $2n$ *battlefields (*$n \in \mathbb{N}\backslash\{0\}$*) such that in the corresponding* OSPBand, we have

$$\inf_{\text{strategies}} \sup_{\text{adversaries}} R_T = \Omega\left(n\sqrt{T}\right). \tag{9.9}$$

Proof. We consider a CB game having the parameters as follows (they are known to the learner):

1. There are $2n$ battlefields divided into two groups: the *big battlefields* indexed by $\{b_1, b_2, \dots, b_n\}$ where each has a value of 1; the *small battlefields* indexed by $\{s_1, s_2, \dots, s_n\}$ where each has a value of $1/2$.

2. The learner has $k = n$ troops. Thus, her strategy set is

$$S := \{\boldsymbol{z} \in \{1, \dots, n\}^{2n} : \textstyle\sum_{i \in [2n]} z_i = n\}.$$

   As a convention, hereinafter, we denote by z_{b_i} and z_{s_i} the allocations toward battlefields b_i and s_i of a strategy $\boldsymbol{z} \in S$.

3. The tie-breaking rule is in favor of the adversary.

Next, we design a special adversary as follows: the adversary has n^2 troops, she chooses a set $A \subseteq \{1, 2, \dots, n\}$ uniformly at random (A is unknown to the learner). At each stage t, the adversary does the following:

1. allocates 0 troop to each small battlefield $s_i, \forall i \in [n]$,

2. chooses an index $i_t \in \{1, \dots, n\}$ uniformly at random (i_t is kept secretly from the learner),

3. allocates n troops to the big battlefield b_j for any $j \neq i_t$,

4. for the big battlefield b_{i_t}:

$$\text{if } i_t \in A, \begin{cases} \text{allocates } n \text{ troops} & \text{with probability } 1/2 - \varepsilon, \\ \text{allocates } 0 \text{ troop} & \text{with probability } 1/2 + \varepsilon, \end{cases}$$

$$\text{if } i_t \notin A, \begin{cases} \text{allocates } n \text{ troops} & \text{with probability } 1/2 + \varepsilon, \\ \text{allocates } 0 \text{ troop} & \text{with probability } 1/2 - \varepsilon. \end{cases}$$

Note importantly that we desire an adversary who maximizes the learner's regret and may not want to optimize her own payoffs; therefore, the strategy of the adversary can be irrational and she may not use all the troops that she has.

We have some intuitive remarks on the payoffs of the learner against such an adversary as follows:

1. The learner can always guarantee to win all the small battlefields (by allocating 1 troop to each one of them) and loses all big battlefields.[41] In that case, for each $i \in [n]$, $\boldsymbol{\ell}_t(b_i) + \boldsymbol{\ell}_t(s_i) = 1 + 0 = 1$. Therefore, her total loss by playing this strategy is $L_t = n$. Let us call this by the *safe-strategy*.

2. The loss that the learner receives in each battlefield is the same when her allocation there belongs to the set $\{1, 2 \ldots, m\}$.[42] In other words, against the adversary described above, in each battlefield, it only matters to the learner whether she allocates 0 troop or more than 0 troop. Formally, for any strategy of the learner, say $\tilde{z} \in S$, there exists a strategy in the set $S^* = \{z \in \{0,1\}^{2n} : \sum_{i \in [2n]} z_i = n\} \subset S$ yielding a better loss than $\tilde{z}$ against this adversary.

3. The optimal (in expectation) strategy that the learner can do against such an adversary is $z^A = (z_{b_1}, \ldots, z_{b_n}, z_{s_1}, \ldots, z_{s_n}) \in S^*$ where $z_{b_i} = 1$, $z_{s_i} = 0$, $\forall i \in A$ and $z_{b_j} = 0$, $z_{s_j} = 1$, $\forall j \notin A$. We call this *the α^A-strategy*. Intuitively, to obtain this strategy, starting from the safe-strategy, the learner moves one troop from the small battlefield s_i to the big battlefield b_i for any $i \in A$. Note importantly that A is unknown to the learner; therefore, she cannot compute this optimal strategy at the beginning of each stage.

The last remark above comes from the following observation: for any $i \in A$, at each stage, if the learner follows a strategy such that she allocates 1 troop to the big battlefield b_i and 0 troop to the small battlefield s_i (i.e., $\alpha_{b_i} = 1, \alpha_{s_i} = 0$), then the expected losses she receives from these battlefields are

$$\begin{aligned}
&\boldsymbol{\ell}_t(b_i)+\boldsymbol{\ell}_t(s_i)\\
=&\frac{1}{n}\left[\left(\frac{1}{2}-\varepsilon\right)\cdot 1+\left(\frac{1}{2}+\varepsilon\right)\cdot 0+\frac{1}{2}\right]+\frac{n-1}{n}(1+0)\\
=&1-\frac{\varepsilon}{n}.
\end{aligned}$$

Therefore, in the optimal case, by playing the α^A strategy, the learner's total loss is $n - |A|\varepsilon/n$, which is optimal.

On the other hand, for any $j \notin A$, if the learner allocates 1 troop to the big battlefield b_j and 0 troop to the small battlefield s_j (i.e., $\alpha_{b_j} = 1, \alpha_{s_j} = 0$), then the expected losses she receives from these battlefields are:

$$\begin{aligned}
&\boldsymbol{\ell}_t(b_j)+\boldsymbol{\ell}_t(s_j)\\
=&\frac{1}{n}\left[\left(\frac{1}{2}+\varepsilon\right)\cdot 1+\left(\frac{1}{2}+\varepsilon\right)\cdot 0+\frac{1}{2}\right]+\frac{n-1}{n}(1+0)\\
=&1+\frac{\varepsilon}{n}.
\end{aligned}$$

[41] She only has n troops; therefore, by allocating them toward the small battlefields, she gives up all big battlefield (note that if there is a tie $0-0$ at a battlefield, the learner still loses that battlefield).

[42] The learner wins a small battlefield as long as she allocates more than 0 troop. In big battlefields, if the adversary allocates n troops, then the learner always loses regardless of what she does; otherwise, if the adversary allocates 0 then the learner always wins as long as she allocates more than 0.

Therefore, the problem that the learner faces at each stage is equivalent to the case where she has the action set $S' = \{0,1\}^n$ containing all the α^A-strategies corresponding to all $A \subset \{1,\dots,n\}$ and that at stage $t+1$, for each $i \in [n]$, the learner needs to guess whether $i \in A$ based on t incurred loss samples. If she guesses correctly, she reduces ε/n from the loss of the safe-strategy; but if he guesses incorrectly, then her total loss increases by ε/n comparing to that of the safe-strategy.

Following Dani et al. (2008), this is equivalent to the statistic problem of deciding between two Bernoulli distributions whose means differ by $2\varepsilon/n$ based on t samples and the initial prior belief is shared equally (i.e., in the prior, each Bernoulli distribution has 50% to be the actual distribution). In this problem, it holds with probability $\Omega(1)$ that an error is made unless there are more than $t = \Omega\left(\left(\frac{n}{2\varepsilon}\right)^2\right)$ samples (see Dani et al. (2008)). Therefore, we have shown that in expectation, regardless of the learner's decision, the loss she suffers from the pair of battlefields b_i, s_i contributes at least $\Omega\left(\min\left\{T, \left(\frac{n}{2\varepsilon}\right)^2\right\} \cdot \frac{2\varepsilon}{n}\right)$ to the total regret. Taking the summation over all battlefields, the total expected regret is $\Omega\left(\min\left\{2T\varepsilon, \frac{n^2}{2\varepsilon}\right\}\right)$ and we obtain the bound in (9.9) by setting $\varepsilon = n/\sqrt{T}$.

□