ϵ-Minimax Solutions of Statistical Decision Problems*

Andrés Aradillas Fernández† José Blanchet‡ José Luis Montiel Olea†
Chen Qiu† Jörg Stoye† Lezhi Tan‡

Abstract

We present an algorithm for obtaining *ϵ-minimax* solutions of statistical decision problems. We are interested in problems where *i*) the statistician is allowed to choose randomly among I decision rules, and *ii*) the statistical model may have a parameter space with infinitely many elements. The minimax solution of these problems admits a convex programming representation over the $(I-1)$-simplex, and the algorithm herein suggested is a version of *mirror subgradient descent*, initialized with uniform weights and stopped after a finite number of iterations. The resulting iterative procedure is known in the computer science literature as the *Multiplicative Weights* update method, and it is used in algorithmic game theory as a practical tool to find approximate solutions of two-person zero-sum games. We apply the suggested algorithm to different minimax problems in the econometrics literature. An empirical application to the problem of optimally selecting *sites* to maximize the external validity of an experimental policy evaluation illustrates the usefulness of the suggested procedure.

1 Introduction

Under Wald (1950)'s *minimax* criterion different statistical decision rules are ranked based on their worst possible expected loss. Searching for a *minimax-optimal* decision rule—i.e., a rule with the

*We would like to thank Karun Adusumilli, Isaiah Andrews, Tim Armstrong, Giannis Fikioris, Kei Hirano, Nicole Immorlica, Lihua Lei, Charles Manski, Guillaume Pouliot, Brenda Quesada Prallon, Vasilis Syrgkanis, David Shmoys, Sophie Sun, Yiwei Sun, Eva Tardos, Alex Torgovitsky, and Davide Viviano for helpful feedback, comments, and suggestions. We gratefully acknowledge financial support from the NSF under grant SES-2315600. First version: December 5th, 2024.

†Department of Economics, Cornell University.

‡Management Science and Engineering Department, Stanford University.

smallest worst-case expected loss—presents two well-known computational challenges. First, evaluating the worst-case performance of a given decision rule typically requires optimizing a nonlinear function (the *risk* function) over the model's parameter space. Second, minimizing the worst-case expected loss over the decision rules under consideration typically requires optimization over a high-dimensional (and possibly infinite-dimensional) space; for example, all measurable functions that map the data into actions. It is known that obtaining the minimax solution of a decision problem—and sometimes even deciding whether a minimax solution exists—is NP hard in general; see Du and Pardalos (1995); Daskalakis, Skoulakis, and Zampetakis (2021).[1]

We consider a particular class of decision problems in which the decision maker is restricted to choose from a menu of I available decision rules, all of which are assumed to have risk between zero and a known positive constant M. Our motivation is that, while it is always theoretically interesting to look for the *best* overall decision rule, there are situations in which it is equally desirable to *"evaluate the performance of relatively simple statistical decision functions that researchers use in practice"* (Dominitz and Manski, 2024) and choose optimally among them. It is known that if we allow the decision maker to choose randomly among its I options, the corresponding minimax problem can be viewed as a nonlinear convex optimization problem over the $(I-1)$-dimensional simplex (Chamberlain, 2000). The connection to convex programming is helpful, but is not a computational panacea: evaluating the objective function of the convex program associated to the minimax problem could remain computationally costly.[2] For instance, if one were to rely on a textbook *mirror descent* routine for convex problems in the simplex (Nemirovski and Yudin, 1983; Bubeck et al., 2015, Chapter 4.3), one would typically need to make infinitely many evaluations of the objective function (and its subgradient) to guarantee that an optimal solution has been found. If one is willing to measure the complexity of an iterative procedure for convex optimization by

[1]See also Montiel Olea, Prallon, Qiu, Stoye, and Sun (2024a) for an example of a minimax problem that is NP hard but that in practice can be solved to provable (sub)optimality using solvers for linear integer programming.

[2]As we will explain later, the objective function in the convex program is the worst-case expected loss of any given random selection over the I decision rules; which, as discussed above requires optimizing a nonlinear function over the model's parameter space.

the number of calls the procedure makes to the objective function—as it is done in the popular framework of Nemirovski and Yudin (1983)—it is thus desirable to search for algorithms that evaluate the objective function as infrequently as possible.

We argue that it is possible to make substantial progress in solving the general class of statistical decision problems herein considered if, instead of insisting in finding an *exact* minimax solution, we make our goal to find an *approximate* solution to the original problem. In particular, our main result (Theorem 1) shows that if we search for an ϵ-minimax optimal decision rule (Ferguson, 1967, Chapter 1, Definition 4)—i.e., a rule with the smallest worst-case expected loss up to a given additive factor ϵ—then we can provably obtain such a rule by using a *mirror subgradient descent routine* (with negative entropy as a mirror map). The routine is initialized at uniform weights over the I original decision rules, and the step size is set to $\eta \equiv \epsilon/2M^2$. Importantly, we show that it suffices to stop the mirror descent routine after $T \geq 4\ln(I)M^2/\epsilon^2$ epochs. Thus, we obtain a concrete bound on the computational cost of our procedure (in terms of the number of evaluations of the objective function and its subgradient). The required number of iterations scales logarithmically on the number of original decision rules, I, which means that the algorithm is still useful in problems in which I is large.

The algorithm suggested herein is known in the computer science literature as the *Multiplicative Weights* method; see Arora, Hazan, and Kale (2012) for a survey.[3] This method is used in problems where a decision maker chooses randomly among I alternatives repeatedly (an *online decision-making* problem), but after each round he obtains a payoff for all of the I available actions.

Contributions: The main contributions of this paper are as follows:

1. We first present theoretical results showing that the Multiplicative Weights method can be used to obtain an ϵ-minimax solution in the class of statistical decision problems herein considered (which typically involve an infinite parameter space). It is well known that statistical

[3]The specific version of the Multiplicative Weights method that arises from the mirror descent routine is known as the *Hedge algorithm* (Arora et al., 2012, Section 2.1).

decision problems can be formulated as a two-person zero sum games, and Freund and Schapire (1999) have shown how to use the Multiplicative Weights method to approximately solve the mixed extension of a two-person zero-sum game where both players have finitely many pure strategies. However, to the best of our knowledge, the Multiplicative Weights method has not been applied to approximately solve the statistical decision problems herein considered, and instead some other algorithms have been suggested in the literature; see, for example, Filar and Raghavan (1982) and our discussion of related literature below. This is surprising in light of the straightforward connection with two-person zero-sum games and the origins of Multiplicative Weights in iterative dynamics for game play—see the notion of κ-exponential fictitious play in Fudenberg and Levine (1995) and the references to the work of Blume (1993) therein.

2. We illustrate the usefulness of the suggested algorithm by analyzing a simple *binary treatment choice problem with partial identification* based on the work of Stoye (2012). We use this simple example to compare the output of the Multiplicative Weights update method with known exact solutions of two types of minimax problems: minimizing worst-case regret and solving an ex-ante Robust Bayes problem using the class of priors in Giacomini and Kitagawa (2021). Our analysis of the stylized treatment choice problem in Stoye (2012) shows that, with an appropriate selection of the I decision rules, the ϵ-minimax decision rule obtained by the Multiplicative Weights algorithm is very similar to the exact solutions of the minimax problem.

3. Finally, we present an empirical application to the problem of optimally selecting *sites* to maximize the external validity of an experimental policy evaluation. This *site selection problem* has been recently introduced in the work of Gechter, Hirano, Lee, Mahmud, Mondal, Morduch, Ravindran, and Shonchoy (2024). When the policy maker is restricted to select only one site for experimentation, the output of the Multiplicative Weights algorithm is a selection

probability for each of the sites available for experimentation. In the empirical application (we provide further details below), the ϵ-minimax solution is very different to uniform random sampling. This suggests that selecting uniformly at random where to experiment, need not maximize the external validity of an experimental policy evaluation.

Related Literature: Different algorithms have been suggested for approximating the solutions of minimax problems like the ones considered in this paper. Some classical references include Troutt (1978); Filar and Raghavan (1982); Kempthorne (1987); Chamberlain (2000); Elliott, Müller, and Watson (2015). One important difference between our work and this existing literature is that—once a desired approximation error ϵ has been selected, and once the bound M on the risk function has been obtained—there are no further inputs that the user needs to specify in order to run the algorithm. This means that we are explicit about the number of iterations, step size, and also the initial condition. Importantly, we are able to guarantee that, upon termination after finitely many rounds, the algorithm provably generates and ϵ-minimax rule—in the sense of Ferguson (1967))—provided our assumptions are satisfied. This is possible because the rich literature studying the Multiplicative Weights algorithm has explicit performance guarantees for the algorithm at any given iteration.[4]

Relatedly, there is also recent interest in approximating the solution of minimax problems in which the strategies for both the statistician and nature are parameterized via neural networks, with weights that are updated iteratively using versions of what is called *subgradient ascent-descent*; see the recent work of Luedtke, Carone, Simon, and Sofrygin (2020) and also Luedtke, Chung, and Sofrygin (2021). These algorithms where two players use subgradient descent are similar to the approaches used when optimizing Generative Adversarial Neural Networks (GANs); see, for

[4]In this sense, one could say that our work follows closely the literature on *convergence analysis* in convex optimization; see, for example, the work of Nemirovski and Yudin (1983). This means that we try to be as explicit as we can on the computational resources that we can credibly rely on (in our case, an oracle that finds the worst-case point in the parameter space for a given decision rule) and then we try to make use of these computational resources as efficiently as possible (in our case, this means that we attempt to call the available oracle as infrequently as possible to obtain an approximation).

example, Kaji, Manresa, and Pouliot (2023). These subgradient ascent-descent algorithms are also commonly used to approximate the equilibrium of two-person zero-sum games by invoking simultaneous no-regret dynamics; see, for example, Section 3.1 in Lewis and Syrgkanis (2018) and the references therein. Convergence rates for these subgradient ascent-descent algorithms, as well some performance guarantees for a finite number of iterations, are available under some conditions. It is known, however, that the (approximate) stationary points of these gradient ascent-descent algorithms are not necessarily ϵ-minimax strategies. Instead, they are close to what the literature refers to as *local min-max* solutions; see the seminal work of Daskalakis et al. (2021).

Empirical Application: In our main application, we study the *site selection problem* in Gechter et al. (2024) and Egami and Lee (2024). Broadly speaking, a policy maker wishes to experimentally evaluate the effects of a new policy, with the end goal of recommending its implementation on a set of different *sites*. There are two types of sites: *policy-relevant* and *experimental* sites. There are also covariates $X_s \in \mathbb{R}^d$ available for each site. The *site selection problem* asks the following question: if the policy maker can pick at most k experimental sites, what are the sites that optimize external validity?

Two recent papers have provided an answer to this question. Gechter et al. (2024) use an elegant decision-theoretic framework to recommend a nonrandomized selection of sites with the goal of maximizing the average welfare of the policy maker. Montiel Olea et al. (2024a) use the framework of Gechter et al. (2024) to show that, under some conditions, selecting the k-sites with the most representative covariates (in a sense they make precise) is minimax (welfare) regret optimal (restricting the policy maker to consider only nonrandomized selection of sites).

We obtain these two recommendations in the specific context of the selection of candidate *migration corridors* for conducting randomized evaluations in Bangladesh. We assume that a hypothetical policy maker is interested in selecting only one site to evaluate an encouragement design where poor rural households with family members who had migrated to a larger urban destination receive a 30–45 minute training about how to register and use a mobile banking service to send instant remittances back home; see Lee, Morduch, Ravindran, Shonchoy, and Zaman (2021) and Gechter et al. (2024).

The site selected based on average welfare and minimax (welfare) regret is not the same. We consider both of them—along with an experimental site where the encouragement design of interest has already been experimentally evaluated—to give the policy maker three concrete decision rules ($I = 3$). We then let the policy maker choose randomly over them to determine where to experiment in order to minimize worst-case regret. We note that although, in general, randomly choosing where to experiment could be viewed as contrived, the randomized selection of experimental sites is typically thought of as the first-best in applied work. For instance, Duflo, Glennerster, and Kremer (2007) note that *"the external validity of randomized evaluations for a given population (say, the population of a country) would be maximized by randomly selecting sites and, within these sites, by randomly selecting treatment and comparison groups."*[5] In this application, the parameter space consists of functions that control the treatment effect heterogeneity across sites. We impose a Lipschitz constraint on these functions (we provide details later), and consider the worst-case regret over this space.

The two main lessons from our application are the following. First, choosing uniformly at random where to experiment does not tend to be ϵ-minimax optimal. Instead, the ϵ-minimax solution seems to adjust the probability of sampling a site based on its baseline covariates. For instance, in our application the experimental site whose covariates are closest (on average) to the covariates of the policy-relevant sites is sampled with the highest probability. Second, there seem to be some cases (for example, when one experimental site is closest to each of the policy-relevant sites) in which the ϵ-minimax solution places almost probability one such a site. This suggests that maximizing the external validity of a randomized evaluation need not be maximized by randomly selecting sites. Instead, it is possible that using baseline covariates, experimenting on the most representative site

[5]Although they remark that this is almost never done because *"randomized evaluations are typically performed in "convenience" samples, with specific populations."*

could be useful for policy purposes. Our approach thus provides an algorithm for deciding how to *randomly* select sites to optimize external validity, taking into account information about baseline covariates.

OUTLINE: The rest of the paper is organized as follows. Section 2 introduces notation, main assumptions, and presents the convex programming representation of the minimax problems herein analyzed. Section 3 defines an ϵ-minimax decision rule and presents the algorithm. Section 4 applies the algorithm to three illustrative examples that involve solving treatment choice problems with partial identification. Our algorithm is the used to solve for ϵ-minimax regret optimal rules; but we also argue that it can be applied to solve other minimax problems, such as (ex-ante) Robust Bayes analysis with the priors suggested by Giacomini and Kitagawa (2021). Section 5 presents the main application. Section 6 discusses some extensions. Section 7 concludes.

2 Minimax Problems

2.1 Notation

A decision maker must choose an action a that belongs to some set $\mathcal{A}$. Prior to choosing the action, he observes data: the realization of a random variable X taking values in a set $\mathcal{X}$. A data-driven choice of action is summarized by a *decision rule*: a mapping from data to actions, which is herein denoted by the function $d : \mathcal{X} \to \mathcal{A}$.

It is common to allow the decision maker to consider every (measurable) function d as a decision rule. However, we restrict our analysis to the case in which the decision maker only considers I decision rules that belong to the finite set $\mathcal{D} \equiv \{d_1, \ldots, d_I\}$. Our motivation is that, while it is always theoretically interesting to look for the *best* overall decision rule, there are situations in which it is equally desirable to *"evaluate the performance of relatively simple statistical decision functions that researchers use in practice"* (Dominitz and Manski, 2024) and choose optimally among them.

An important aspect of our analysis is that we allow the decision maker to choose randomly from the set of decision rules $\mathcal{D}$ and we represent such a random choice by an element in the $I-1$ simplex:

$$\Delta(\mathcal{D}) \equiv \left\{ (p_1, ..., p_I) \in \mathbb{R}^I \;\middle|\; \sum_{i=1}^{I} p_i = 1, p_i \geq 0 \right\}.$$

It is well known that allowing the decision maker to choose randomly is usually to his advantage.[6] Moreover, there are two additional reasons why we would like to allow for the possibility of *randomization*. The first one is that in the main application we will consider in the paper (the *site selection* problem described in Section 5), the random choice of experimental sites is viewed as the default practice in applied work. The second reason is that, as we will explain in Section 3 (Remark

[6] Consider a "matching pennies" game with two players, each with two actions: left and right. Suppose that column player gets M when matched and $-M$ when unmatched. If neither player is allowed to choose actions randomly, the worst-case payoff obtained by the column player is $-M$ regardless of the action chosen. If the column player can randomize, but the row player cannot, the worst-case payoff for the column player if he chooses each action at random with probability $1/2$ is zero.

4), allowing for random choice of actions can reduce the computational burden of selecting a good decision rule.

A *risk function* is used to summarize the performance of each decision rule $d_i \in \mathcal{D}$. This performance is contingent on the data generating process, which we parameterize by an element θ belonging to some space Θ. Thus, we write the risk function of each decision rule $d \in \mathcal{D}$ as a mapping $R : \mathcal{D} \times \Theta \to \mathbb{R}$. We refer to θ as a parameter, and to Θ as the parameter space. We are particularly interested in the case in which Θ is an infinite set; for example, when Θ equals all of $\mathbb{R}^d$. We also want to allow for the possibility that each element in the parameter space is an infinite dimensional object (for example, when θ itself is a function). We impose the following assumption on the risk function:

Assumption 1. There exists a known constant $0 < M < \infty$ such that for any $d \in \mathcal{D}$ and $\theta \in \Theta$, $0 \leq R(d, \theta) \leq M$.

In Section 4 we explain how this assumption can be verified for each of the illustrative examples we consider. We view Assumption 1 as a minimal regularity condition for the minimax problem to be well-behaved. We also note that the assumption holds if each of the I decision rules under consideration has a finite worst-case risk.

In a slight abuse of notation, we extend the original domain of the risk function—which was defined over decision rules in $\mathcal{D}$—to all possible random selections in $\Delta(\mathcal{D})$. We do this by defining, for any $p \in \Delta(\mathcal{D})$ and $\theta \in \Theta$, the function:

$$R(p, \theta) \equiv \sum_{i=1}^{I} p_i R(d_i, \theta). \tag{1}$$

We view a decision problem as a triplet $(\mathcal{D}, \Theta, R(\cdot, \cdot))$ and we define the "minimax value" of the decision problem as the scalar

$$\bar{v} \equiv \inf_{p \in \Delta(\mathcal{D})} \sup_{\theta \in \Theta} R(p, \theta). \tag{2}$$

A random selection $p^* \in \Delta(\mathcal{D})$ is said to be a *minimax* decision rule if

$$\sup_{\theta \in \Theta} R(p^\star, \theta) = \bar{v}. \tag{3}$$

The use of the minimax criterion as a solution concept in statistical decision problems is traditional, dating back to Wald (1950). An axiomatic foundation of the minimax criterion in decision problems under uncertainty is presented in Gilboa and Schmeidler (1989). Manski (2021) argues that the primary challenge to use the minimax criterion and Wald (1950)'s statistical decision theory is computational.

2.2 Exact minimax solutions via convex programming

We first show that the minimax solution of the decision problems considered in this paper can be computed via convex programming. This observation is based on an analogous result in Chamberlain (2000); see Equation 5, p. 630, and the discussion therein. The argument is as follows. For $p \in \Delta(\mathcal{D})$, define the nonlinear function

$$f(p) \equiv \sup_{\theta \in \Theta} R(p, \theta). \tag{4}$$

This function is the upper envelope—over all possible values in the parameter space—of the risk of p.

Lemma 1. *The function $f : \Delta(\mathcal{D}) \to \mathbb{R}$ is convex. Furthermore, fix an arbitrary $p_0 \in \Delta(\mathcal{D})$. If there exists $\theta_0 \in \Theta$ such that $R(p_0, \theta_0) = f(p_0)$, then the vector g_0 in $\mathbb{R}^I$ given by*

$$g_0 \equiv (R(d_1, \theta_0), \ldots, R(d_I, \theta_0))^\top. \tag{5}$$

is a subgradient of f at p_0.[7]

[7] If $f : \Delta(\mathcal{D}) \to \mathbb{R}$ is convex, a vector g_0 is said to be a subgradient of f at a point p_0 if $f(p) \geq f(p_0) + g_0^\top (p - p_0), \forall p \in$

Proof. The result follows Chamberlain (2000), but we provide a proof for the sake of exposition. See Appendix A.1. □

Lemma 1 shows that solving the minimax problem in (2) can be viewed as a nonlinear convex program over the $(I-1)$ simplex. A popular routine for solving this type of convex optimization problems is the mirror descent method of Nemirovski and Yudin (1983). We briefly describe a textbook version of this routine (Bubeck et al., 2015, Section 4.3, p. 301). To make sure that the subgradient is well defined, we make the following assumption.

Assumption 2. For any $p \in \Delta(\mathcal{D})$, there exists $\theta_p \in \Theta$ such that

$$\sum_{i=1}^{I} p_i R(d_i, \theta_p) = \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_i R(d_i, \theta).$$

The assumption says that for any $p \in \Delta(\mathcal{D})$ it is possible to find an element θ_p such that $R(p, \theta_p) = f(p)$. This means that there is an algorithm that is capable to i) evaluate the function $f(p)$ and to ii) find a maximizer that evaluates to $f(p)$. Assumption 2 also imposes substantial regularity to search for a "worst-case" parameter in the parameter space, and requires that worst-case risk is attained. Later we discuss the extent to which this assumption can be relaxed, by requiring that we can get a δ-approximation to $f(p)$. See Remark 3.

Mirror Descent Routine: The following is a typical mirror descent routine for finding the minimum of (4). The routine is taken from (Bubeck et al., 2015, Section 4.3, p. 301) and simply adjusts the notation to our problem. Initialize $w_0 \in \mathbb{R}^I$ to be the vector that contains ones in all of its entries. We will denote such vector by $\mathbf{1}$. Fix a step-size $\eta > 0$. For every $t \in \mathbb{N}$:

$\Delta(\mathcal{D})$. See pp. Rockafellar (1970) 214-215.

1. Set

$$\begin{aligned}\phi_t &\equiv \mathbf{1}^\top w_{t-1}\\ &= \sum_{i=1}^{I} w_{i,t-1}.\end{aligned}$$

2. Obtain a point $p_t \in \Delta(\mathcal{D})$ by computing for each $= 1, \ldots, I$:

$$p_{i,t} \equiv \frac{w_{i,t-1}}{\phi_t}.$$

3. Given $p_t = (p_{1,t}, ..., p_{I,t})^\top \in \Delta(\mathcal{D})$, find $\theta_t \in \Theta$ such that

$$\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) = \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{i,t} R(d_i, \theta).$$

   Such a point in the parameter space exists by Assumption 2.

4. Define the vector

$$g_t \equiv (R(d_1, \theta_t), \ldots, R(d_I, \theta_t))^\top.$$

   By Lemma 1, this vector is a subgradient of f at p_t if Assumption 2 holds.

5. Update the weights w_{t-1} with the coordinate-by-coordinate multiplicative rule:

$$w_{i,t} \equiv w_{i,t-1} \cdot \exp\left(-\eta\, g_{i,t}\right).$$

Under this mirror descent routine, the vector p_t gets updated as $p_{i,t+1} = w_{i,t}/\phi_{t+1}$.

We note that the connection to convex programming is helpful, but should not be viewed as a computational panacea. Evaluating the objective function of the convex program and its subgradient could remain computationally costly. Verifying Assumption 2 requires optimizing a nonlinear function over the model's parameter space. As noted above, the solution of such nonlinear optimization problem is used to evaluate the subgradient of the objective function $f(p)$. It is known that the rate of convergence of mirror descent for convex problems in the simplex improves over regular subgradient descent (Bubeck et al., 2015, Section 4.3). However, in order to guarantee that a minimax solution has been found one would typically need to run the routine above for infinitely many epochs. This could be computational costly if one needs to fulfill Assumption 2.

Moreover, the use of (nonlinear) convex programming to find exact minimax solutions of general decision problems might mean that, in some cases, one incurs in a higher computational cost than needed. For example, there are some classical results in the game theory literature (Dantzig, 1951; Adler, 2013; Owen, 2013, Section III.1, p. 36) that show that when Θ has finitely many elements, it is possible to express the minimax problem in (2) as a linear program. Proposition 1 in Section 6 shows that every minimax problem where the statistician chooses randomly among I alternatives can be expressed as a linear program in $I + 1$ variables with as many constraints as elements in Θ (even if Θ has infinitely many elements). This shows that, in some cases, it might be more efficient to use this linear programming representation.

3 Approximate Solutions for Minimax Problems

Computing the exact minimax rule of a decision problem could be computationally costly. This section presents a definition of an *approximate* minimax solution, and an off-the-shelf implementation of mirror descent—used routinely in different areas of computer science and machine learning—to finding it.

3.1 ϵ-Minimax Decision Rules

Definition 1. [Ferguson (1967), p. 44] A random selection $p^\star_\epsilon \in \Delta(\mathcal{D})$ is an "ϵ-minimax" decision rule for the decision problem $(\mathcal{D}, \Theta, R(\cdot,\cdot))$ if

$$\sup_{\theta\in\Theta} R(p^\star_\epsilon, \theta) \leq \inf_{p\in\Delta(\mathcal{D})} \sup_{\theta\in\Theta} R(p,\theta) + \epsilon = \bar{v} + \epsilon.$$

We note that the risk of an ϵ-minimax decision rule is smaller—up to an additive factor of size ϵ—than the worst-case risk of any other decision rule. That is:

$$R(p^\star_\epsilon, \theta) \leq \sup_{\theta\in\Theta} R(p,\theta) + \epsilon, \; \forall \theta \in \Theta, \; \forall p \in \Delta(\mathcal{D}).$$

The definition of a minimax decision rule further implies that

$$\bar{v} \leq \sup_{\theta\in\Theta} R(p^\star_\epsilon, \theta) \leq \bar{v} + \epsilon.$$

3.2 Multiplicative Weights Algorithm for finding ϵ-Minimax Rules

In this section, we show that running the mirror descent routine described in Section 2.2 can be used to provably find ϵ-minimax solutions for the decision problems herein considered. To be more explicit, consider the following pseudocode for mirror descent, but stopped after T epochs.

Our concrete suggestion is to set the step size to $\eta \equiv \epsilon/2M^2$ and to stop the routine after $T \geq 4\ln(I)M^2/\epsilon^2$ epochs. This routine is known as the *Multiplicative Weights* update algorithm, and it is a popular algorithm in computer science, that has found different applications in machine learning; see Arora et al. (2012). The specific version of the Multiplicative Weights algorithm used in this paper uses an exponential function of each of the coordinates of the gradient to update the weights. This version of Multiplicative Weights is commonly referred to as the *Hedge* algorithm. See Section 2.1 in Arora et al. (2012).

Algorithm 1 Mirror Descent, stopped after T epochs (a.k.a. *Multiplicative Weights*).

Input: Step-size $\eta > 0$; and number of epochs $T \in \mathbb{N}$.
Initialize $w_0 \in \mathbb{R}^I$ by setting $w_{i,0} = 1$ for all $i \in \{1, \dots, I\}$.
for $t = 1, 2, \dots$ **do**
    Compute $\phi_t := \sum_{i=1}^{I} w_{i,t-1}$
    For each $i \in \{1, \dots, I\}$, compute

$$p_{i,t} := \frac{w_{i,t-1}}{\phi_t}$$

    Find $\theta_t \in \Theta$ such that

$$\theta_t := \arg\sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{i,t} R(d_i, \theta)$$

    Define the vector

$$g_t := (R(d_1, \theta_t), \dots, R(d_I, \theta_t))^\top$$

    Consider the multiplicative weights update:

$$w_{i,t} := w_{i,t-1} \cdot \exp(-\eta \cdot g_{i,t})$$

end for

For any nonnegative real number x, let $\lceil x \rceil$ denote the "ceiling function"; that is smallest integer larger than x. Our main result is the following.

Theorem 1. *Suppose Assumptions 1-2 hold. Furthermore, assume there exists a minimax decision rule $p^\star \in \Delta(\mathcal{D})$. If $\epsilon \leq M$, $\eta \equiv \epsilon/2M^2$, and $T \equiv \lceil 4M^2 \ln(I)/\epsilon^2 \rceil$, then the random choice of decision rules that assigns probability*

$$p_i^\epsilon \equiv \frac{1}{T} \sum_{t=1}^{T} p_{i,t}$$

to each decision rule d_i—where p_t corresponds to the t-th iteration of the mirror descent routine in Algorithm 1—is ϵ-minimax in the sense of Definition 1. Moreover

$$\bar{v}^\epsilon \equiv \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right),$$

where θ_t corresponds to "nature's best response" in the t-th iteration of the mirror descent routine *in Algorithm 1, is an ϵ-approximation to $\bar{v}$: that is, $\bar{v} \leq \bar{v}^{\epsilon} \leq \bar{v} + \epsilon$.*

Proof. See Appendix A.2 $\square$

Theorem 1 presents a concrete computational strategy to approximately solve the statistical decision problems considered in this paper. The only "tuning" parameter that needs to be chosen is ϵ, which controls the approximation error. We note that in cases where is difficult to commit to a value of ϵ explicitly, one can *solve* for the value of ϵ if there is a specific target for the runtime of the algorithm and we know the time it takes for each iteration to run.

Theorem 1 shows that the average value of p_t (over the T rounds) provides an ϵ-minimax rule. The theorem also shows that the average worst-case payoff obtained in each round is an ϵ-approximation to the minimax value of the decision problem.

We make some remarks about Theorem 1.

Remark 1 (Rate optimality). It is known that the rate $\ln(I)/\epsilon^2$ obtained in Theorem 1 cannot be improved upon; see Section 4 in Arora et al. (2012) and also Gravin, Peres, and Sivan (2016). We note, however, that the number of epochs in Theorem 1 can be lowered by almost a factor of two. In particular, Theorem 2 in Appendix A.4 shows that adopting the proof strategy of Freund and Schapire (1999) allows to establish a result analogous to that of Theorem 1, but with $\eta \equiv \ln\left(1 + \frac{\epsilon}{M}\right)/M$ and $T \equiv 2\lceil M(M+\epsilon)\ln(I)/\epsilon^2 \rceil$.

Remark 2 (Finite Θ). The algorithm described above is a cheaper alternative to finding the exact solution of a convex program. Even when Θ has J elements, obtaining an approximate minimax solution could be computationally cheaper relative to obtaining an exact minimax solution via the linear program in Proposition 1. As discussed in Remark 5, the computational cost of using the fastest solver for linear program to find an exact minimax solution is of order $(1+J+I)^{2.055}$ time. We note that Algorithm 1 makes $4M^2\lceil \ln(I) \rceil/\epsilon^2$ calls to nature's oracle. Suppose that the runtime of the oracle is $r(I, J)$. In each round, the algorithm evaluates the risk of the I actions available to the decision maker. Thus, the runtime of the algorithm is of order

$$M^2 I \ln(I) r(I, J)/\epsilon^2.$$

If the calls to the oracle that computes nature's best response are not expensive, and if M/ϵ^2 is not too large, the time needed in order to compute the approximate solution to the minimax problem could be smaller than that time needed to obtain the exact solution.

Remark 3 (Approximate Nature's best response)*.* It is possible to extend the proof to the case that θ_t is not the exact best response, but an approximate one. More precisely, consider θ_t^δ such that

$$\left(\sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{i,t} R(d_i, \theta) \right) - \delta \leq \sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t^\delta) \leq \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{i,t} R(d_i, \theta).$$

This extension can be (roughly) completed by doing the following adjustments to the proof.

Firstly, in Step 2, we say that

$$\sup_{\theta \in \Theta} \sum_{i=1}^{I} \tilde{p}_i^\star R(d_i, \theta) \leq \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right).$$

We can change this to

$$\sup_{\theta \in \Theta} \sum_{i=1}^{I} \tilde{p}_i^\star R(d_i, \theta) \leq \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t^\delta) + \delta \right).$$

Then, by Step 1, this is bounded by above by

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t^\delta) + \frac{\epsilon}{2} + \frac{\ln(I)}{T} \left(\frac{4M^2}{\epsilon} \right) + \delta.$$

Then, subbing in to the previous Step 2, the left hand side is bounded by above by

$$\bar{v} + \frac{\epsilon}{2} + \frac{\ln(I)}{T}\left(\frac{4M^2}{\epsilon}\right) + \delta.$$

Choosing T as we have done gives an $\epsilon + \delta$ approximation.

Remark 4 (Minimax Solution without randomization). Finally, note that even if one were interested in computing the minimax optimal rule among $\{d_1, \ldots, d_I\}$, one would need I calls to the oracle (one for computing the worst-case performance of each rule). Surprisingly, the ϵ-minimax solution among randomized rules calls the oracle $\lceil 4M^2 \ln(I)/\epsilon^2 \rceil$ times. When I is large, the difference could be substantial.

4 Illustrative Examples

4.1 ϵ-Minimax Regret in a treatment choice problem with partial identification

Consider the following example taken from Stoye (2012) and Yata (2021). A policy maker uses experimental data to decide whether to implement a new policy in a target population of interest. The treatment effect of action $a = 1$ is $\mu^* \in \mathbb{R}$, while the effect of action $a = 0$ is normalized to be equal to 0. Thus, the policy maker's expected payoff equals $W(a, \mu^*) \equiv a \cdot \mu^*$.

The data available to the policy maker is an estimated treatment effect, $\hat{\mu}$, for the experimental population. The policy maker assumes that

$$\hat{\mu} \sim N(\mu, \sigma^2), \tag{6}$$

where $\sigma > 0$ is known and where $\mu \in \mathbb{R}$ is the true effect of the policy in the population where the experiment was conducted. The policy maker is concerned about the external validity of the experiment at hand. This is captured by allowing the effect of the policy in the experimental population (μ) to be different from the effect in the target population (μ^*). The policy maker is willing to work under the assumption that $|\mu^* - \mu| \leq k$ for some known $k \geq 0$. In this example, $\theta = (\mu, \mu^*)^\top$ and $\Theta \equiv \{(\mu, \mu^*) \in \mathbb{R}^2 \mid |\mu - \mu^*| \leq k\} \subseteq \mathbb{R}^2$.

A decision rule for the policy maker is a mapping $d : \mathbb{R} \to [0, 1]$ from the observed experimental data (6) to an action $a \in [0, 1]$. The action is interpreted as the fraction of the target population that will be treated. Consider the regret loss associated to $W(a, \mu^*)$ given by $L(a, \theta) \equiv \mu^*[\mathbf{1}\{\mu^* \geq 0\} - a]$. Define the risk function

$$R(d, \theta) \equiv \mathbb{E}_\theta[L(d, \theta)].$$

Exact Minimax Solution Over all Decision Rules: Let $\mathcal{D}^*$ denote the set of all decision rules. Stoye (2012) derived a solution to the minimax (regret) problem

$$\inf_{d \in \mathcal{D}^*} \sup_{\theta \in \Theta} R(d, \theta), \tag{7}$$

as a function of (σ^2, k). Stoye (2012) showed that when $k \geq \sqrt{\pi/2}\sigma$, Equation (7) equals $k/2$. Montiel Olea, Qiu, and Stoye (2024b) further showed that when $k \geq \sqrt{\pi/2}\sigma$ there are infinitely many minimax-regret optimal rules. One such solution takes the form

$$d^\star_{MQS}(\hat{\mu}) = \begin{cases} 0, & \hat{\mu} < -\rho^\star, \\ \frac{\hat{\mu}+\rho^\star}{2\rho^\star}, & -\rho^* \leq \hat{\mu} \leq \rho^* \\ 1, & \hat{\mu} > \rho^*, \end{cases},$$

where $\rho^* \in (0, k)$ uniquely solves the nonlinear equation:

$$\left(\frac{1}{2k}\right)\rho^* - \frac{1}{2} + \Phi\left(-\frac{\rho^*}{\sigma}\right) = 0, \tag{8}$$

see Theorem 3 in Montiel Olea et al. (2024b).

Approximate Minimax Regret Solution over a Class of Threshold Rules: Suppose that instead of considering all decision rules, we focus on a class $\mathcal{D} \subset \mathcal{D}^*$ that contains only "threshold" rules; that is, decision rules of the form

$$d_i(\hat{\mu}) \equiv \mathbf{1}\{\hat{\mu} \geq c_i\},$$

where $c_i \in \mathbb{R}$. For concreteness, we consider 500 different values for c_i equally spaced in the interval $[-k, k]$.

Algebra shows that, in this example, the largest worst-case risk among all threshold rules in $\mathcal{D}$ is bounded above by $M \equiv \sigma \max_{x \geq 0} x\Phi\left((2k/\sigma) - x\right)$, where $\Phi(\cdot)$ denotes the Normal c.d.f.[8] Since the expected loss is nonnegative, Assumption 1 is satisfied.

We can also show that, for a given $p \in \Delta(\mathcal{D})$, the values $(\mu, \mu^*) \in \Theta$ that verify Assumption 2 can be obtained by solving three optimization problems. Define the parameter μ^*_+ to be the solution of the problem

$$\max_{\mu^* \geq 0} \mu^* \left(\sum_{i=1}^{I} p_i \Phi\left(\frac{c_i - \mu^*}{\sigma} + \frac{k}{\sigma} \right) \right),$$

and $\mu_+ \equiv \mu^*_+ - k$. Define the parameter μ^*_- to be the solution of the problem

$$\max_{\mu^* \leq 0} -\mu^* \left(\sum_{i=1}^{I} p_i \Phi\left(\frac{\mu^* - c_i}{\sigma} + \frac{k}{\sigma} \right) \right),$$

and $\mu_- = \mu^*_- + k$. Set θ_p to be the maximizer of

$$\{R(p, \mu_+, \mu^*_+), R(p, \mu_-, \mu^*_-)\}.$$

Since we have verified Assumption 1 and 2, we proceed to applying Algorithm 1. We consider

[8]The formula corresponds to the worst-case risk of the rule that uses the threshold $c_i = k$ (or -k).

the case in which $\sigma = 1$ and $k = 2$. The value of $M = 2.5294$. We set $\epsilon = .1$. The number of epochs in Theorem 1 then becomes

$$T = 4\lceil M^2 \ln(I)/\epsilon^2 \rceil = 15,905.$$

The runtime of Algorithm 1 is 57.93 seconds (on a personal MacBook Pro 2019 @2.4GHz 8 Core-Intel Core i9). Figure (2) presents a comparison of d^*_{MQS} and the ϵ-minimax rule. The value of $\bar{v}^\epsilon$ is 1.0017.

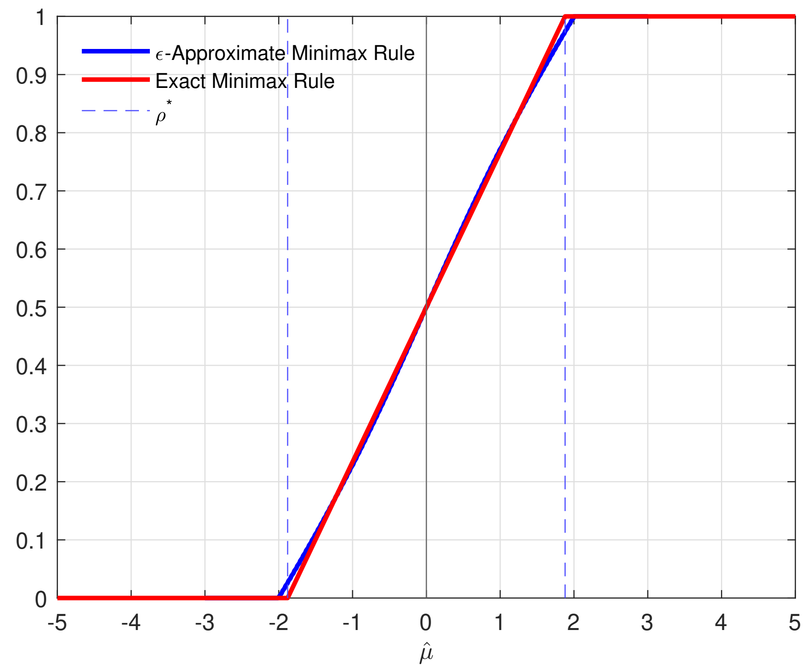

Figure 1: ϵ-Minimax Decision Rule via the Multiplicative Weights algorithm. The graph is generated using $\sigma = 1$, $k = 2$. The value of ρ^* in Equation 8 is 1.8797.

4.2 ϵ-Robust Bayes rules in a treatment choice problem with partial identification

Consider the same example as in Section 4.1, but instead of focusing on minimax-regret optimality as in Stoye (2012), we are interested in computing ex-ante Robust Bayes rules as in Aradillas Fernández, Montiel Olea, Qiu, Stoye, and Tinda (2024).

Let π be a prior over $(\mu, \mu^\star)$. We are interested in obtaining the rule that minimizes worst-case expected risk over the class of priors suggested by Giacomini and Kitagawa (2021). We will denote this class of priors by Γ. Broadly speaking, the priors in this class fix a marginal prior over μ, but allow for arbitrary priors over $\mu^*|\mu$ (as long as the joint distribution over (μ, μ^*) is supported on Θ). For this example, we will first consider the "two-point prior" for μ analyzed in Aradillas Fernández et al. (2024). That is, we assume that the prior of μ is supported on the set $\mathcal{M} = \{-\bar{\mu}, \bar{\mu}\}$. We first assume that the policy maker has a discrete uniform prior π_μ on $\mathcal{M}$, meaning that $\pi_\mu(\mu = \bar{\mu}) = \pi_\mu(\mu = -\bar{\mu}) = 1/2$.

Just as we did in Section 4.1, we consider the regret loss $L(a, \theta) \equiv \mu^\star[\mathbf{1}\{\mu^\star \geq 0\} - a]$ and the risk function

$$R(d, \theta) \equiv \mathbb{E}_\theta[L(d, \theta)].$$

However, we are now interested in the average ("Bayesian") risk of a decision rule; which we define as

$$r(d, \pi) \equiv E_\pi[R(d, \theta)].$$

Let $\mathcal{D}^*$ be the set of all decision rules. The minimax problem of interest is thus

$$\inf_{d \in \mathcal{D}^*} \sup_{\pi \in \Gamma} r(d, \pi). \tag{9}$$

We follow the literature and refer to any decision rule that solves this problem as either ex-ante Γ-minimax or ex-ante Robust Bayes.

Aradillas Fernández et al. (2024) showed that, under some conditions, the problem in Equation (9) for the two-point priors on μ described before has infinitely many solutions. One such solution takes the form

$$d^{\star}(\hat{\mu}) = \begin{cases} 0, & \hat{\mu} < -\frac{\sigma^2\rho^{\star}}{\bar{\mu}} \\ \frac{\bar{\mu}\hat{\mu}+\sigma^2\rho^{\star}}{2\sigma^2\rho^{\star}}, & -\frac{\sigma^2\rho^{\star}}{\bar{\mu}} \leq \hat{\mu} \leq \frac{\sigma^2\rho^{\star}}{\bar{\mu}} \\ 1, & \hat{\mu} > \frac{\sigma^2\rho^{\star}}{\bar{\mu}} \end{cases},$$

where $\rho^{\star}$ uniquely solves

$$\int_0^1 \Phi\left(\frac{2\rho^{\star}x - \rho^{\star} - (\bar{\mu}/\sigma^2)}{\bar{\mu}/\sigma}\right) dx = \frac{-\bar{\mu}+k}{2k}.$$

We compare this Γ-minimax optimal rule with the ϵ-approximation obtained via the Multiplicative Weights algorithm. We again consider the class $\mathcal{D}$ of decision rules of the form

$$d_i = \mathbf{1}\{\hat{\mu} \geq c_i\},$$

where $c_i \in \mathbb{R}$. We again start with an equally spaced grid of 500 points over $[-k, k]$.

In order to apply the Multiplicative Weights Update algorithm we extend the Bayes risk $r(d, \pi)$ to any element $p \in \Delta(\mathcal{D})$ by defining

$$r(p, \pi) \equiv \sum_{i=1}^{I} p_i r(d_i, \pi) = \sum_{i=1}^{I} p_i \mathbb{E}_{\pi}\left[R(d_i, \mu, \mu^{\star})\right] = \mathbb{E}_{\pi}\left[\sum_{i=1}^{I} p_i R(d_i, \mu, \mu^{\star})\right].$$

We note that Assumption 1 is satisfied with the same M as in Subsection 4.1. In order to verify Assumption 2, we note that takes the results in Giacomini and Kitagawa (2021) show that

$$\sup_{\pi\in\Gamma} \mathbb{E}_{\pi}\left[\sum_{i=1}^{I} p_i R(d_i, \mu, \mu^{\star})\right]$$

equals

$$\mathbb{E}_{\pi_{\mu}}\left[\sup_{\mu^{\star}\in[\mu-k,\mu+k]} \sum_{i=1}^{I} p_i R(d_i, \mu, \mu^{\star})\right] \equiv \mathbb{E}_{\pi_{\mu}}\left[\bar{\Lambda}(\mu, p_1, ..., p_I)\right].$$

This relation immediately gives the prior $\pi \in \Gamma$ associated to the worst-case Bayes risk of any vector $p \in \Delta(\mathcal{D})$. In particular, the subgradient used in the updates is

$$g_t^i = \pi_\mu \cdot R(d_i, \bar{\mu}, \bar{\mu}_t^\star) + \pi_\mu \cdot R(d_i, -\bar{\mu}, (-\bar{\mu})_t^\star),$$

where $\bar{\mu}_t^\star$ and $(-\bar{\mu})_t^\star$ are the corresponding values of $\mu^\star$ for $\mu = \bar{\mu}$ and $\mu = -\bar{\mu}$, that solve

$$\bar{\Lambda}(\mu, p_1, ..., p_I) \equiv \sup_{\mu^\star \in [\mu - k, \mu + k]} \sum_{i=1}^{I} p_i R(d_i, \mu, \mu^\star).$$

We can show that the solutions of μ^* (as a function of μ) are given by

$$\mu^\star = \begin{cases} \mu + k, & \frac{\mu+k}{2k} \geq \sum_{i=1}^{I} p_i \Phi\left(-\frac{c_i - \mu}{\sigma}\right) \\ \mu - k, & \frac{\mu+k}{2k} < \sum_{i=1}^{I} p_i \Phi\left(-\frac{c_i - \mu}{\sigma}\right) \end{cases}$$

We consider the case in which $\sigma = 1$, $k = 2$, and $\bar{\mu} = 0.5$. We set $\epsilon = 0.1$. The number of epochs in Theorem 1 is again

$$T = 4\lceil M^2 \ln(I)/\epsilon^2 \rceil = 15,905.$$

The algorithm runs for $T = 15,905$ iterations and finishes in under 50 seconds (on a personal ASUS Vivobook Pro 15 @ 2.5GHz Intel Core Ultra 9 185H).

Figure 2 shows the true solution versus its ϵ-approximate solution. Qualitatively, the two are very close. In fact, the minimax values are close as well, with the ϵ-approximation having a minimax value of 0.9376 and the true solution having a minimax value of 0.9375. Note that here, the term referred to as $\rho^\star$-adjusted is

$$\rho^\star\text{-adjusted} = \frac{\sigma^2 \rho^\star}{\bar{\mu}}. \tag{10}$$

We also consider a 500-point uniform prior (i.e., $\pi_\mu = 1/500$) supported on an equally spaced grid within $[-k, k]$. In this case, no analytical solution is available. We still keep $\sigma = 1$ and

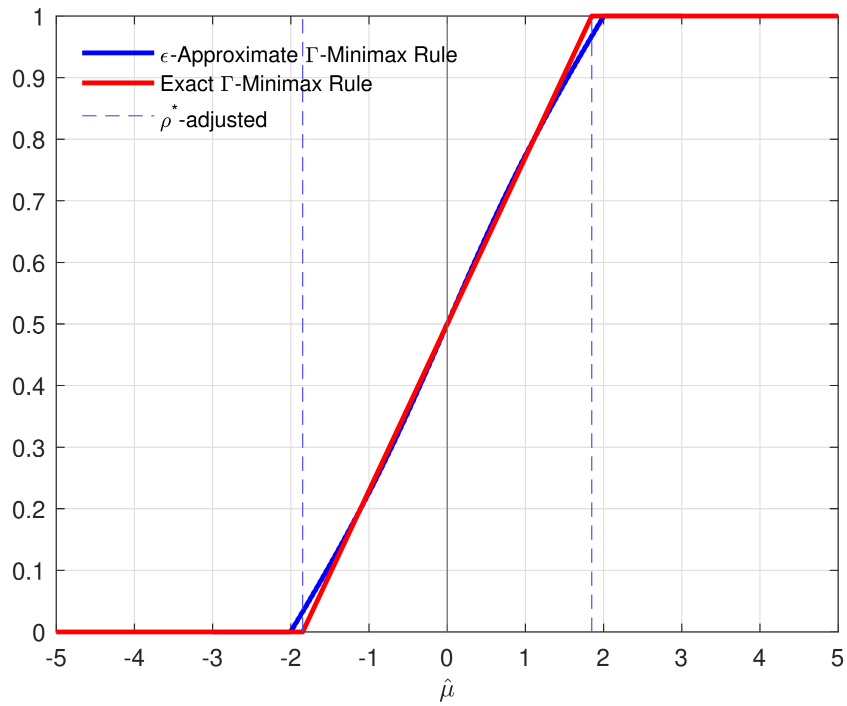

Figure 2: ϵ-Minimax Decision Rule for the 2-point Robust Bayes problem via the Multiplicative Weights algorithm. The graph is generated using $\sigma = 1$, $k = 2$. The approximate $\rho^\star$-adjusted is 1.84.

$k = 2$. Keeping $\epsilon = 0.1$, the algorithm runs for $T = 15,905$ iterations and finishes in 840 seconds (on a personal ASUS Vivobook Pro 15 @ 2.5GHz Intel Core Ultra 9 185H). Figure 3 shows the ϵ-approximate solution.

4.3 ϵ-minimax regret for a simple site selection problem

A policy maker wishes to experimentally evaluate the effects of a new policy, with the end goal of recommending its implementation on a set of different *sites*. The sites are indexed by $s \in \mathcal{S} \equiv \{1, \ldots, S\}$. There are two types of sites: *policy-relevant* (with the set of them denoted $\mathcal{S}_P$) and *experimental* sites (denoted $\mathcal{S}_E$). There are also covariates $X_s \in \mathbb{R}$ available for each site $s \in \mathcal{S}$. This is the *site selection problem* in Gechter et al. (2024) and Egami and Lee (2024).

In order to illustrate the usefulness of our results, consider the following simple example discussed

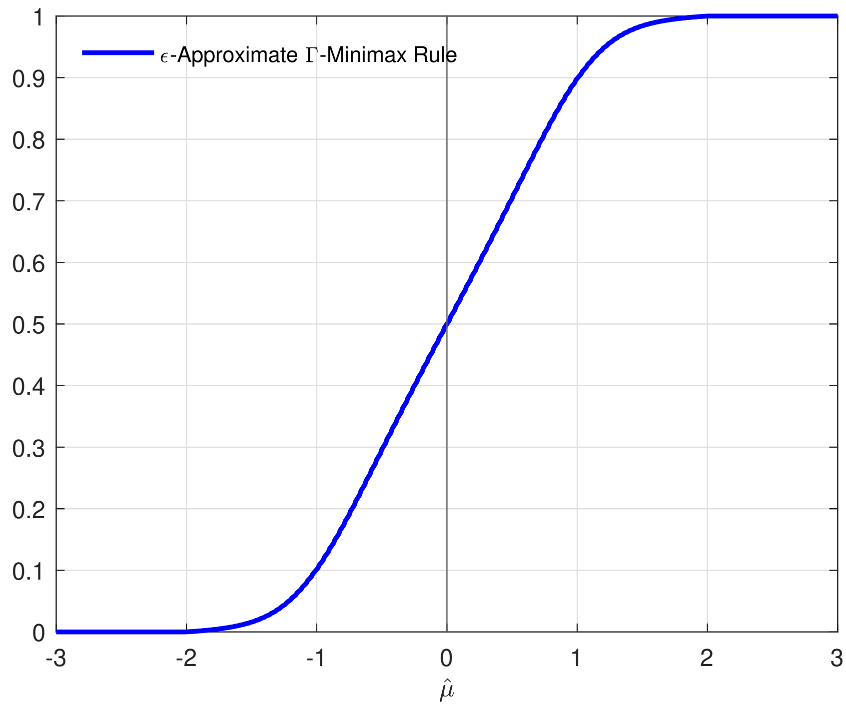

Figure 3: ϵ-Minimax Decision Rule via the MWU algorithm for the 500-point Robust Bayes problem supported along $[-k, k]$.

in Montiel Olea et al. (2024a) where there are a total of four sites partitioned as follows:

$$\mathcal{S}_E \equiv \{1, 4\}, \quad \mathcal{S}_P \equiv \{2, 3\}.$$

Suppose, in addition, that there is one covariate available for each site, satisfying $X_s = s$ for sites $s \in \{1, 2, 3\}$ and $X_4 = 4 + \Delta$, for some $\Delta \geq 0$. The parameter Δ controls the distance between Site 4 and the policy-relevant sites with indexes $\{2, 3\}$. A larger value of Δ means that Site 4 is farther from the experimental sites than Site 1.[9]

Let τ_s denote the true effect of the policy of interest on site s. We impose a Lipschitz condition on the effects $(\tau_1, \tau_2, \tau_3, \tau_4)$:

$$|\tau_s - \tau_{s'}| \leq C||X_s - X_{s'}||, \quad \forall s, s' \in \mathcal{S}.$$

[9]To be more precise, the sum of distances from Site 4 to Sites 2 and 3, is larger than for Site 1.

Call the space of vectors $\tau \equiv (\tau_1, \tau_2, \tau_3, \tau_4) \in \mathbb{R}^4$ that satisfy the equation above Lip_C.

Where to experiment? Suppose that the policy maker can only experiment in one of the two sites in $\mathcal{S}_E$. In order to further simplify our analysis, we will herein assume that selecting a site for experimentation perfectly reveals the effect of the policy of interest on such site. Let $\mathscr{S}$ denote the index of the site where the policy maker decides to experiment. After choosing $\mathscr{S}$, the policymaker must decide how to best use $\tau_{\mathscr{S}}$ for making policy choices on sites $s \in \mathcal{S}_E$. In what follows, let $a(\tau_{\mathscr{S}}) \equiv (a_2(\tau_{\mathscr{S}}), a_3(\tau_{\mathscr{S}})) \in [0,1]^2$ denote the probability of implementing the policy in Sites 2 and 3 after observing $\tau_{\mathscr{S}}$.

Restrictions on decision rules: We assume that the policy maker considers only two decision rules, which we denote by d_1 and d_2. Under the first decision rule (d_1), the policy maker experiments on Site 1 and uses a version of the least-randomizing treatment rules $a(\tau_1) \equiv (a_2(\tau_1), a_3(\tau_1))$ described in Montiel Olea et al. (2024b); that is:

$$\mathscr{S} = 1, \quad a_2(\tau_1) \equiv \begin{cases} 1 & \text{if } \tau_1 > C \\ \frac{C+\tau_1}{2C} & \text{if } \tau_1 \in [-C, C] \\ 0 & \text{if } \tau_1 < -C \end{cases}, a_3(\tau_1) \equiv \begin{cases} 1 & \text{if } \tau_1 > 2C \\ \frac{2C+\tau_1}{4C} & \text{if } \tau_1 \in [-2C, 2C] \\ 0 & \text{if } \tau_1 < -2C. \end{cases}$$

Under the second decision rule (d_2), the policy maker experiments on Site 4:

$$\mathscr{S} = 4,$$

and uses treatment rules $a(\tau_4) \equiv (a_2(\tau_4), a_3(\tau_4))$ analog to $a_2(\tau_1), a_3(\tau_1)$:

$$a_2(\tau_4) \equiv \begin{cases} 1 & \text{if } \tau_4 > (2+\Delta)C \\ \frac{(2+\Delta)C+\tau_4}{2(2+\Delta)C} & \text{if } \tau_4 \in [-(2+\Delta)C, (2+\Delta)C] \\ 0 & \text{if } \tau_4 < -(2+\Delta)C \end{cases}$$

$$a_3(\tau_4) \equiv \begin{cases} 1 & \text{if } \tau_4 > (1+\Delta)C \\ \frac{(1+\Delta)C+\tau_4}{2(1+\Delta)C} & \text{if } \tau_4 \in [-(1+\Delta)C,(1+\Delta)C] \\ 0 & \text{if } \tau_1 < -(1+\Delta)C. \end{cases}$$

We assume that the welfare of each decision rule is the average welfare across policy-relevant sites:

$$W(\mathscr{S}, a(\tau_{\mathscr{S}}), \tau) \equiv \frac{1}{2}\left(a_2(\tau_{\mathscr{S}})\tau_2 + a_3(\tau_{\mathscr{S}})\tau_3\right)$$

and that regret is given by:

$$R(\mathscr{S}, a(\tau_{\mathscr{S}}), \tau) \equiv \frac{1}{2}\left(\tau_2\left(\mathbf{1}\{\tau_2 \geq 0\} - a_2(\tau_{\mathscr{S}})\right) + \tau_3\left(\mathbf{1}\{\tau_3 \geq 0\} - a_4(\tau_{\mathscr{S}})\right)\right).$$

In order to apply the Multiplicative Weights Update algorithm, we allow the policy maker to randomize between the two decision rules d_1 and d_2 described above. This modeling choice allows the policy maker to choose randomly where to experiment. Our framework can thus accommodate the conventional view that *"the external validity of randomized evaluations for a given population (say, the population of a country) would be maximized by randomly selecting sites and, within these sites, by randomly selecting treatment and comparison groups"*, see Duflo et al. (2007).

We let p denote the probability of decision rule d_1. We extend the original regret function $R(\mathscr{S}, a(\tau_{\mathscr{S}}), \tau)$ to any element $p \in [0,1]$ by defining

$$R(p, \tau) = pR(1, a(\tau_1), \tau) + (1-p)R(4, a(\tau_4), \tau).$$

We are interested in using the Multiplicative Weights Update algorithm to approximately solve the minimax problem

$$\inf_{p\in[0,1]} \sup_{\tau\in \mathrm{Lip}_C} R(p,\tau). \tag{11}$$

The results in Montiel Olea et al. (2024a) show that Assumption 1 is satisfied with

$$M = \left(\frac{3 + 2\Delta}{4} \right) C.$$

Figure 4 below shows the number of epochs according to the formula in Theorem 1. The formula is obtained by setting $\epsilon = .1$, $C = 1$, and considering $\Delta \in \{0, .1, .2, \ldots, 2\}$. Figure 5 reports the total runtime of the Multiplicative Weights Update algorithm (based on a personal MacBook Pro 2019 @2.4GHz 8 Core-Intel Core i9).

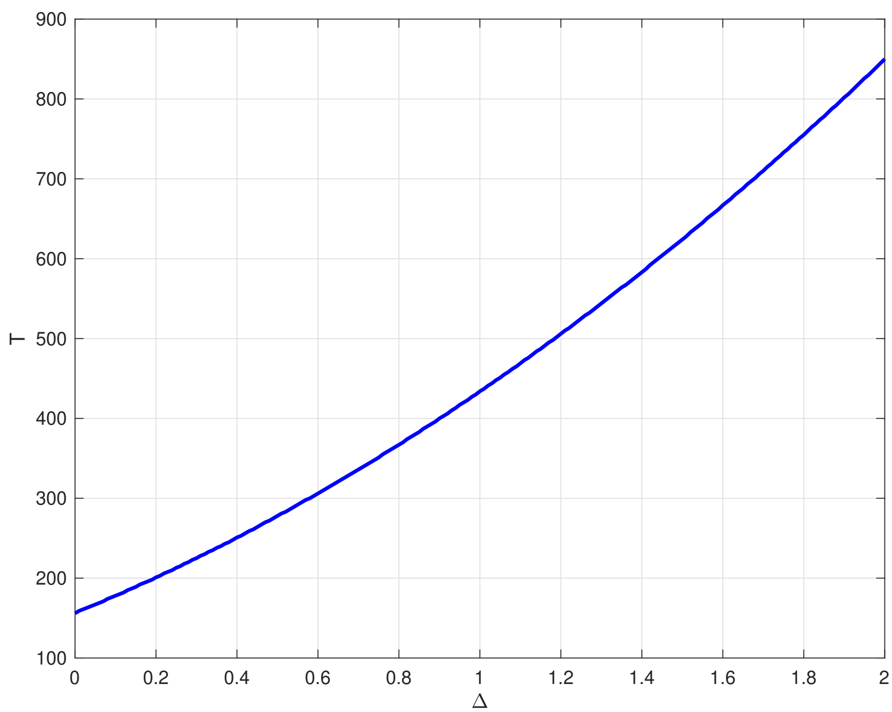

Figure 4: Number of epochs according to the formula in Theorem 1. The graph is generated using $C = 1$, and $\epsilon = .1$.

Figure 6 below shows the ϵ-approximation to the minimax problem in Equation (11). We make three remarks about this figure.

First, when $\Delta \geq 1$, we can solve analytically for the minimax value of this problem. In fact, an

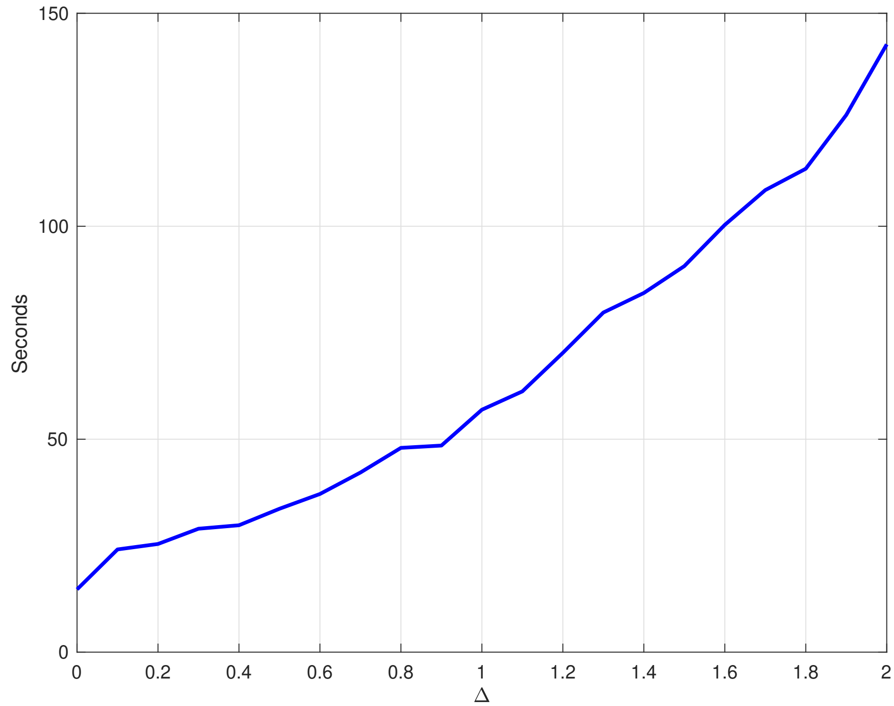

Figure 5: Total runtime of the Multiplicative Weights Update algorithm. The graph is generated using $C = 1$, and $\epsilon = .1$.

extension of the results in Montiel Olea et al. (2024a) can be used to show that even if we allow the policy maker to randomize between decision rules, it is (minimax-regret) optimal to experiment on Site 1 (the solution to the k-median problem in Montiel Olea et al. (2024a) with $k = 1$). The minimax value of the site selection problem can be shown to equal $3C/4$ (solid, black line). The solid red line shows that the Multiplicative Weights algorithm performs as expected, and delivers an ϵ (.1) approximation to the minimax value of the site selection problem. The range between the magenta, dashed line and the solid red line is guaranteed to contain the true minimax value.

Second, when $\Delta = 0$, Montiel Olea et al. (2024a) present an exact solution for the site selection problem without restrictions to the class of decision rules. The minimax value reported therein is $C/2$ (which is achieved by using treatment rules $a_2(\cdot) = a_3(\cdot)$). The approximate minimax value reported by the Multiplicative Weights algorithm is slightly above that number. We note that an important difference relative to the findings of Montiel Olea et al. (2024a) in the context of this example, is that the the decision rules under consideration do not exhibit symmetry in the treatment rules; that is, $a_2(\cdot) \neq a_3(\cdot)$.

Third, the difference between the solid black line and the solid red line for $\Delta \in (0,1)$ approximates the value of randomization, and shows that even in cases where the distance from Site 1 to $\mathcal{S}_P$ is smaller than that of Site 4, it can be suboptimal to select Site 1 with probability one.

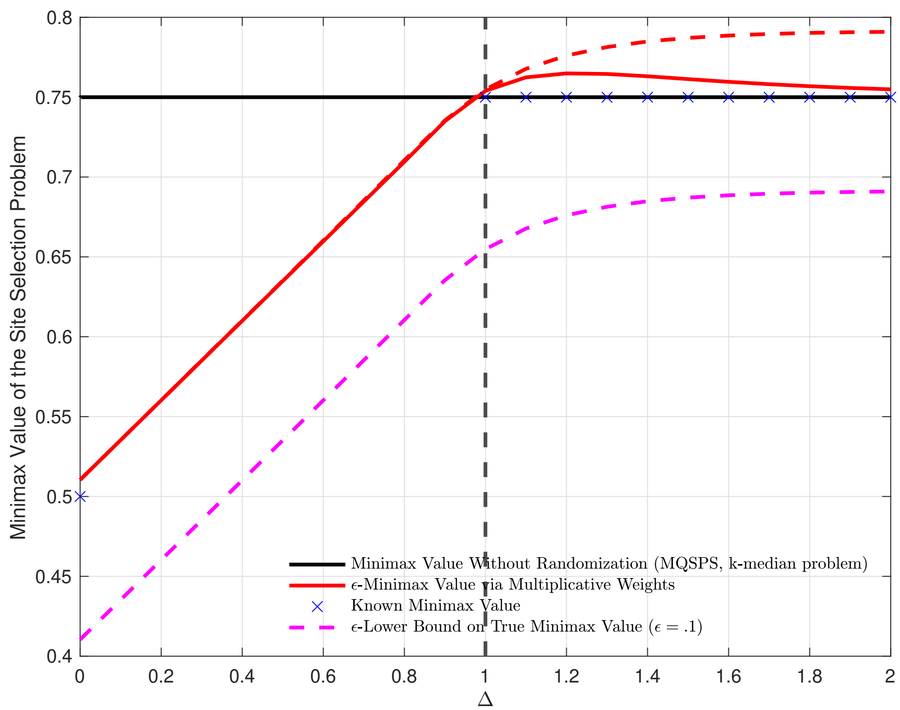

Figure 6: ϵ-Minimax Value of the Site Selection Problem via the Multiplicative Weights algorithm. The graph is generated using $C = 1$, $\epsilon = .1$.

Finally, Figure 7 presents the ϵ-minimax decision rule for the minimax problem in (11). The most surprising result, in our opinion, is that the probability of selecting Site 1, is not monotonic in Δ. In fact, for $\Delta \in [0, .8]$, the ϵ-minimax rule chooses Site 1 with probability in between .4 and .5.

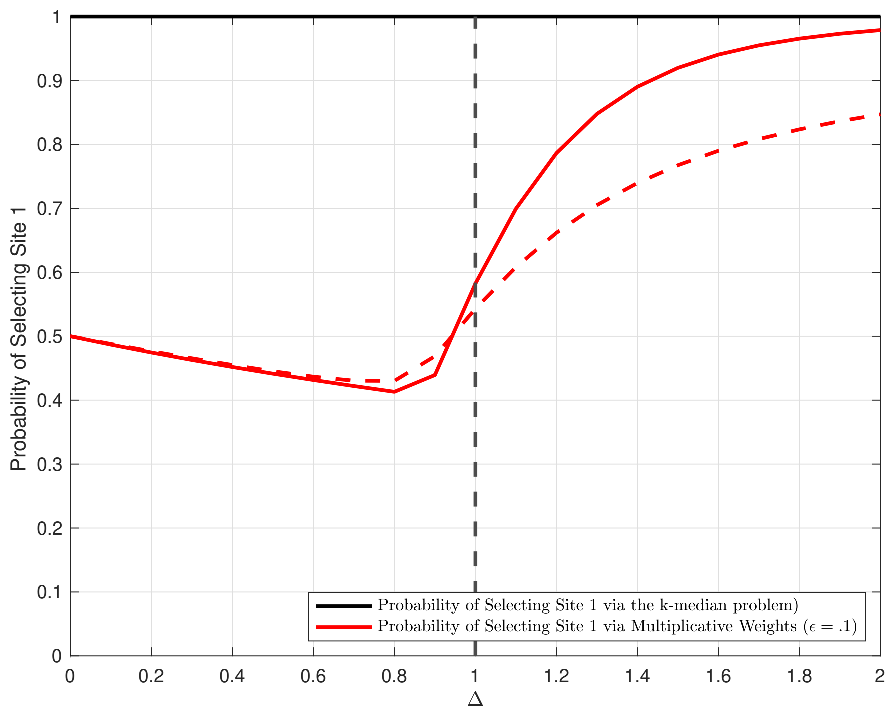

Figure 7: ϵ-Minimax rule for the Site Selection Problem via the Multiplicative Weights algorithm. The graph is generated using $C = 1$, $\epsilon = .1$.

5 Application

Lee et al. (2021) conducted a randomized controlled trial in Bangladesh to estimate the effects of encouraging rural households to receive money transfers from migrant family members. They specifically conducted an encouragement design where poor rural households with family members who had migrated to a larger urban destination receive a 30–45 minute training about how to register and use the mobile banking service "bKash" to send instant remittances back home.

The experiment was conducted in the Gaibandha district, one of Bangladesh's poorest regions. It focused on households that had migrant workers in the Dhaka district, the administrative unit in which the capital of Bangladesh is located. Lee et al. (2021) measure several outcomes of both receiving households and sender migrants; see their Figures 3 and 4. To give a concrete example of the measured outcomes, one question of interest is whether families that adopt the mobile banking technology are more (or less) likely to declare that the *monga*—the seasonal period of hunger in September through November—was not a problem for their household. Table 9, Column 7, p. 60 in Lee et al. (2021) presents results for this specific variable showing that households that used a bKash account in the treatment group are 9.2 percentage points more likely to declare that *monga* was not a problem. The standard error of the estimator is 4.5 percentage points.

Is the corridor selected by Lee et al. (2021) a good choice for a researcher who is concerned about external validity?[10] There are two recent papers that have provided an answer to this question. Gechter et al. (2024) use an elegant decision-theoretic framework to argue that the *Dhaka-Noakhali* corridor would have been a better choice from the perspective of maximizing average welfare. Montiel Olea et al. (2024a) use the framework of Gechter et al. (2024) to argue that the *Dhaka-Pabna* corridor would have been a better choice from the minimax (welfare) regret criterion perspective (restricting the policy maker to consider only nonrandomized selection of corridors). The *Dhaka-Pabna* corridor is also recommended by the *synthetic purposive sampling* approach in Egami and Lee (2024). One important comment is that the *Dhaka-Pabna* corridor is the most representative in terms of covariates, in the sense that it minimizes the average distance (measured using the euclidean distance between covariates) to the 41 migration corridors analyzed in Gechter et al. (2024).

In our application, we consider a situation where a policy maker is considering the three sites mentioned above to run an experiment: Dhaka-Gaibandha (the original site in Lee et al. (2021)), Dhaka-Noakhali (the site suggested by Gechter et al. (2024)) and Dhaka-Pabna (the site suggested in Montiel Olea et al. (2024a)). Each of these sites (migration corridor) have site characteristics $X_s \in \mathbb{R}^d$, with $d = 13$.[11] We index these three sites by $1, 2, 3$ respectively and refer to the set $\mathcal{S}_E \equiv \{1, 2, 3\}$ as the set of experimental sites. Once we exclude these three sites, we have 38

[10] Following Gechter et al. (2024), we name the corridors using a destination-origin format; for example, the migration corridor studied in Lee et al. (2021) is "Dhaka-Gaibandha".

[11] The covariates include mean household income, mean household size, migrant density, mean remittances. See Figure 2 in Montiel Olea et al. (2024a).

migration corridors. We use the distance between the covariates of each of these sites and Dhaka-Pabna to order them in increasing order and index them with integers 4 to 41. Figure 8 presents the distances. The figure shows that for most of the sites the corridor Dhaka-Pabna is the "closest" in terms of the Euclidean distance between covariates.

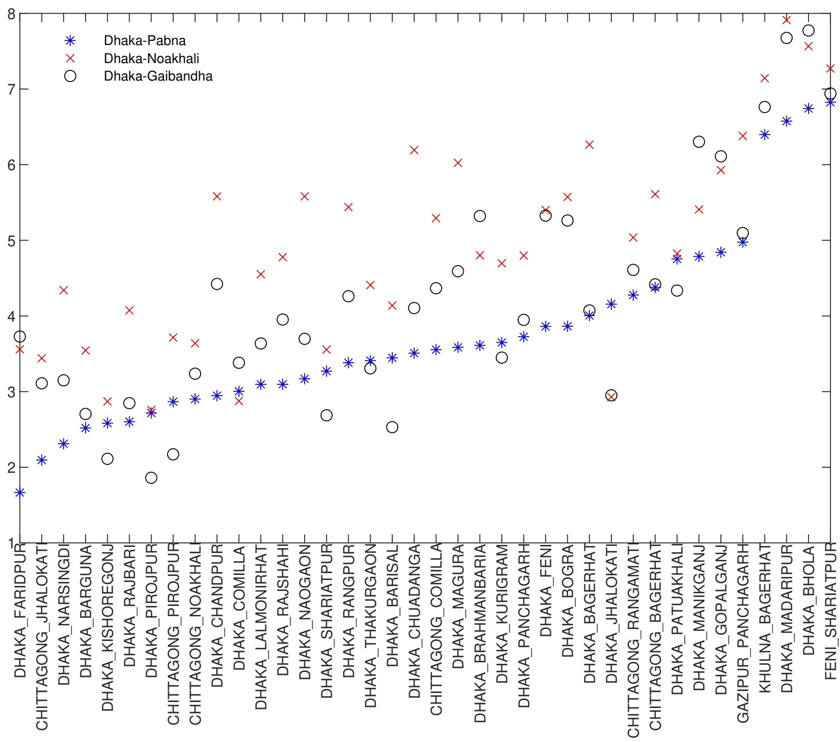

Figure 8: Distances from each of the experimental sites to each of the policy-relevant sites.

We assume that the sites $\mathcal{S}_p \equiv \{4, \ldots, 41\}$ in the x-axis of Figure 8 are the policy-relevant sites. This means that policy maker is interested in deciding whether the training program discussed in Lee et al. (2021) should be rolled out in these sites. We assume that the outcome variable of interest for the policy maker is the likelihood that the households declare that the *monga* was not a problem.

Treatment Effect Heterogeneity: Treatment effect heterogeneity is allowed, but only via the observable site characteristics. The effects of the policy in each site, denoted by τ_s, are restricted to be a Lipschitz function (with respect to a Euclidean norm $||\cdot||$) with known constant C; that is,

$\tau_s = \tau(X_s)$ where

$$|\tau(x) - \tau(x')| \leq C||x - x'||, \quad \forall x, x' \in \mathbb{R}^{13}.$$

One first issue that we need to address in order to conduct our exercise is the value of C that will be used in our application. We do this by using the available point estimates of the treatment effect of the program in Lee et al. (2021). Let x_{DG} denote the covariates of the corridor Dhaka-Gaibandha. Assume that the we entertained the possibility that the true effect, $\tau(x_{DG})$, coincides with the estimated effect 9.2. We consider a "low C" regime.

Suppose that we want to consider a value of C that imposes that if 9.2 were the true effect, then even the corridor that is the most different (in terms of covariates) to Dhaka-Gaibandha the effect of the program must be nonnegative. Intuitively, this means that even for the most different corridor, the effect must remain nonnegative. Dhaka-Bhola is the most different site and $\|x_{\mathrm{DG}} - x_{\mathrm{DB}}\| = 7.7736$. Since the Lipschitz restriction imposes that

$$\tau(x_{\mathrm{DG}}) - C\|x_{\mathrm{DG}} - x_{\mathrm{DB}}\| \leq \tau(x_{\mathrm{DB}}),$$

we could pick C as

$$C = 9.2/7.7736 \approx 1.1834.$$

Treatment Rules: The policy maker makes two choices. First, the policy maker must pick one site on which to experiment. Second, the policy maker must decide how to make treatment choices in all the sites of interest given the available data. Extending the illustrative example discussed in Section 4.3, we assume that if the policy maker decides to experiment on site s, the available data becomes $\widehat{\tau}_s$, with

$$\widehat{\tau}_s \sim \mathcal{N}(\tau_s, \sigma_s^2) \tag{12}$$

and, as in Gechter et al. (2024), we assume σ_s^2 is known. In order to conduct our exercise, we assume that σ_s is the same for all experimental sites, and that it matches the standard error of the estiamted effect of the program in the Dhaka-Gaibhanda site. That is $\sigma_s = 4.5$ for all $s \in \mathcal{S}_E$.

The treatment rule is a mapping $T : \mathbb{R} \to [0,1]^{\#\mathcal{S}_P}$. For $s \in \mathcal{S}_E$ we further denote by T_s the specific policy choice for site s. We refer to a tuple (s, T) as a policy, and we use d to denote it. We consider three nonrandomized policies

$$\mathcal{D} \equiv \{d_1, d_2, d_3\}\,.$$

Under policy d_s, the policy maker experiments on site $s \in \mathcal{S}_E$ and its recommendation for any policy relevant site is $\mathbf{1}\{\widehat{\tau}_s \geq 0\}$. That is, the policy maker makes the same policy recommendation for every policy-relevant site depending on the sign of $\widehat{\tau}_s$. [12] We focus on this special form of policy rule because we think it captures the policy recommendations that are given based on randomized controlled trials.

We consider the following regret function for the policy d_s,

$$R(d_s, \tau) \equiv \frac{1}{\#\mathcal{S}_P} \sum_{s' \in \mathcal{S}_P} \left(\tau(X_{s'})(\mathbf{1}\{\tau(X_{s'}) \geq 0\} - \mathbb{E}_{\tau(X_s)}\left[\mathbf{1}\{\widehat{\tau}_s \geq 0\}\right])\right). \tag{13}$$

This expression can be simplified to:

$$R(d_s, \tau) \equiv \frac{1}{\#\mathcal{S}_P} \sum_{s' \in \mathcal{S}_P} \left(\tau(X_{s'}) \left(\mathbf{1}\{\tau(X_{s'}) \geq 0\} - \Phi\left(\frac{\tau(X_s)}{\sigma_s}\right)\right)\right). \tag{14}$$

The minimax (regret) problem that we are interested in solving is

$$\inf_{p \in \Delta^2} \sup_{\tau \in \mathrm{Lip}_C(\mathbb{R}^{13})} \sum_{s=1}^{3} p_s R(d_s, \tau), \tag{15}$$

where $\mathrm{Lip}_C(\mathbb{R}^{13})$ refers to the space of all Lipschitz functions $f : \mathbb{R}^{13} \to \mathbb{R}$ with constant C.

[12] The results in Montiel Olea et al. (2024a) suggest that this type of policy is likely to be suboptimal. The policy maker could improve its welfare by allowing the treatmetment choice to be randomnly selected, depending on the distance between the policy-relevant site and the experimental site.

5.1 Results

We report results for the case in which $C = 1.1834$. We consider four different scenarios that vary in terms of the number of sites that are policy relevant. The scenarios we consider have either 1, 5, 15, or 38 policy-relevant sites. In each of these cases, we choose to include the sites that are closest to Dhaka-Pabna. For example, when we include only one policy-relevant site we include Dhaka-Faridpur. We do this because, in light of the results in Montiel Olea et al. (2024a), the best nonrandomized choice of experimental site is Dhaka-Pabna. And we would like to use this application to understand how the probability of selecting this site changes as we include sites that perhaps are closer to some of the other experimental sites under consideration.

Figure 9 presents the ϵ-minimax selection of sites obtained via the Multiplicative Weights algorithm. Note first that when there is only one policy-relevant site (and this site is closest to Dhaka-Pabna) the probability of choosing Dhaka-Pabna is close to 1. This is measured by the height of the first yellow bar in Figure 9. We think this is an interesting result as it shows that even if randomization is allowed, it is possible that choosing the site that is most representative for the policy-relevant site is still approximately minimax regret optimal.

The results with five policy relevant sites are also worth discussing. By construction, the five policy-relevant sites that we considered are those that are closest to Dhaka-Pabna. According to Figure 8, Dhaka-Pabna is the nearest neighbor for all of them, with the exception of Dhaka-Kishoregonj. For the latter site, the nearest neighbor is Dhaka-Gaibandha. Figure 9 shows that with 5 sites the ϵ-minimax selection of experimental sites places probability slightly above .2 to Dhaka-Gaibandha (corresponding to the height of the second blue bar) and probability close to .7 to Dhaka-Pabna (corresponding to the height of the second yellow bar).

Finally, we discuss the cases in which the experimental sites are 15 and 38. We note that in both cases the ϵ-minimax solutions are very similar, though the time required by the algorithm and the number of iterations are not; see Tables 1. The recommended probability of experimenting on Dhaka-Pabna is close to .6 (height of the last yellow bar). The recommended probability of

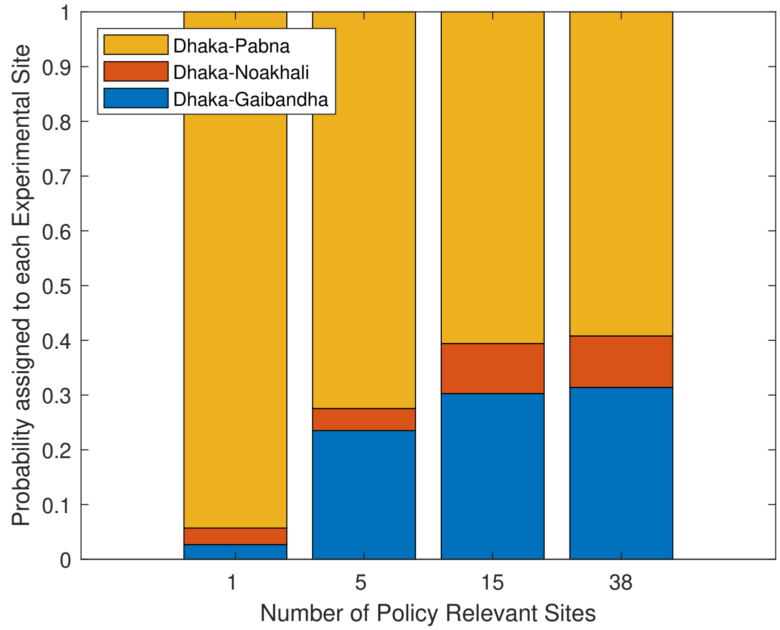

Figure 9: ϵ-Minimax decision rule for the Site Selection Problem via the Multiplicative Weights algorithm. The graph is generated using $C = 1.1834$, $\sigma = 4.5$, and $\epsilon = .1$.

experimenting on Dhaka-Noakhali is close to .1. Interestingly, we note that the ordering of the probabilities is also consistent with the ordering of the three experimental sites (in terms of how frequently they are the nearest neighbor of each of the policy-relevant sites).

Number of Sites	Runtime (seconds)	Iterations
1	$1,765$	$14,063$
5	$1,554$	$11,449$
15	$7,427$	$12,974$
38	$38,741$	$21,248$

Table 1: Runtime (seconds) and Number of Iterations ($C = 1.1834$).

6 Additional Results

6.1 Semi-Infinite Linear Programming for Minimax Problems

Proposition 1 shows that every minimax problem where the statistician chooses randomly among I alternatives can be expressed as a linear program in $I+1$ variables with as many constraints as elements in Θ.

Proposition 1. *Suppose the minimax value $\bar{v}$ in* (2) *is finite. Then $\bar{v}$ is the value function of the following semi-infinite linear program:*

$$\inf_{v\in\mathbb{R},p\in\mathbb{R}^I} v \tag{16}$$

subject to

$$\begin{aligned} v - \sum_{i=1}^{I} p_i R(d_i,\theta) &\geq 0, \forall\theta\in\Theta, \\ p_i &\geq 0, \forall i = 1,...,I, \\ \sum_{i=1}^{I} p_i &= 1. \end{aligned}$$

Moreover, if the random selection $p^\star \in \Delta(\mathcal{D})$ solves (16), *then $p^\star$ is a minimax decision rule.*

Proof. See Appendix A.3. □

Remark 5. Proposition 1 shows that the minimax problem in (2) can be viewed as a linear program with $I+1$ variables and as many restrictions as elements in the parameter space (hence, potentially a semi-infinite linear program). We first note that even when Θ is finite, this linear program may be computationally costly to solve. For instance, suppose Θ has J elements. Then, we can rewrite (16) as the following linear program in standard form:

$$\min_{v,p,\lambda} v$$

subject to the following constraints:

$$v - \sum_{i=1}^{I} p_i R(d_i, \theta) = \lambda_j, \quad j = 1, ..., J,$$

$$\sum_{i=1}^{I} p_i = 1,$$

$$p_i \geq 0, \forall i = 1, ..., I,$$

$$\lambda_j \geq 0, \forall j = 1, ..., J.$$

This can be written in matrix notation as

$$\min \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}^T}_{(I+J+1)\times 1} \begin{pmatrix} v \\ p \\ \lambda \end{pmatrix}$$

subject to the following constraint:

$$\underbrace{\begin{bmatrix} 1 & -\bar{R}(\theta_1) & -1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 1 & -\bar{R}(\theta_J) & 0 & .. & 0 & -1 \\ 0 & \mathbf{1}^\top & 0 & ... & ... & 0 \end{bmatrix}}_{(J+1)\times(1+I+J)} \begin{pmatrix} v \\ p \\ \lambda \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ ... \\ 0 \\ 1 \end{pmatrix}}_{(J+1)\times 1}$$

where $\bar{R}(\theta) \equiv (R(d_1, \theta), ..., R(d_I, \theta))$. Jiang, Song, Weinstein, and Zhang (2020) show that the fastest known LP solver for general (dense) linear programs can solve such a program in an order of approximate $(1 + I + J)^{2.055}$ time. This means that when $I + J$ is large, it could be computationally costly to solve the linear program in (16). This gives a further motivation for the use of

ϵ-approximate minimax solutions.

7 Conclusion

This paper presented an algorithm for obtaining ϵ-*minimax* solutions of statistical decision problems where the statistician is allowed to choose randomly among I decision rules. The notion of an ϵ-minimax decision rule was taken from Ferguson (1967) (Chapter 1, Definition 4) and it refers to a decision rule whose worst-case expected loss exceeds the minimax value of the decision problem by at most an additive factor of ϵ.[13]

Once we allow for randomized selection over the I decision rules, the minimax problem admits a convex programming representation over the $(I-1)$-simplex, an observation which has been previously documented in the literature by Chamberlain (2000). Both the objective function and the subgradient of this problem are in general difficult to evaluate. The reason being that the objective function of the convex problem involves solving a nonconvex maximization problem to find the worst-case performance (over the model's parameter space) of a specific randomized selection over the I rules. This type of problem arises commonly in the convex optimization literature; see Bubeck et al. (2015) and the seminal work of Nemirovski and Yudin (1983). The algorithm herein suggested is a version of *mirror subgradient descent*, initialized with uniform weights and stopped after a finite number of iterations. The early stopping of the algorithm tries to minimize the number of calls to the objective function and its subgradient.

The iterative procedure arising from this mirror descent routine described in this paper is known in the computer science literature as the *Multiplicative Weights* update method, and it is used in algorithmic game theory as a practical tool to find approximate solutions of two-person zero-sum games. The paper applies the suggested algorithm to different minimax problems in the

[13]We note that the definition given in Ferguson (1967) differs of the usage of ϵ-minimax decision rules in other contexts. Most notably, from the work of Manski and Tetenov (2016), who use the term ϵ-minimax to refer to a decision rule whose worst-case regret is at most ϵ.

econometrics literature. In some of these problems, the minimax solution is known, and we show numerically that in these examples the ϵ-minimax solution is practically the same as the true minimax solution.

Finally, we apply the algorithm to the *site selection problem* of Gechter et al. (2024); namely, how to optimally selecting *sites* to maximize the external validity of an experimental policy evaluation. Our algorithm allows the researcher to choose randomly where to experiment, but adjusting optimally for the available baseline covariate information.

We think there are several interesting areas for future work, both from an applied and from a more theoretical perspective. From a purely applied angle, applying this algorithm could be helpful in settings in which the space of decision rules is already discrete, but large. For example, as in the recent interesting work of Christy and Kowalski (2024). Relatedly, the algorithm herein suggested might also be helpful in applications in which the parameter space is "large". For example, in the site selection application we considered the parameter space is a subset of $\mathbb{R}^S$ where S is the total number of experimental and policy-relevant sites. This means that that our algorithm could be useful in extending the scope of certain minimax problems, such as the one described in the recent work of Armstrong, Kline, and Sun (2024). More generally, the algorithm we presented is part of large literature in computer science and algorithmic game theory. We think that there is an opportunity to apply some of the procedures suggested in this literature (including some of the recent procedures that parameterize the strategy space using neural networks).

From a more theoretical perspective, it would be interesting to further explore the differences between ϵ-minimax strategies and the notion of a local min-max point in Daskalakis et al. (2021). There are very interesting results about the relation between this notion and the stationary points of sugbradient ascent-descent dynamics. But it would be great to understand, theoretically and empirically, what are the potential benefits of searching for these type of points as opposed to ϵ-minimax strategies.

References

Adler, I. (2013): "The equivalence of linear programs and zero-sum games," *International Journal of Game Theory*, 42, 165–177.

Aradillas Fernández, A., J. L. Montiel Olea, C. Qiu, J. Stoye, and S. Tinda (2024): "Robust Bayes Treatment Choice with Partial Identification," *arXiv e-prints*, arXiv–2408.

Armstrong, T., P. M. Kline, and L. Sun (2024): "Adapting to misspecification," Tech. rep., National Bureau of Economic Research.

Arora, S., E. Hazan, and S. Kale (2012): "The multiplicative weights update method: a meta-algorithm and applications," *Theory of computing*, 8, 121–164.

Blackwell, D. A. and M. A. Girshick (1954): *Theory of games and statistical decisions*, John Wiley, New York.

Blume, L. E. (1993): "The statistical mechanics of strategic interaction," *Games and economic behavior*, 5, 387–424.

Bubeck, S. et al. (2015): "Convex optimization: Algorithms and complexity," *Foundations and Trends® in Machine Learning*, 8, 231–357.

Chamberlain, G. (2000): "Econometric applications of maxmin expected utility," *Journal of Applied Econometrics*, 15, 625–644.

Christy, N. and A. E. Kowalski (2024): "Starting Small: Prioritizing Safety over Efficacy in Randomized Experiments Using the Exact Finite Sample Likelihood," *arXiv preprint arXiv:2407.18206*.

Dantzig, G. B. (1951): "A proof of the equivalence of the programming problem and the game problem," *Activity analysis of production and allocation*, 13.

Daskalakis, C., S. Skoulakis, and M. Zampetakis (2021): "The complexity of constrained min-max optimization," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 1466–1478.

Dominitz, J. and C. F. Manski (2024): "Comprehensive OOS Evaluation of Predictive Algorithms with Statistical Decision Theory," Tech. rep., National Bureau of Economic Research.

Du, D.-Z. and P. M. Pardalos (1995): *Minimax and applications*, vol. 4, Springer Science & Business Media.

Duflo, E., R. Glennerster, and M. Kremer (2007): "Using randomization in development economics research: A toolkit," *Handbook of development economics*, 4, 3895–3962.

Egami, N. and D. D. I. Lee (2024): "Designing Multi-Context Studies for External Validity: Site Selection via Synthetic Purposive Sampling," *Working Paper*.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly optimal tests when a nuisance parameter is present under the null hypothesis," *Econometrica*, 83, 771–811.

Ferguson, T. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, vol. 7, Academic Press New York.

Filar, J. A. and T. Raghavan (1982): "An algorithm for solving S-games and differential S-games," *SIAM Journal on Control and Optimization*, 20, 763–769.

Freund, Y. and R. E. Schapire (1999): "Adaptive game playing using multiplicative weights," *Games and Economic Behavior*, 29, 79–103.

Fudenberg, D. and D. K. Levine (1995): "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control*, 19, 1065–1089.

Gechter, M., K. Hirano, J. Lee, M. Mahmud, O. Mondal, J. Morduch, S. Ravindran, and A. S. Shonchoy (2024): "Selecting Experimental Sites for External Validity," *Working Paper*.

Giacomini, R. and T. Kitagawa (2021): "Robust Bayesian inference for set-identified models," *Econometrica*, 89, 1519–1556.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of mathematical economics*, 18, 141–153.

Gravin, N., Y. Peres, and B. Sivan (2016): "Tight lower bounds for multiplicative weights algorithmic families," *arXiv preprint arXiv:1607.02834*.

Jiang, S., Z. Song, O. Weinstein, and H. Zhang (2020): "Faster dynamic matrix inverse for faster lps," *arXiv preprint arXiv:2004.07470*.

Kaji, T., E. Manresa, and G. Pouliot (2023): "An adversarial approach to structural estimation," *Econometrica*, 91, 2041–2063.

Kempthorne, P. J. (1987): "Numerical specification of discrete least favorable prior distributions," *SIAM Journal on Scientific and Statistical Computing*, 8, 171–184.

Lee, J. N., J. Morduch, S. Ravindran, A. Shonchoy, and H. Zaman (2021): "Poverty and migration in the digital age: Experimental evidence on mobile banking in Bangladesh," *American Economic Journal: Applied Economics*, 13, 38–71.

Lewis, G. and V. Syrgkanis (2018): "Adversarial generalized method of moments," *arXiv preprint arXiv:1803.07164*.

Luedtke, A., M. Carone, N. Simon, and O. Sofrygin (2020): "Learning to learn from data: Using deep adversarial learning to construct optimal statistical procedures," *Science Advances*, 6, eaaw2140.

Luedtke, A., I. Chung, and O. Sofrygin (2021): "Adversarial Monte Carlo meta-learning of optimal prediction procedures," *Journal of Machine Learning Research*, 22, 1–67.

Manski, C. F. (2021): "Econometrics for decision making: Building foundations sketched by Haavelmo and Wald," *Econometrica*, 89, 2827–2853.

Manski, C. F. and A. Tetenov (2016): "Sufficient trial size to inform clinical practice," *Proceedings of the National Academy of Sciences*, 113, 10518–10523.

Montiel Olea, J. L., B. Prallon, C. Qiu, J. Stoye, and Y. Sun (2024a): "Externally Valid Selection of Experimental Sites via the k-Median Problem," *https://arxiv.org/abs/2408.09187*.

Montiel Olea, J. L., C. Qiu, and J. Stoye (2024b): "Decision Theory for Treatment Choice Problems with Partial Identification," *arXiv preprint arXiv:2312.17623*.

Nemirovski, A. and D. Yudin (1983): *Problem Complexity and Method Efficiency in Optimization*, A Wiley-Interscience publication, Wiley.

Owen, G. (2013): *Game theory*, Emerald Group Publishing.

Rockafellar, R. T. (1970): "Convex analysis," .

Stoye, J. (2012): "Minimax regret treatment choice with covariates or with limited validity of experiments," *Journal of Econometrics*, 166, 138–156.

Troutt, M. D. (1978): "Algorithms for non-convex S n-games," *Mathematical Programming*, 14, 332–348.

Wald, A. (1950): *Statistical Decision Functions*, New York: Wiley.

Yata, K. (2021): "Optimal Decision Rules Under Partial Identification," ArXiv:2111.04926 [econ.EM], `https://doi.org/10.48550/arXiv.2111.04926`.

A Proofs of Main Results

A.1 Proof of Lemma 1

Take $p, p' \in \Delta(\mathcal{D})$. Note that

$$\begin{aligned} f(\alpha p + (1-\alpha)p') &= \sup_{\theta \in \Theta} R(\alpha p + (1-\alpha)p', \theta) \\ &= \sup_{\theta \in \Theta} \left(\sum_{i=1}^{I} (\alpha p_i + (1-\alpha)p'_i) R(d_i, \theta) \right) \\ &= \sup_{\theta \in \Theta} \left(\alpha \sum_{i=1}^{I} p_i R(d_i, \theta) + (1-\alpha) p'_i R(d_i, \theta) \right) \\ &\leq \alpha \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_i R(d_i, \theta) + (1-\alpha) \sup_{\theta \in \Theta} p'_i R(d_i, \theta) \\ &= \alpha f(p) + (1-\alpha) f(p'). \end{aligned}$$

Thus, the function $f(\cdot)$ is convex.

We now show that g_0 is a subgradient of f at p_0. That is, that for any $p \in \Delta(\mathcal{D})$:

$$f(p) \geq f(p_0) + g_0^\top (p - p_0).$$

Let p be an arbitrary point in $\Delta(\mathcal{D})$. By definition

$$\begin{aligned} f(p_0) &\equiv \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{0,i} R(d_i, \theta) \\ &= \sum_{i=1}^{I} p_{0,i} R(d_i, \theta_0) \\ &= \sum_{i=1}^{I} (p_{0,i} - p_i) R(d_i, \theta_0) + \sum_{i=1}^{I} p_i R(d_i, \theta) \\ &= g_0^{\top}(p_0 - p) + \sum_{i=1}^{I} p_i R(d_i, \theta) \\ &\leq g_0^{\top}(p_0 - p) + f(p). \end{aligned}$$

Thus, g_0 is a subgradient of $f(\cdot)$ at p_0.

A.2 Proof of Theorem 1

The proof is based on an extension of the arguments in Theorem 2.1 and Theorem 2.3 in Arora et al. (2012). For the sake of exposition, we divide our proof in three steps.

STEP 1: Fix the step-size η. First, we show that after T rounds the algorithm guarantees that, for all decision rules $d_i \in \{d_1, ..., d_I\}$, we have obtained average payoffs bounded above by the average payoff of any decision rule d_i plus an error term. More precisely:

$$\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right) \leq \frac{1}{T} \sum_{t=1}^{T} R(d_i, \theta_t) + M^2 \eta + \frac{\ln(I)}{T\eta} \tag{17}$$

To show this, we use a similar argument to Theorem 2.1 in Arora et al. (2012). Let $M > 0$ be the bound on the risk function in Assumption 1. We define the normalized sub-gradient, $m_t \equiv g_t / M$.

Then, recall the definition of ϕ_t, we have that

$$\begin{aligned}
\phi_{t+1} &= \sum_{i=1}^{I} w_{i,t} \\
&= \sum_{i=1}^{I} w_{i,t-1} \exp\left(-\eta g_{i,t}\right) \\
&= \sum_{i=1}^{I} w_{i,t-1} \exp\left(-\eta M m_{i,t}\right) \\
&\leq \sum_{i=1}^{I} w_{i,t-1} \left(1 - \eta M m_{i,t} + \eta^2 M^2 m_{i,t}^2\right) \\
&= \phi_t - \phi_t \eta M \sum_{i=1}^{I} p_{i,t} m_{i,t} + \phi_t \eta^2 M^2 \sum_{i=1}^{I} p_{i,t} m_{i,t}^2 \\
&= \phi_t \left(1 - \eta M \sum_{i=1}^{I} p_{i,t} m_{i,t} + \eta^2 M^2 \sum_{i=1}^{I} p_{i,t} m_{i,t}^2\right) \\
&\leq \phi_t \exp\left(-\eta M \sum_{i=1}^{I} p_{i,t} m_{i,t} + \eta^2 M^2 \sum_{i=1}^{I} p_{i,t} m_{i,t}^2\right).
\end{aligned}$$

The first inequality follows from the fact that $\exp(-x) \leq 1-x+x^2$, for $|x| \leq 1$.[14] The last inequality follows from $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$. By induction after T rounds, and using the fact that w_0 was initialized to be a vector of ones (i.e., $w_0 := \mathbf{1}$, we have

$$\begin{aligned}
\phi_{T+1} &\leq \phi_1 \exp\left(-\eta M \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t} + \eta^2 M^2 \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t}^2\right) \\
&= I \exp\left(-\eta M \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t} + \eta^2 M^2 \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t}^2\right). \qquad (18)
\end{aligned}$$

[14]Note that:

$$\begin{aligned}
\exp(-x) &= 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + ... \\
&\leq 1 - x + x^2
\end{aligned}$$

if and only if:

$$0 \leq \frac{x^2}{2!} + \frac{x^3}{3!} - \frac{x^4}{4!} + ...$$

Note that $x^n \geq x^{n+1}$ for all $|x| \leq 1$ and $n \in \mathbb{N}$, and so the statement above holds.

Also notice that

$$\phi_{T+1} \geq w_{i,t+1} = \prod_{t=1}^{T} \exp\left(-\eta g_{i,t}\right), \tag{19}$$

After taking logs of both sides in (18) and (19), we have

$$-\sum_{t=1}^{T} g_{i,t} \leq \frac{\ln(I)}{\eta} - M \sum_{t=1}^{T}\sum_{i=1}^{I} p_{i,t} m_{i,t} + \eta M^2 \sum_{t=1}^{T}\sum_{i=1}^{I} p_{i,t} m_{i,t}^2.$$

Since $m_{i,t} = g_{i,t}/M = R(d_i, \theta_t)/M$, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \leq & \frac{1}{T}\sum_{t=1}^{T} R(d_i, \theta_t) + \frac{M^2\eta}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_{i,t} m_{i,t}^2 + \frac{\ln(I)}{T\eta} \\
\leq & \frac{1}{T}\sum_{t=1}^{T} R(d_i, \theta_t) + M^2\eta + \frac{\ln(I)}{T\eta} \\
& (\text{since } m_{i,t}^2 \leq 1) \\
\leq & \frac{1}{T}\sum_{t=1}^{T} R(d_i, \theta_t) + \frac{\epsilon}{2} + \frac{2\ln(I)M^2}{T\epsilon} \\
& (\text{since } \eta = \epsilon/2M^2).
\end{aligned}$$

STEP 2: Let $p_i^\epsilon \equiv \frac{1}{T}\sum_{t=1}^{T} p_{i,t}$. We show that after T rounds, we have that:

$$\bar{v} \leq \sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i^\epsilon R(d_i, \theta) \leq \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t)\right) \leq \bar{v} + \frac{\epsilon}{2} + \frac{\ln(I)}{T}\left(\frac{2M^2}{\epsilon}\right),$$

where $\bar{v}$ is the minimax value of the decision problem.

To show this, note that the lower bound holds by definition. For the upper bound, note:

$$\begin{aligned}\sup_{\theta\in\Theta}\sum_{i=1}^{I} p_i^{\epsilon} R(d_i,\theta) &= \sup_{\theta\in\Theta}\sum_{i=1}^{I}\left(\frac{1}{T}\sum_{t=1}^{T} p_{i,t}\right) R(d_i,\theta) \\ &= \sup_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_{i,t}R(d_i,\theta)\right) \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_{i,t}R(d_i,\theta_t)\right), \end{aligned} \tag{20}$$

where the inequality uses the fact that θ_t is nature's best response to p_t.

Let $p^{\star}$ be the minimax strategy of the decision problem (which exists by assumption of the Theorem). Since, by Step 1, the right hand side of the equation above is bounded by above by Equation (17) for any d_i, it follows that for any $p_i \in \Delta^{I-1}$, 20 is bounded above by

$$\frac{1}{T}\sum_{t=1}^{T} p_i R(d_i,\theta_t) + \frac{\epsilon}{2} + \frac{\ln(I)}{T}\left(\frac{4M^2}{\epsilon}\right).$$

In particular, we have that

$$\sum_{i=1}^{I} p_i^{\star} R(d_i,\theta_t) \leq \sup_{\theta\in\Theta}\sum_{i=1}^{I} p_i^{\star} R(d_i,\theta) = \bar{v}.$$

Consequently,

$$\sup_{\theta\in\Theta}\sum_{i=1}^{I} p_i^{\epsilon} R(d_i,\theta) \leq \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_{i,t}R(d_i,\theta_t)\right) \leq \bar{v} + \frac{\epsilon}{2} + \frac{\ln(I)}{T}\left(\frac{2M^2}{\epsilon}\right).$$

This gives the desired result.

STEP 3: By taking the smallest integer T such that

$$\frac{\ln(I)}{T}\left(\frac{2M^2}{\epsilon}\right) \leq \frac{\epsilon}{2},$$

or, equivalently,

$$T = \lceil 4M^2 \ln(I)/\epsilon^2 \rceil.$$

We then conclude that

$$\bar{v} \leq \sup_{\theta \in \Theta} \left(\sum_{i=1}^{I} p_i^\epsilon R(d_i, \theta) \right) \leq \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right) \leq \bar{v} + \frac{\epsilon}{2} + \frac{\epsilon}{2} = \bar{v} + \epsilon,$$

where $p_i^\epsilon \equiv \frac{1}{T} \sum_{t=1}^{T} p_{i,t}$ (as defined before). Since

$$p^\epsilon \equiv (p_1^\epsilon, ..., p_I^\epsilon) \in \Delta^{I-1},$$

we conclude that $\tilde{p}^\star$ is an ϵ-minimax decision rule and that $\bar{v}^\epsilon \equiv \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right)$ is an ϵ-approximation to $\bar{v}$.

A.3 Proof of Proposition 1

Let $v^\star$ be the value of the program in (16). First, we need to show that $\bar{v} \leq v^\star$. Let $C^\star$ denote the choice set of the program in (16). If the choice is empty, then $v^\star = \infty$, and the statement will be vacuously true. Therefore, assume that $C^\star \neq \emptyset$.

Take any $(v, p) \in C^\star$. By definition, for any such $(v, p) \in C^\star$, we must have

$$v \geq \sum_{i=1}^{I} p_i R(d_i, \theta), \forall \theta \in \Theta.$$

In particular,

$$v \geq \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_i R(d_i, \theta).$$

Therefore,

$$v^\star \equiv \inf_{(v,p)\in C^\star} v \geq \inf_{(v,p)\in C^\star} \sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i R(d_i,\theta) \geq \inf_{p\in\Delta(\mathcal{D})} \sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i R(d_i,\theta) = \bar{v},$$

where the last inequality follows from the fact that if $(v,p) \in C^*$ then $p \in \Delta(\mathcal{D})$.

Now, we need to show that $v^* \leq \bar{v}$. Suppose this is not the case; i.e., $\bar{v} < v^*$. Pick any $\epsilon > 0$ such that $\bar{v} + \epsilon < v^*$. By definition of $\bar{v}$, for any $\epsilon > 0$ there must exist $p_\epsilon \in \Delta(\mathcal{D})$ such that

$$\bar{v} \leq \sup_{\theta\in\Theta} \sum_{i=1}^{I} p_\epsilon R(d_i,\theta) \leq \bar{v} + \epsilon < v^*.$$

Let $v_\epsilon \equiv \bar{v} + \epsilon$. Note that $(v_\epsilon, p_\epsilon) \in C^\star$. Therefore, $v_\epsilon \geq v^* > v_\epsilon$. A contradiction.

Finally, suppose $p^\star \in \Delta(\mathcal{D})$ solves (16). By definition,

$$\sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i^\star R(d_i,\theta) \geq \bar{v}$$

We just need to show that

$$\sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i^\star R(d_i,\theta) \leq \bar{v}$$

Suppose not, and therefore, we have

$$\sup_{\theta\in\Theta} \sum_{i=1}^{I} p_i^\star R(d_i,\theta) > \bar{v}$$

By definition of $\bar{v}$, we have that $\forall \epsilon > 0$, $\exists p_\epsilon^\star \in \Delta(\mathcal{D})$ such that

$$\sup_{\theta\in\Theta} \sum_{i=1}^{I} p_{\epsilon,i} R(d_i,\theta) < \bar{v} + \epsilon$$

Then choose $\epsilon > 0$ such that

$$\bar{v} + \epsilon < \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_i^\star R(d_i, \theta)$$

Then,

$$\sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{\epsilon,i} R(d_i, \theta) < \sup_{\theta \in \Theta} \sum_{i=1}^{n} p_i^\star R(d_i, \theta)$$

Take $\tilde{v}_\epsilon \equiv \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_{\epsilon,i} R(d_i, \theta)$. Since $p^\star$ solves (16), this means

$$v^\star \geq \sum_{i=1}^{I} p_i^\star R(d_i, \theta), \forall \theta \in \Theta$$

Note that $(\tilde{v}_\epsilon, p_\epsilon) \in C^\star$ and

$$\tilde{v}_\epsilon < \sup_{\theta \in \Theta} \sum_{i=1}^{I} p_i^\star R(d_i, \theta) \leq v^\star$$

Therefore, this means $(p^\star, v^\star)$ cannot solve (16), which is a contradiction. Therefore, we must have $\bar{v} = v^\star$. Thus, p^* is minimax. □

A.4 Reducing the number of epochs

In this section, we show that it is possible to reduce the number of epochs used in Algorithm 1, while preserving the quality of the approximation. The result is obtained by adapting the proof technique for the convergence of the Hedge algorithm used in Freund and Schapire (1999). Before stating the theorem, we first recall the definition of Kullback-Leibler divergence between two discrete distributions defined on the same support: suppose distributions p_1, p_2 are supported on I discrete points, and they can be expressed as vectors $p_1 = (p_{i,1})_{i=1}^{I}, p_2 = (p_{i,2})_{i=1}^{I}$, then the KL divergence between them is defined as

$$\mathbf{KL}(p_1, p_2) = \sum_{i=1}^{I} p_{i,1} \ln \frac{p_{i,1}}{p_{i,2}}.$$

Theorem 2. *Suppose Assumptions 1-2 hold. Furthermore, assume there exists a minimax decision rule $p^\star \in \Delta(\mathcal{D})$. If $\eta \equiv \ln\left(1 + \frac{\epsilon}{M}\right)/M$, and $T \equiv 2\lceil M(M+\epsilon)\ln(I)/\epsilon^2 \rceil$, then the random choice of* *decision rules that assigns probability*

$$\frac{1}{T}\sum_{t=1}^{T} p_{i,t}$$

to each decision rule d_i—where p_t corresponds to the t-th iteration of the mirror descent routine in Section 2.2—is ϵ-minimax in the sense of Definition 1.

Proof. It suffices to modify Step 1 in Theorem 1. For this purpose, fix the step-size $\eta > 0$. First, we show that after T rounds that for any $p \in \Delta(D)$, our average payoffs can be bounded above by the payoff of p plus an error term. More precisely:

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_{i,t} R(d_i,\theta_t)\right) \leq \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{I} p_i R(d_i,\theta_t)\right) + \frac{\epsilon}{2} + \frac{\ln(I)}{T}\cdot\frac{M(M+\epsilon)}{\epsilon}. \tag{21}$$

Note that:

$$\begin{aligned}\mathbf{KL}(p,p_{t+1}) - \mathbf{KL}(p,p_t) &= \sum_{i=1}^{I} p_i \ln\frac{p_{i,t}}{p_{i,t+1}} \\ &= \sum_{i=1}^{I} p_i \ln\frac{\phi_{t+1}}{\phi_t \exp(-\eta g_{i,t})} \\ &= \eta\sum_{i=1}^{I} p_i g_{i,t} + \ln\frac{\phi_{t+1}}{\phi_t}.\end{aligned}$$

Moreover, since $p_{i,t} = w_{i,t-1}/\phi_t$ and $\phi_{t+1} = \sum_{i=1}^{I} w_{i,t} = \sum_{i=1}^{I} w_{i,t-1}\exp(-\eta g_{i,t})$

$$\frac{\phi_{t+1}}{\phi_t} = \sum_{i=1}^{I} p_{i,t}\exp(-\eta g_{i,t}).$$

Let $m_{i,t} = g_{i,t}/M$ and $\tilde{\eta} = M\eta$. We then have

$$\begin{aligned}
\mathbf{KL}(p, p_{t+1}) - \mathbf{KL}(p, p_t) &= \eta \sum_{i=1}^{I} p_i g_{i,t} + \ln \sum_{i=1}^{I} p_{i,t} \exp(-\eta g_{i,t}) \\
&= \tilde{\eta} \sum_{i=1}^{I} p_i m_{i,t} + \ln \sum_{i=1}^{I} p_{i,t} \exp(-\tilde{\eta} m_{i,t}) \\
&\leq \tilde{\eta} \sum_{i=1}^{I} p_i m_{i,t} + \ln \sum_{i=1}^{I} p_{i,t} \left(1 - (1 - e^{-\tilde{\eta}}) m_{i,t}\right) \\
&= \tilde{\eta} \sum_{i=1}^{I} p_i m_{i,t} + \ln \left(1 - (1 - e^{-\tilde{\eta}}) \sum_{i=1}^{I} p_{i,t} m_{i,t}\right) \\
&\leq \tilde{\eta} \sum_{i=1}^{I} p_i m_{i,t} - (1 - e^{-\tilde{\eta}}) \sum_{i=1}^{I} p_{i,t} m_{i,t},
\end{aligned}$$

where the first inequality applies $e^{-ax} \leq 1 - (1 - e^{-a})x, \forall x \in [0,1], a > 0$ [15], while the last line follows from the fact that $\ln(1 - x) \leq -x$ for all $x \leq 1$. Induction for T rounds yields:

$$\mathbf{KL}(p, p_{T+1}) - \mathbf{KL}(p, p_1) \leq \tilde{\eta} \sum_{t=1}^{T} \sum_{i=1}^{I} p_i m_{i,t} - (1 - e^{-\tilde{\eta}}) \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t}.$$

Since $p_{i,1} = 1/I$ for all $i = 1, \ldots, I$, then $\mathbf{KL}(p, p_1) \leq \ln(I)$ for all $p \in \Delta(D)$. Therefore,

$$0 \leq \mathbf{KL}(p, p_{T+1}) \leq \tilde{\eta} \sum_{t=1}^{T} \sum_{i=1}^{I} p_i m_{i,t} - (1 - e^{-\tilde{\eta}}) \sum_{t=1}^{T} \sum_{i=1}^{I} p_{i,t} m_{i,t} + \ln(I).$$

[15] It suffices to notice that e^{-ax} is a convex function of x, and $1 - (1 - e^{-a})x$ is the line passing $(0, 1), (1, e^{-a})$.

Consequently:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_{i,t}R(d_i,\theta_t) &\le \frac{\tilde{\eta}}{1-e^{-\tilde{\eta}}}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) + \frac{\ln(I)M}{T(1-e^{-\tilde{\eta}})} \\
&\le \frac{1-e^{-2\tilde{\eta}}}{2e^{-\tilde{\eta}}(1-e^{-\tilde{\eta}})}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) + \frac{\ln(I)M}{T(1-e^{-\tilde{\eta}})} \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) + \frac{1-e^{-\tilde{\eta}}}{2e^{-\tilde{\eta}}}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) + \frac{\ln(I)M}{T(1-e^{-\tilde{\eta}})} \\
&\le \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) + \frac{e^{\tilde{\eta}}-1}{2}\cdot M + \frac{\ln(I)Me^{\tilde{\eta}}}{T(e^{\tilde{\eta}}-1)},
\end{aligned}
$$

Since $\eta = \ln(1+\epsilon/M)/M$, then $\tilde{\eta} = \ln(1+\epsilon/M)$. Taking into the definition of η, we get (21).

□

B Connection to S Games

The minimax problem we study is closely related to what blackwell1954 call an S-Game. Player I, the statistician in our case, has a finite number of pure strategies $d \in \mathcal{D} \equiv \{d_1,\ldots,d_I\}$. Player II, nature, may have infinitely many pure strategies. For each $\theta \in \Theta$, the strategy of nature can be represented as a vector in $\mathbb{R}^I$, $s(\theta) = (R(d_1,\theta), R(d_2,\theta), ..., R(d_I,\theta))$. Define

$$
S := \{(R(d_1,\theta), R(d_2,\theta), ..., R(d_I,\theta)) \in \mathbb{R}^I | \theta \in \Theta\} \tag{22}
$$

and $M(i,s) = s_i$, then $\Gamma_P = (\mathcal{D}, S, M)$ is a S-game with payoff matrix $M(i,s)$. The index P stands for *pure* as we are only considering pure strategies. The mixed extension of the S-game is equivalent to $\Gamma_m = (\Delta, R, M)$, where Δ is the set of discrete probability distribution over $\mathcal{D}$, and R is the set of all countable convex linear combination of points in S. When S is closed and convex, our minimax problem is exactly to solve the best mixed strategy for player I.

(Blackwell and Girshick, 1954, Theorem 2.4.2) states that i) Every S game has a value, and the first player has a good (a minimax) strategy; and ii) If S is closed and convex, player II has a pure good strategy. Further, (Blackwell and Girshick, 1954, Theorem 1.8.1) indicates that if game Γ_P has a pure value, then its mixed extension Γ_m also has the same value.

Combining these results, we have that i) there exists $\bar{v}$,

$$\inf_{d\in\mathcal{D}} \sup_{\theta\in\Theta} R(d,\theta) = \sup_{\theta\in\Theta} \inf_{d\in\mathcal{D}} R(d,\theta) = \bar{v}, \tag{23}$$

$$\text{and} \quad \inf_{p\in\Delta(\mathcal{D})} \sup_{\theta\in\Theta} R(p,\theta) = \sup_{\theta\in\Theta} \inf_{p\in\Delta(\mathcal{D})} R(p,\theta) = \bar{v}, \tag{24}$$

and there exists *minimax* decision rule $p^* \in \Delta(D)$ such that

$$\sup_{\theta\in\Theta} R(p^*,\theta) = \bar{v}.$$

Further, consider mixed strategies for the nature,

$$\inf_{p\in\Delta(D)} \sup_{q\in\mathcal{P}(\Theta)} \int_\Theta R(p,\theta)dq(\theta) = \sup_{\theta\in\Theta} \inf_{q\in\mathcal{P}(\Theta)} \int_\Theta R(p,\theta)dq(\theta) = \bar{v}, \tag{25}$$

where $\mathcal{P}(\Theta)$ is the set of all mixed strategies of the nature. This means the assumption in Theorem 1 can be verified from our Assumptions. ii) If the set S defined in (22) is convex and closed, there exists exactly one θ^* that is *maxmin* strategy for nature, i.e.,

$$\inf_{p\in\Delta(D)} R(p,\theta^*) = \bar{v}.$$

Otherwise, the results in Blackwell and Girshick (1954) show that that the *maxmin* strategy for nature is supported on at most I points.

B.1 Maximin Strategy

Our proof for Theorem 1 also gives a surprising side result: When the game has a value, i.e. (25) holds, we can derive approximate max-min strategy for the nature.

Definition 2. For simplicity, we denote

$$R(p, q) := \int_{\Theta} R(p, \theta) dq(\theta).$$

A distribution $q_\epsilon^* \in \mathcal{P}(\Delta)$ is called an "ϵ-maximin" strategy for the game $(\Delta(\mathcal{D}), \mathcal{P}(\Theta), R(\cdot, \cdot))$ if

$$\inf_{p \in \Delta(D)} R(p, q_\epsilon^*) \geq \sup_{q \in \mathcal{P}(\Theta)} \inf_{p \in \Delta(D)} R(p, q) - \epsilon = \bar{v} - \epsilon.$$

In our proof, we derived an intermediate result that (17),

$$\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right) \leq \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{I} p_i R(d_i, \theta_t) + M^2 \eta + \frac{\ln(I)}{T\eta}, \forall p \in \Delta(\mathcal{D})$$

Recall assumption 2 states that for all t, $\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) = \sup_{\theta \in \Theta} R(p_t, \theta)$, so

$$\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{I} p_{i,t} R(d_i, \theta_t) \right) = \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in \Theta} R(p_t, \theta) \geq \inf_{p \in \Delta(\mathcal{D})} \sup_{\theta \in \Theta} R(p, \theta) = \bar{v},$$

we get

$$\bar{v} \leq \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{I} p_i R(d_i, \theta_t) + M^2 \eta + \frac{\ln(I)}{T\eta}, \forall p \in \Delta(\mathcal{D})$$

By taking $\eta = \epsilon / 2M^2, T = \lceil 4M^2 \ln(I) / \epsilon^2 \rceil$, we get

$$\bar{v} - \epsilon \leq \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{I} p_i R(d_i, \theta_t), \forall p \in \Delta(\mathcal{D}).$$

Now, if we choose q^*_ϵ to be a discrete distribution that

$$q^*_\epsilon(\theta) = \frac{|\{t \in [T] : \theta_t = \theta\}|}{T},$$

Then,

$$\inf_{p \in \Delta(\mathcal{D})} R(p,q) R(p,q^*_\epsilon) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I} p_i R(d_i,\theta_t) \geq \bar{v} - \epsilon = \sup_{q \in \mathcal{P}(\Theta)} \inf_{p \in \Delta(\mathcal{D})} R(p,q) - \epsilon,$$

which means q^*_ϵ is an "ϵ-maximin" strategy for the nature.