On the numerical approximation of minimax regret rules via fictitious play*

Patrik Guggenberger
Department of Economics
Pennsylvania State University

Jiaqi Huang
Department of Economics
Pennsylvania State University

First version December 2024. This version March 10, 2025

Abstract

Given the lack of analytical solutions for minimax regret treatment rules in most scenarios of empirical interest, finding numerical approximations is of key interest. To do so when potential outcomes are in $\{0,1\}$ we suggest discretizing the action space of nature and applying an algorithm based on Robinson's (1951) pioneering work on iterative solutions for two-person zero sum games with finite action space. Crucially, an application of Bayes' rule avoids the need to evaluate regret of each candidate treatment rule in each iteration. After every iteration of the algorithm we obtain a bound for the distance of the maximal regret of the current treatment rule to the minimax regret value. In the case where potential outcomes are in $[0,1]$ we apply the coarsening approach to the approximation of a minimax regret rule from the binary case.

As one application we consider a policymaker who has to choose between two treatments after observing a dataset with potentially unequal sample sizes per treatment. To significantly decrease computation time we leverage the general algorithm with theoretical insights about certain symmetry conditions that can be imposed on the treatment rules. As another application, we consider testing several innovations against the status quo.

Keywords: fictitious play, finite sample theory, minimax regret, numerical approximation, statistical decision theory, treatment assignment

JEL classification: C44

*We would like to thank Toru Kitagawa, Chuck Manski, and Pepe Montiel for very helpful discussion and references. We thank seminar participants in Berlin (Humboldt University), Cornell, Freiburg, and UC Irvine for helpful comments.

1 Introduction

Consider a policymaker who has to pick one of T treatments after being informed about treatment outcomes by a sample of size N from the population. If there is not one treatment policy δ that is best,[1] in expected outcome say, uniformly over all possible data generating processes (DGPs) then it is not unambiguous how an optimal treatment policy should be defined. Typically, there is not just one rule that is admissible. One could pursue a Bayesian analysis and optimize for a given prior over the space of all possible DGPs. One could pursue a max-min analysis that looks for a treatment rule that maximizes over all possible rules the minimal (over all DGPs) expected outcome. While the Bayesian route is subjective and would lead to potentially poor finite sample performance when the prior is false, the max-min criterion is often uninformative in the sense of all treatment policies being max-min. A growing literature therefore focuses on finding *minimax regret* rules, see Wald (1950), Savage (1954), and Manski (2004), that is, rules that minimize over all treatment policies the maximal regret over all DGPs, where regret for a given policy and DGP measures the gap between the best possible expected outcome and the expected outcome obtained for the chosen policy. Unfortunately, there are very few examples where minimax regret rules are analytically known and therefore in many examples of empirical interest they cannot currently be used by policymakers.[2] Manski (2021) concludes that "The primary challenge to use of statistical decision theory is computational." While there have been some contributions with regards to numerical implementation of minimax rules, see e.g. Chamberlain (2000) who provides an approach via convex minimization, the computational burden of the currently available algorithms, at least in the types of examples considered in the paper here, is overwhelming especially with the complexity of the space of DGPs, the number of actions by the policymaker or the sample size N increasing.[3]

[1] A treatment rule δ is a mapping from the set of possible samples onto the space of probability distributions on the set $\{1, ...T\}$, see (2.1) for a precise definition.

[2] Exceptions include e.g. Schlag (2006), Manski (2007), Stoye (2009, 2012), Tetenov (2012), Montiel Olea, Qiu, and Stoye (2023), Yata (2023), Chen and Guggenberger (2024), Kitagawa, Lee, and Qiu (2024), and additional references in these papers. The analytical finite sample results derived in these papers apply only under very restrictive assumptions, like an unrestricted (except for certain bounds) symmetric parameter space for the DGPs, an equal number of observations on outcomes for every treatment, or a normality assumption for observed outcomes in the partially identified examples. Guggenberger, Mehta and Pavlov (2024) derive minimax policies in the case where a policymaker is concerned with an α-quantile rather than expected outcome. See also Manski and Tetenov (2023). Robust Bayes methods constitute an additional approach and have been studied in the context of partially-identified models in Giacomini and Kitagawa (2022) and Aradillas Fernández, Montiel Olea, Qiu, Stoye, and Tinda (2024). Important references for optimal treatment choice in an asymptotic framework include Hirano and Porter (2009) and Kitagawa and Tetenov (2019).

[3] Meaningful progress has been made in some cases when one is willing to search for minimax regret rules in *constrained* classes of treatment rules. See e.g. Manski (2004), Manski and Tetenov (2016, 2019,

The objective of this paper is to provide a feasible procedure for the numerical approximation of minimax regret rules. We first provide an algorithm based on Robinson (1951) to approximate minimax regret rules that applies in the case where potential outcomes Y_t, for each $t = 1, ..., T$, are elements of $\{0, 1\}$ and the set of possible DGPs that nature chooses from has been discretized. In the case where $Y_t \in [0, 1]$ we obtain an approximation to a minimax regret rule via the so-called coarsening method applied to the solution from the binary case.

For the binary case, we adapt an algorithm from Robinson (1951) who considers a zero-sum two player game in which each player has finitely many actions and randomization over these actions is allowed for. The algorithm iteratively updates the strategy chosen by a player. More precisely, players take turns and each time when it is a player's turn she picks one of the finitely many actions that is a best response to the current mixed strategy of the other player that equals the empirical distribution of the strategies used by that player up to that period. The player then updates her strategy by setting it equal to the empirical distribution function over all best responses she played up to that point. Robinson (1951) shows that as the number of iterations goes to infinity the payoff that each player secures converges to the value of the game. In the game theory literature this approach is known as "fictitious play". In our scenario, the two players are the policymaker that picks a treatment rule and nature that picks the DGP with the payoff to the policymaker equal to negative regret. To conform with the setup in Robinson (1951) we discretize the set of possible DGPs that nature can choose from.[4] We show that as the discretization gets finer the minimax regret of the discretized game converges to the one of the original game. In the binary case, the set of nonrandomized treatment rules for the policymaker is automatically finite. Applied to our setting, Robinson's (1951) result implies that the maximal regret the policymaker might suffer approaches the minimax regret value under the sequence of treatment rules produced by the algorithm. We consider modifications of Robinson's (1951) approach by using updating weights proposed by Leslie and Collins (2006) that seem to lead to faster convergence of the algorithm.

In each step of our proposed algorithm a best response has to be calculated. Crucially, calculating the policymaker's best response to a given strategy by nature can be done, in many cases of interest, computationally trivially, via Bayes' rule and comparisons of the

2021), Kitagawa and Tetenov (2018), and Dominitz and Manski (2024). The latter paper obtains a simple expression for the maximum regret of a simple intuitive "midpoint prediction" for best prediction under square loss with missing data.

[4] We make the discretization of nature's action space explicit in our discussion. But note that any numerical approach to approximate minimax regret rules via an iterative algorithm that relies on repeatedly calculating best responses by nature through grid search also effectively uses discretization.

conditional means. We do so in our main applications. Note that the main computational bottleneck of other methods proposed in the literature, e.g. Chamberlain's (2000) approach based on convex optimization, is caused by the need to either evaluate the risk of all treatment rules or the risk of a given treatment rule for all possible actions by nature for every iteration of the algorithm. As illustrated below, in our examples we deal with cases where the policymaker may be tasked to randomize among 10^{10000} different treatment rules (or even many more). While our approach can successfully tackle that challenge any procedure that attempts to evaluate the risk of each treatment rule must necessarily fail due to the computational burden.

For a given strategy of the player we calculate nature's best response via grid search which poses the main computational hurdle. The computational complexity of our procedure (and any other procedure we are aware of) increases exponentially in the dimension of nature's action space. We are unaware of any particular structure, like convexity, to nature's maximization problem and currently use grid search to solve it.

Importantly, at every iteration step, the algorithm produces a bound for how far the maximal regret of the currently proposed treatment rule differs from the actual minimax regret value under discretization of nature's action space. This bound converges to zero as the number of iterations increases.

We also consider the case where the policymaker uses a priori information that restricts the parameter space of nature, for example, by restricting the possible set of means of Y_t. We show that this can be straightforwardly incorporated into our general framework. Also, the set of rules the policymaker may choose from can also be restricted in our framework to incorporate for example certain policy restrictions. The latter however may rule out the applicability of Bayes' rule and may thus lead to a substantial increase in the computation time.

In the case $Y_t \in [0, 1]$ we immediately obtain an approximate minimax regret rule via the *coarsening approach* (see e.g. Schlag (2006) and Stoye (2009)) and an approximate minimax regret rule from the binary case. We illustrate how to adapt the coarsening approach to various sampling designs, arbitrary number of treatments and incorporation of a priori information on possible DGPs by the policymaker.

In given applications of the proposed algorithm, we suggest searching for *symmetries* in the model that allow for further reduction of computation time. Namely, often it is possible to show that Nash equilibria of the game between the policymaker and nature can be found in a class of treatment rules restricted by certain symmetry conditions. This then implies that minimax regret rules can be found as well in that restricted class.

We study the following examples. In the first example, we consider the case of a policymaker who needs to pick one of two treatments after having observed N_t observations on each treatment, $t = 1, 2$, where N_1 and N_2 may be different, and aims at maximizing expected outcome. First, we consider the case where outcomes are either successes or failures, i.e. $Y_t \in \{0, 1\}$. It is shown that in this case minimax rules can be found in a class of treatment rules restricted by a symmetry condition that substantially reduces the class of rules that need to be searched over. In a nutshell, the "symmetric" rules δ considered are such that the value of $\delta(w_N)$ for a sample $w_N = (n_1, n_2)$ (where n_t denotes the number of successes for treatment t in the sample) already pins down the value of $\delta(w'_N)$ for another sample $w'_N = (n'_1, n'_2)$ such that $n_t + n'_t = N_t$ for $t = 1, 2$. That is, pinning down the values of a symmetric rule on about one half of the arguments already determines the entire rule. We then provide an algorithm that leverages the insights of fictitious play with the restrictions imposed by symmetry.[5] To give some measure about the computational complexity that is being tackled here, one can show that in the case where $N_1 = N_2 = 300$ there are $2^{N_1(N_2+1)/2+N_2/2} = 2^{45300} \approx 10^{13636.7}$ symmetric "nonrandomized" treatment rules one needs to search over.[6] By Stoye (2009, Corollary 1) it is known that the minimax regret value in that case equals .006940. In 51 minutes of computation time with 2000 iterations, the algorithm produces a treatment rule (a weighted average over the $10^{13636.7}$ symmetric nonrandomized treatment rules) whose maximal regret equals .006939! We illustrate that our algorithm can lead to substantial improvements in terms of maximal regret after quite few iterations relative to an empirical success rule.

The second example we study is testing innovations against a status quo. Stoye (2009) provides analytical expressions for minimax regret rules (given as solutions to a univariate equation) when one innovation is compared to the status quo, but there is no known analytical solution for the case of more than one innovation. As in the previous example, we derive certain symmetry conditions that help reduce the computational effort and then apply the general algorithm for that example. Additional applications appear in a Supplementary Appendix.

Relative to most of the extant literature on computational approaches to approximate minimax regret rules our approach avoids the most costly step of the algorithm that requires

[5]For a very restricted subcase we are able to provide analytical results for minimax rules, namely the case where $\min_{t=1,2}\{N_t\} = 0$, that is the case where one only observes data on outcomes from one treatment.

[6]For symmetric rules, defined properly in (3.4), one needs to set $\delta_2(n_1, n_2) = 1/2$ when $n_1/N_1 = n_2/N_2 = 1/2$. Therefore, when we say "nonrandomized" in the context of symmetric rules we mean $\delta_2(n_1, n_2) \in \{0, 1\}$ except for the case $n_1/N_1 = n_2/N_2 = 1/2$.

All computations reported in the paper are done using the Gauss software on a Dell desktop computer with an Intel(R) Core(TM) i7-14700 2.10 GHz Processor.

evaluation of risk over all treatment rules. Namely, consider a finite set of treatment rules with f elements to choose from and denote by p a discrete probability distribution over that set, i.e. a mixed strategy. It is well known that the function $h(p) = \sup_{s \in \mathbb{S}} R(p, s)$ is convex over the $f-1$-dimensional simplex, where R denotes regret (defined in (2.2) below) and $\mathbb{S}$ denotes the set of all possible distributions for potential outcomes. Chamberlain (2000) suggests using tools from convex programming to find a minimum for the function $h(p)$. However, this step is computationally costly when f is large. Instead, in our algorithm, when a best response by the policymaker needs to be calculated for the current mixed strategy used by nature, we rely on Bayes' rule. The computational hurdle is further reduced by dividing the problem into a computationally feasible numerical step and the coarsening trick.

Regarding numerical approaches for approximation of minimax regret rules, besides the early contribution by Chamberlain (2000), Masten (2023) considers nonlinear optimization packages like KNITRO and applies them to the case of *unbalanced* samples with $T = 2$ for several small sample sizes with $\max_{t \in \{1,2\}}\{N_t\} \leq 5$. See Dominitz and Manski (2024) for a comprehensive discussion of computational methods. Recently, Aradillas Fernández, Blanchet, Montiel Olea, Qiu, Stoye, and Tan (2024) consider a particular convex optimization routine called "mirror descent" to approximate ϵ-minimax rules. The policymaker iteratively updates her strategy using "multiplicative weights" based on the subgradient of h while nature computes a best response in each iteration to the policymaker's current strategy. For a bounded risk function it is shown that it takes $O(\ln f/\epsilon^2)$ iterations to obtain an ϵ-minimax rule. From results in Ben-Tal, Margalit, and Nemirovski (2001) it follows that the algorithm is optimal with respect to number of iterations among all iterative, first-order algorithms for convex optimization over the $f-1$-dimensional simplex up to the logarithmic factor $\ln f$. In contrast, we do not provide results on the maximal number of iterations our algorithm requires to produce an ϵ-minimax rule in a broad class of problems. We are concerned about the overall computational effort, which combines both the number of iterations and the computational effort in each iteration, in a given application. There are examples, where the convergence rate of fictitious play is rather slow, see e.g. Daskalakis and Pan (2014). But in the examples we consider in the paper, we find that the overall computational effort to obtain an ϵ-minimax rule a is very competitive. We provide a comparison with "mirror descent" in the context of our first example.

Daskalakis, Deckelbaum, and Kim (2024) focuses on computing the value of a game using the so-called "excessive gap technique". Using that approach requires optimization for each iteration of the algorithm which can be challenging in the applications that we consider given the strategy space for the policymaker is growing very fast. An additional related contribution regarding convex optimization is Bubeck (2015). An important reference for an algorithm to approximate a least favorable prior distribution is Kempthorne (1987).

Note that the general problematic one faces here is related to a testing scenario where several competing tests are available that all control the size. If the power curves of the tests intersect then which test to pick becomes a non-unambiguous task. Elliott, Mueller, and Watson (2015) consider finding tests that maximize weighted average power and in a scenario where the null hypothesis is composite they propose algorithms that find tests that are nearly optimal. It is shown that finding such optimal tests can be expressed as a minimax regret problem. Therefore, the numerical algorithm proposed in our paper might potentially be applied also in this scenario but would require several modifications

The paper is organized as follows. Section 2 provides a general description of the numerical procedure to approximate minimax regret rules based on fictitious play. In Subsection 2.1 the discretized model of nature's parameter space is introduced and subsection 2.2 discusses the algorithm to approximate minimax regret rules. Subsection 2.3 discusses how approximate minimax regret rules can be obtained in the case where outcomes live on the unit interval via the so-called coarsening approach of approximate minimax regret rules from the binary case. We then consider several examples of the general theory. Section 3 introduces the example of treatment choice with unbalanced sample sizes. Analytical minimax regret rules are provided for the special case where $\min_{t=1,2}\{N_t\} = 0$. Subsection 3.1 shows that there exists minimax regret rules that satisfy a symmetry condition. Subsection 3.2 specializes and leverages the insights from the general algorithm from Section 2 by incorporating symmetry. We discuss the results from the application of the algorithm in Subsection 3.3 where we also consider the case where the action space is restricted by a priori information of the policymaker. Finally in Section 4 we consider the example of "testing innovations" with more than one alternative. The proofs of all statements are given in the Appendix. A Supplementary Appendix contains further results on the examples in the paper.

2 The model, its approximation, and an algorithm based on "fictitious play"

A player has to choose one of several actions $t \in \mathbb{T} := \{1, 2, ..., T\}$ with randomization being allowed for. In many applications, the player is a policymaker and $\mathbb{T}$ is the set of treatments that she considers for a population.

By Y_t for $t \in \mathbb{T}$ we denote the random outcome if the player chooses action t. The vector of potential outcomes $(Y_t)_{t\in\mathbb{T}}$ is assumed to be an element in the set S. In several examples in the literature, $S = \{0, 1\}^T$ or $S = [0, 1]^T$, that is, the case, where the range of outcomes is the same under every action $t \in \mathbb{T}$, and where in the first case outcomes can be interpreted as success or failure. The algorithm we propose applies to the case where $S = \{0,1\}^T$.

The setup can be interpreted as a game of the player against an antagonistic nature that chooses a joint distribution $s \in \mathbb{S}$ for $(Y_t)_{t\in\mathbb{T}}$, a so-called "state of the world", where $\mathbb{S}$ denotes the set of possible distributions on S. Oftentimes, the player is assumed to care about expected outcome. For $t \in \mathbb{T}$ define $\mu_t = E_s Y_t$.

Before choosing an action, the player observes a sample $w_N = (t_i, y_{t_i,i})_{i=1,...,N}$ of size N with $y_{t_i,i}$, for given t_i, denoting an independent draw from $Y_{t_i,i}$ for $i = 1, ..., N$ (generated from a certain $s \in \mathbb{S}$). As in Stoye (2009) several *sampling designs* are possible. For example, under *fixed assignment*, assume integers N_t for $t \in \mathbb{T}$ are given and the sample w_N consists of N_t independent observations under each action $t \in \mathbb{T}$ with the sum of the N_t equal to N. Under *random assignment*, $P(t_i = t) = p_t$ for all $i = 1, ..., N$ and $t \in \mathbb{T}$, for some probabilities p_t whose sum over t equals 1. Lastly, consider the case of *testing innovations*, where the policymaker knows the mean outcome μ_T of treatment T of a status quo treatment. She needs to consider whether to switch from the status quo to one of $T-1$ alternative treatments whose mean outcome is unknown. The sample $w_N = (t_i, y_{t_i,i})_{i=1,...,N}$ consists of N_t independent observations under each action $t \in \{1, ..., T-1\}$ with the sum of the N_t for $t = 1, ..., T-1$ equal to N. Or alternatively one could generate the sample via random assignment.

Wlog we can assume that the marginal distributions of Y_t are independent of each other. Then, in the case where $S = \{0,1\}^T$ the joint distribution of $(Y_t)_{t\in\mathbb{T}}$ is fully pinned down by the mean vector $\mu = (\mu_1, ..., \mu_T) \in [0,1]^T$. Thus, from now on, in this case, an action by nature (rather than being presented as a joint DGP for the potential outcomes) is identified with picking a mean vector $\mu \in [0,1]^T$. The policymaker may have certain a priori knowledge that allows her to restrict μ to a subset $M \subset [0,1]^T$.

We next define the notion of a *treatment rule*. The task for the player is to choose a vector $\delta(w_N) \in [0,1]^T$ for every possible sample realization w_N, whose components add up to one and represent the probabilities with which action t is chosen. Namely, which treatment the player ultimately assigns is determined as an independent draw from the discrete random variable $B = B(\delta(w_N)) \in \mathbb{T}$ that equals t with probability $\delta_t(w_N)$ for $t \in \mathbb{T}$, where $\delta_t(w_N)$ denotes the t-th component of $\delta(w_N)$.

For a given sampling design, denote by $\mathbb{D}$ the set of rules δ that the policymaker can choose from. Unless otherwise specified, $\mathbb{D}$ is taken as the set of all (measurable) mappings

$$w_N \mapsto \delta(w_N) \in \Delta^{T-1} := \{x = (x_1, ..., x_T)' \in [0,1]^T,\ x_t \geq 0 \text{ for } t = 1, ..., T,\ \textstyle\sum_{t=1}^T x_t = 1\} \tag{2.1}$$

with $\delta(w_N)$ defining a probability distribution over the set $\mathbb{T}$. Note that we are therefore including randomized rules. In certain applications, $\mathbb{D}$ may be restricted. For example, Ishigara and Kitagawa (2024) use only nonrandomized linear aggregation rules to gain tractability. In empirical applications, policy restrictions may rule out certain elements in Δ^{T-1} from consideration.

For given $\delta \in \mathbb{D}$ and $s \in \mathbb{S}$ denote by $u(\delta, s)$ the payoff for the player. In this paper, $u(\delta, s) = E_s Y_{B(\delta(w_N))}$ i.e. the player cares about the expected outcome.[7] Given a choice $u(\delta, s)$ we use the optimality concept of *minimax regret* to choose a rule for the player. In particular, define regret of a rule δ when the state of nature is s as

$$R(\delta, s) = \max_{d \in \mathbb{D}} u(d, s) - u(\delta, s). \tag{2.2}$$

The objective is to find a rule δ^*, if one exists, that satisfies

$$\delta^* \in \arg\min_{\delta \in \mathbb{D}} \max_{s \in \mathbb{S}} R(\delta, s). \tag{2.3}$$

For the remainder of this paper, we assume that the player is concerned about expected outcome. Minimax regret rules have been derived analytically only in very few cases. We propose an algorithm to generate a treatment rule whose maximal regret gets arbitrary close to the minimax regret value. The algorithm is based on a result in Robinson (1951).

Robinson's (1951) algorithm

Robinson (1951) considers a finite two-player zero-sum game with payoff matrix $A = (a_{ij})$ for $i = 1, .., f$ and $j = 1, ..., g$. When the row-player (player 1 from now on) chooses action i and the column player (player 2 from now on) chooses action j then player 1 gets a_{ij} and player 2 gets $-a_{ij}$. Each player can choose a mixed strategy over her f and g, respectively, actions, that is, player 1 can choose a probability distribution p_i, $i = 1, .., f$, such that $p_i \geq 0$ and $\sum_{i=1}^{f} p_i = 1$ (and q_j, $j = 1, .., g$, such that $q_j \geq 0$ and $\sum_{j=1}^{g} q_j = 1$ for player 2, respectively.) By the minimax theorem there exists distributions such that the inequality (that holds for all distributions)

$$\min_j \sum_{i=1}^{f} a_{ij} p_i \leq \max_i \sum_{j=1}^{g} a_{ij} q_j \tag{2.4}$$

holds as an equality. Robinson (1951) provides an iterative procedure where in the n-th iteration, mixed strategies δ^n and ν^n for players 1 and 2, defined by $(p_{in})_{i=1,..,f}$ and

[7] Another possibility is $u(\delta, s) = q_{s,\alpha}(Y_{B(\delta(w))})$, where $q_{s,\alpha}(Y_{B(\delta(w))})$ denotes the α-quantile of $Y_{B(\delta(w))}$ when the state of the world is $s \in \mathbb{S}$, see Manski (1988).

$(q_{jn})_{j=1,...,g}$, respectively, are picked such that

$$\lim_{n\to\infty}\min_j \sum_{i=1}^{f} a_{ij}p_{in} = \lim_{n\to\infty}\max_i \sum_{j=1}^{g} a_{ij}q_{jn} \tag{2.5}$$

and the common limit equals the value of the game. The procedure works as follows. Start with any one of the f possible actions of player 1. The mixed strategy δ^1 puts weight one on that action. Pick any one of the g possible actions of player 2 that is a best response to δ^1. The distribution ν^1 puts weight one on that action. For the iteration, assume δ^n defined by $(p_{in})_{i=1,..,f}$ and ν^n defined by $(q_{jn})_{j=1,...,g}$ are given. Pick any one of the f possible actions of player 1 that is a best response to ν^n i.e. maximizes player 1's payoff. The mixed strategy δ^{n+1} is then defined as $\frac{n-1}{n}\delta^n + \frac{1}{n}I(\delta^n_{BR})$, where δ^n_{BR} denotes a best response by player 1 and $I(\cdot)$ denotes a point mass on the action in the brackets. Pick any one of the g possible actions of player 2 that is a best response to δ^{n+1}, i.e. maximizes player 2's payoff. The distribution ν^{n+1} is defined as $\frac{n-1}{n}\nu^n + \frac{1}{n}I(\mu^n_{BR})$, where μ^n_{BR} denotes a best response by player 2. Theorem 1 by Robinson (1951) establishes that this algorithm guarantees that (2.5) holds. Furthermore, Fudenberg and Levine (1998, Proposition 2.2) establish that if the distributions used in each round converge for both players, then the limits must constitute a mixed strategy Nash equilibrium.[8]

We first consider the binary case where $S = \{0,1\}^T$ and $\mu = (\mu_1, ..., \mu_T) \in M \subset [0,1]^T$. The policymaker and nature play the roles of players 1 and 2, respectively, in Robinson's (1951) setup above. The elements a_{ij} in the payoff matrix A represent negative regret when the policymaker and nature choose actions i and j, respectively. Player 1 aims to maximize negative regret while player 2 aims to minimize it.

In the next subsection we suggest discretizing nature's parameter space to conform with Robinson's (1951) requirement of only finitely many actions for both players. Given outcomes are assumed to be in $\{0,1\}$ and given there are T possible treatments on which data might be observed, there are only finitely many different samples w_N that can arise. Denote by W the number of different samples, $(w_{N1}, ...w_{NW})$ say, that can arise. Any treatment policy δ can then by summarized as a $T \times W$ matrix (or a TW vector), where the w-th column contains $\delta(w_{Nw}) \in \Delta^{T-1}$ for $w = 1, ..., W$. If one takes as the actions for the policymaker the set of nonrandomized treatment rules then this results in a finite set of actions[9] and mixing these actions one obtains all treatment rules as defined in (2.1). In some applications the

[8] This procedure may not work for games with generic payoffs or more than two players, because the empirical distribution of the strategies may not converge. See references in Fudenberg and Levine (1998).

[9] For example, if the sample contains N_t observations for treatment t, $t \in \mathbb{T}$ then $W = 2^N$, where $N = N_1 + ... + N_T$.and there are T^W many different nonrandomized treatment rules. That number grows very fast in both N and T.

finite set of actions may contain also randomized rules. In all applications considered below one needs to choose a finite set of actions, denoted by D, such that, when mixing over the actions in D is allowed for, then one obtains the set of all treatment rules $\mathbb{D}$ the policymaker can choose from, that is

$$\mathbb{D} = \{\textstyle\sum_{i=1}^{f} p_i \delta_i, \text{ for all } i = 1, ..., f, \ \delta_i \in D, \ p_i \geq 0, \text{ and } \sum_{i=1}^{f} p_i = 1\}. \tag{2.6}$$

The algorithm applies straightforwardly to any such set of action D and therefore policy restrictions can be easily implemented.

2.1 Approximation via discretization

In this subsection we define and discuss an ε-discretization of the action space of nature. Denote by $||\cdot||$ the Euclidean norm.

Definition (ε-discretization)**.** *We call a finite subset $M_\varepsilon \subset M$ an ε-discretization of the action space of nature if for any $\mu \in M$ we can find a $\mu_\varepsilon \in M_\varepsilon$ such that $\|\mu_\varepsilon - \mu\| < \varepsilon$.*

Comment. Often one can take M_ε as $\{\mu = (\mu_1, ..., \mu_T) \in M; \ \mu_t = \overline{n}_t \varepsilon$ for any $\overline{n}_t \in \mathbb{N}$ for all $t = 1, ..., T\}$ i.e. one takes an evenly-spaced grid with step size ε. However, such a construction will not always lead to an ε-discretization, for example, the resulting set M_ε could even be empty.

We now establish that as $\varepsilon \to 0$ the minimax regret value of the game in which nature is restricted to DGPs from an ε-discretization M_ε converges to the minimax regret value of the original game. We assume that the regret function $R(\delta, \cdot)$ is continuous in $\mu \in M$ uniformly in $\delta \in \mathbb{D}$, i.e. $\forall \lambda > 0$ there exists $\eta > 0$ such that when $\|\mu - \mu'\| < \eta$ for $\mu, \mu' \in M$ then for all $\delta \in \mathbb{D}$ we have $|R(\delta, \mu) - R(\delta, \mu')| < \lambda$.

Lemma 1 *Consider the case $S = \{0, 1\}^T$ and $\mu \in M$ for M compact. Define*

$$V \equiv \inf_{\delta \in \mathbb{D}} \max_{\mu \in M} R(\delta, \mu) \text{ and } V_m \equiv \inf_{\delta \in \mathbb{D}} \max_{\mu \in M_{\varepsilon_m}} R(\delta, \mu), \tag{2.7}$$

the values of the original and discretized games, respectively. Assume the regret function $R(\delta, \cdot)$ is continuous in μ uniformly in δ. Then, $\varepsilon_m \to 0$ as $m \to \infty$ implies

(i) $V_m \to V$ and

(ii) $\max_{\mu \in M} R(\delta_{\varepsilon_m}, \mu) \to V$ for any rule δ_{ε_m} that satisfies $V_m - \max_{\mu \in M_{\varepsilon_m}} R(\delta_{\varepsilon_m}, \mu) \to 0$.

Comments. 1. Continuity of $R(\delta, \cdot)$ is a weak assumption that we verify in several examples below. Together with compactness of M the condition implies that $\max_{\mu \in M} R(\delta, \mu)$

exists for any $\delta \in \mathbb{D}$. Note that in Lemma 1 we allow for the possibility that a minimax rule may not exist. We define the value V using the infimum over all treatment rules. Often the inf in the definition of V (and V_m) can be replaced by a min, e.g. if $\max_{\mu \in M} R(\delta, \mu)$ depends continuously on δ and $\mathbb{D}$ is compact.

2. Lemma 1(i) states that the minimax regret value for the case where nature's parameter space equals M_{ε_m} converges to the value of the game when the parameter space equals M as $\varepsilon_m \to 0$. Lemma 1(ii) states that if one uses a rule δ_{ε_m}, that is approximately minimax regret in the case where nature's parameter space is M_{ε_m}, in a situation where nature's parameter space equals M then the resulting minimax regret value converges to the value of the game with parameter space M as $\varepsilon_m \to 0$. That is, δ_{ε_m} is "asymptotically" minimax regret in the case where nature's parameter space equals M in the sense that maximal regret for the rule δ_{ε_m} converges to the minimax regret value as one lets the "stepsize" ε_m in the discretization go to zero. We are not claiming that the rule δ_{ε_m} converges to a minimax regret rule in the original model as $\varepsilon_m \to 0$, the convergence statement is about maximal regret. In this paper, we are proposing an algorithm that produces a treatment rule whose maximal regret with parameter space equal to M_{ε_m} converges to the minimax regret value of the discretized model (which in turn converges to the minimax regret value of the original model as $\varepsilon_m \to 0$.)

3. We make the discretization of nature's action space explicit in our discussion. But note that any numerical approach to approximate minimax regret rules via an iterative algorithm that relies on repeatedly calculating best responses by nature through grid search also effectively uses discretization.

2.2 Algorithm and its convergence properties

In this subsection we describe in detail the algorithm based on Robinson (1951) used to approximate minimax regret rules in the discretized setup, that is, we consider the case where $S = \{0, 1\}^T$ and $\mu = (\mu_1, ..., \mu_T) \in M_\varepsilon$ for a small $\varepsilon > 0$. The discretization M_ε of M is kept fixed throughout the iterations. We also allow for a possible modification suggested by Leslie and Collins (2006) regarding the weights used in the updating of the strategies. Recall that as specified in (2.6) we denote by D a finite set of actions, such that, when mixing over the actions in D is allowed for, then one obtains the set of all treatment rules $\mathbb{D}$ the policymaker can choose from. For example, D could be chosen as the set of nonrandomized treatment rules which is finite and satisfies (2.6).

The algorithm provides a sequence of treatment rules and strategies by nature. For $n \geq 1$ we denote by δ^n the treatment rule in the n-th iteration, by ν^n the empirical distribution of strategies for nature up to time n, and by R^n a lower bound on the minimax regret value obtained in the n-th iteration. To apply the algorithm, pick a small threshold $\xi > 0$ and a sequence of updating weights α_n for $n = 1, 2, ...$ that will be specified further below.

Algorithm 1 ***Initialization:***

i) Define δ^1 as one of the actions in D.

ii) Initialize $\nu^0 = 0$ and $R^0 = 0$.

Iteration. For $n = 1, 2, 3...$ ***DO:***

i) Find a best response $\mu_{BR}^n \in M_\varepsilon$ by nature to δ^n that is, a mean vector that maximizes regret given δ^n :

$$R(\delta^n, \mu_{BR}^n) = \max_{\mu \in M_\varepsilon} R(\delta^n, \mu). \tag{2.8}$$

ii) If

$$R(\delta^n, \mu_{BR}^n) - R^{n-1} < \xi \tag{2.9}$$

*then **break.** Use the rule δ^n.*

iii) Update nature's mixed strategy to be a weighted average of strategies used up to this point, i.e.

$$\nu^n = (1 - \alpha_n)\nu^{n-1} + \alpha_n I(\mu_{BR}^n), \tag{2.10}$$

where $I(x)$ denotes a point mass of size 1 at the point $x \in [0, 1]^T$.

iv) Compute a best response δ_{BR}^n in D by the policymaker to ν^n, that is,

$$R^n := R(\delta_{BR}^n, \nu^n) = \min_{\delta \in D} R(\delta, \nu^n). \tag{2.11}$$

v) Update the treatment rule by

$$\delta^{n+1} = (1 - \alpha_{n+1})\delta^n + \alpha_{n+1}\delta_{BR}^n. \tag{2.12}$$

Comments. 1. Regarding the iteration step iv) note that for a best response δ_{BR}^n against ν^n it is necessary and sufficient that whenever

$$E(Y_t|w_N) > E(Y_{t'}|w_N) \text{ for } t, t' \in \mathbb{T} \tag{2.13}$$

holds, δ_{BR}^n gives zero weight to the "dominated" treatment t'.[10] Importantly, an application of **Bayes' rule** then allows for a straightforward solution to the minimization problem in iv) that provides a very significant computational advantage relative to using convex optimization in more general setups. Namely, assume nature picks the mixed strategy ν^n

[10] One can always find a nonrandomized best response δ_{BR}^n. Unless of course one imposes certain restrictions on the class of treatment rules that impose randomization.

that puts nonzero weights p_m on exactly the mean vectors $\overline{\mu}_m = (\mu_{1m}, ..., \mu_{Tm}) \in M_\varepsilon$ for $m = 1, ..., \overline{m}$. Then,

$$
\begin{aligned}
&E(Y_t|w_N) \\
&= P(Y_t = 1|w_N) \\
&= \textstyle\sum_{m=1}^{\overline{m}} P(Y_t = 1 \text{ \& nature picks } \overline{\mu}_m|w_N) \\
&= \textstyle\sum_{m=1}^{\overline{m}} P(Y_t = 1|\text{nature picks } \overline{\mu}_m \text{ \& } w_N)P(\text{nature picks } \overline{\mu}_m|w_N) \\
&= \textstyle\sum_{m=1}^{\overline{m}} P(Y_t = 1|\text{nature picks } \overline{\mu}_m)P(w_N|\text{nature picks } \overline{\mu}_m)p_m/P(w_N) \\
&= \textstyle\sum_{m=1}^{\overline{m}} \mu_{tm} P(w_N|\text{nature picks } \overline{\mu}_m)p_m/P(w_N),
\end{aligned}
\tag{2.14}
$$

and therefore (2.13) holds iff

$$
\textstyle\sum_{m=1}^{\overline{m}} \mu_{tm} P(w_N|\text{nature picks } \overline{\mu}_m)p_m > \sum_{m=1}^{\overline{m}} \mu_{t'm} P(w_N|\text{nature picks } \overline{\mu}_m)p_m. \tag{2.15}
$$

For a given sample w_N and choices of p_m and $\overline{\mu}_m$ for $m = 1, ..., \overline{m}$ the expressions in (2.15) can be easily calculated (where $P(w_N|\text{nature picks } \overline{\mu}_m)$ obviously depends on the sampling design that is being employed).

2. Choosing the weights $\alpha_n = 1/n$ corresponds to the algorithm proposed in Robinson (1951). In that case, every iteration involves optimizing against the empirical distribution function of strategies chosen up to this point. Other choices of α_n satisfying $\alpha_n > 0$, $\alpha_n \to 0$, and $\sum_{n=1}^{\infty} \alpha_n = \infty$ have been suggested in Leslie and Collins (2006), e.g. $\alpha_n = (C + n)^{-\eta}$ for some constants $\eta \in (0, 1]$ and $C > 0$ (which nests $\alpha_n = 1/n$ when $C = 0$ and $\eta = 1$) or $\alpha_n = 1/\log(C+n)$. Based on our experience in several examples, we find that $\alpha_n = (C+n)^{-\eta}$ with $\eta = .7$ and $C = 5$ works quite well.

3. In the applications of the algorithm considered below regret goes to zero as the sample size N goes to infinity. Therefore, a "small" threshold ξ in one setting may not be small in another setting. It therefore makes sense to modify the break command in (2.9) by also making sure that the approximation error relative to the actual regret is small, that is to check whether $(R(\delta^n, \mu_{BR}^n) - R^{n-1})/R^{n-1} < \xi$.

4. In practice, after the algorithm terminates, one does not necessarily use the treatment rule obtained in the last iteration. One tracks the maximum regret level for the treatment rules generated during the entire iteration process, and chooses the treatment rule that has the minimum level of maximum regret among those rules. This is because there is no guarantee that maximal regret will decrease in each iteration, so it is possible that the rule generated in the last iteration is not the best one amongst the rules considered during all iterations.

5. **Bounds:** Note that $R^n = R(\delta^n_{BR}, \nu^n)$ denotes the minimal regret against a particular mixed strategy by nature and therefore is nonbigger than the minimax regret value. Analogously, $R(\delta^n, \mu^n_{BR})$ denotes maximal regret against a particular mixed strategy by the policymaker and therefore is nonsmaller than the minimax regret value. If $R^n = R(\delta^n_{BR}, \nu^n)$ and $R(\delta^n, \mu^n_{BR})$ are close to each other, they are both necessarily close to the minimax regret value of the discretized model. In any iteration of the algorithm $R(\delta^n, \mu^n_{BR}) - R^n$ provides an upper bound for how far away the maximal regret of the currently considered treatment rule is from the minimax regret value.

If one lets n go to infinity (and never uses the break command in step ii) of the iteration) then by Robinson's (1951, Theorem 1) in the case of $\alpha_n = 1/n$ and by Leslie and Collins (2006, Corollary 5) for the other cases in the proposition below, R^n converges to the minimax regret value of the discretized game. We formulate this important result as a proposition.

Proposition 1 *Assume the sequence α_n satisfies $\alpha_n > 0$, $\alpha_n \to 0$, and $\sum_{n=1}^{\infty} \alpha_n = \infty$. Then the maximal regret from the sequence of treatment rules that is generated from Algorithm 1, $\max_{\mu \in M_\varepsilon} R(\delta^n, \mu)$, converges as $n \to \infty$ to the minimax regret value of the ε-discretized model.*

Comment. 1. Leslie and Collins (2006) provide some evidence that for the more general choices of weights α_n one might be able to speed up the convergence rate relative to the choice $\alpha_n = 1/n$. An important topic for future research is to theoretically work out the convergence speed as a function of the chosen weights.

2.3 Coarsening

Assume one has found a minimax regret rule δ (or approximate minimax regret rule δ_ε) in the setup with $S = \{0,1\}^T$ and mean vector $\mu \in M \subset [0,1]^T$ for a certain sampling design, such as fixed assignment with N_t observations for $t \in \mathbb{T}$ or random assignment with $P(t_i = t) = p_t$ for all $i = 1, ..., N$ and $t \in \mathbb{T}$. We show that then the so-called coarsening approach can be used to generate a minimax regret rule (or approximate minimax regret rule) for the case $S = [0,1]^T$ and the same restriction on the mean vector. The "coarsening trick" proceeds as follows:

Coarsen the outcomes $y_{t_1,1}, ..., y_{t_N,N}$ in the sample $w_N = (t_i, y_{t_i,i})_{i=1,...,N}$ by replacing each $y_{t_i,i}$ by an independent draw $\widetilde{y}_{t_i,i}$ from $\widetilde{Y}_{t_i,i} \in \{0,1\}$ from a Bernoulli random variable with success probability $y_{t_i,i}$ for $i = 1, ..., N$. Denote by $\widetilde{w}_N = (t_i, \widetilde{y}_{t_i,i})_{i=1,...,N}$ the coarsened sample. Define a new treatment rule δ^C by setting $\delta^C(w_N) = \delta(\widetilde{w}_N)$ (or by $\delta^C_\varepsilon(w_N) = \delta_\varepsilon(\widetilde{w}_N)$).

Part i) of the following proposition states that then the treatment rule δ^C is minimax regret with $S = [0,1]^T$ with mean restriction M. The coarsening approach has been introduced when $T = 2$ and $M = [0,1]^T$, see, e.g. Cucconi (1968), Gupta and Hande (1992), Schlag (2003, 2006), and Stoye (2009), and applied with arbitrary T in Chen and Guggenberger (2024) but it can also be employed here with mean restriction M. Part ii) of the proposition shows that coarsening can also be applied to approximate minimax regret rules in the case $S = \{0,1\}^T$ and given M to obtain approximate minimax regret rules in the case $S = [0,1]^T$ with same mean restriction M.

Proposition 2 *Assume the sample is generated by fixed, random assignment or testing innovations.*

i) Assume δ *is a minimax regret rule when* $S = \{0,1\}^T$ *and* $\mu \in M \subset [0,1]^T$*. Then the treatment rule* δ^C *defined by* $\delta^C(w_N) = \delta(\widetilde{w}_N)$ *is minimax regret with* $S = [0,1]^T$ *and* $\mu \in M \subset [0,1]^T$.

ii) Assume for some $\varepsilon > 0$, δ_ε *is an* ε*-approximate minimax regret rule when* $S = \{0,1\}^T$ *and* $\mu \in M \subset [0,1]^T$*, meaning,* $\min_{\delta \in \mathbb{D}} \max_{\mu \in M} R(\delta,\mu) + \varepsilon \geq \max_{\mu \in M} R(\delta_\varepsilon, \mu)$*. Then the treatment rule* δ_ε^C *defined by* $\delta_\varepsilon^C(w_N) = \delta_\varepsilon(\widetilde{w}_N)$ *is an* ε*-approximate minimax regret rule when* $S = [0,1]^T$ *and* $\mu \in M \subset [0,1]^T$*, meaning,* $\min_{\delta \in \mathbb{D}} \max_{s \in \mathbb{S} \text{ s.t. } \mu \in M} R(\delta, s) + \varepsilon \geq \max_{s \in \mathbb{S} \text{ s.t. } \mu \in M} R(\delta_\varepsilon^C, s)$.[11]

Comments. 1. The proof of Proposition 2 i) adapts the proofs in Schlag (2006) and Stoye (2009) from the case with equal sample sizes, $T = 2$ and $M = [0,1]^2$ to the case considered here with arbitrary T and M and the N_t not necessarily all equal. The proposition could likely be generalized to alternative sampling designs.

2. Proposition 2 ii) is an important result. It establishes that approximate minimax regret rules can be obtained for complex problems through a combination of a numerical approach applied in a simpler setup and the coarsening approach. In particular, it applies to the setup of the example considered in Section 3.

3 Treatment assignment with unbalanced samples

In this section we look at an important special case of the general setup in the previous section, namely the case of a policymaker who needs to choose between two treatments after being informed by a sample consisting of N_t independent observations from each treatment $t = 1, 2$. That is, here, $T = 2$ and $\mathbb{T} = \{1, 2\}$. We assume that the policymaker's objective is

[11] Recall that by $\mathbb{S}$ denotes the set of possible distributions for $(Y_1, ...Y_T)$ on S. Recall also that wlog we can assume that the marginals Y_t for $t \in \mathbb{T}$ are independent.

expected outcome, $u(\delta, s) = E_s Y_{B(\delta(w_N))}$ and that realizations are either successes or failures, that is $S = \{0,1\}^2$. We can then apply the coarsening approach to deal with the general case $S = [0,1]^2$. Given that observations are independent wlog we can assume that under $s \in \mathbb{S}$ the marginals (Y_1, Y_2) are independent and that therefore the joint distribution is fully pinned down by the mean vector $(\mu_1, \mu_2) \in M = [0,1]^2$. We first consider the case here where the policymaker is completely agnostic and does not have any a priori information about the mean vector. Then, in Section 3.3.1 we apply the algorithm to find approximate minimax regret rules when M is a strict subset of $[0,1]^2$. From now on, we identify any $s \in \mathbb{S}$ with its resulting mean vector.

An analytical formula for minimax rules is known *only* in the case where the N_t are equal, that is, in the case of "matched treatment assignment". The following statement holds for any integer $T \geq 2$.

Proposition 3 *Take $\mathbb{S}$ as the set of all distributions on $\{0,1\}^T$. Denote by n_t the number of successes for treatment t in the sample. Define*

$$W = \arg\max_{t \in \mathbb{T}} n_t \tag{3.1}$$

the set of all those treatments for which most successes are observed in the sample. Define

$$\delta_t(w_N) = 1/|W| \text{ if } t \in W \text{ and } \delta_t(w_N) = 0 \text{ otherwise,} \tag{3.2}$$

where for a set W we denote by $|W|$ the number of elements in W and δ_t for $t \in \mathbb{T}$ denotes the t-th component of the treatment rule δ. If the N_t are all equal then $\delta \in \mathbb{D}$ is a minimax regret treatment rule.

Comments. 1. Proposition 3 was proven by Schlag (2006) and Stoye (2009) for the case $T = 2$ and for arbitrary finite T by Chen and Guggenberger (2024). The proof uses the well-known Nash equilibrium approach, see e.g. Berger (1985), that is based on the insight that the action of the policymaker in any Nash equilibrium in the zero sum game between the policymaker and nature (with nature's payoff being regret and the policymaker's payoff being negative regret) is automatically a minimax regret rule. See Chen and Guggenberger (2024) for a reproduction of the simple proof of that result. Note that in the zero sum game nature is allowed to randomize over its actions where an action by nature is simply a vector of means $(\mu_1, \mu_2) \in [0,1]^2$. From now on we denote by $\Delta\mathbb{S}$ the space of randomized rules for nature.[12]

[12]Note though that being part of a Nash equilibrium is a sufficient but not a necessary condition for a

When the N_t are not all the same then no analytical results are known about minimax regret rules - even in the case $T = 2$. We make a tiny bit of progress in that regard and next provide an analytical formula for minimax regret rules in the special case where $\min_{t=1,2}\{N_t\} = 0$. Wlog in the following proposition assume $N_1 = 0$ and $N_2 > 0$.

Proposition 4 *In the case where $N_1 = 0$ and $N_2 > 0$ the treatment rule δ defined by $\delta_2(0, n_2) = n_2/N_2$ is minimax regret.*

Comments. 1. The proof establishes that a least favorable distribution by nature is given by randomizing uniformly between the two mean vectors $(0, 1/2)$ and $(1, 1/2)$.

2. Note that for any n_2 and n_2' in $\{0, 1..., N_2\}$ such that $n_2 + n_2' = N_2$, we have $\delta_2(0, n_2) + \delta_2(0, n_2') = (n_2 + n_2')/N_2 = 1$. Thus, the treatment rule satisfies the type of symmetry condition that we formally introduce below in (3.3).

3. We put substantial effort in deriving analytical derivations for minimax regret rules in more general cases but were not successful. Therefore, in what follows below, we will instead approximate minimax regret rules by a version of the algorithm from Section 2 based on fictitious play.

In the next subsection, we show that minimax regret rules can be found in a class of treatment rules restricted by a symmetry condition. We then provide an algorithm that leverages the insights of fictitious play with the restrictions imposed by symmetry, and finally report results from the algorithm for several choices of (N_1, N_2).

From now on, we represent a sample w_N by the number of successes n_t for each treatment, that is, (with some abuse of notation) we write $w_N = (n_1, n_2)$.

3.1 Symmetry restriction

In the context of treatment assignment with unbalanced samples when $T = 2$ and $S = \{0, 1\}^2$ we make the following definition.

Definition (symmetric treatment rule). *We call a treatment rule δ symmetric when*

$$\delta_2(w_N) + \delta_2(w_N') = 1 \tag{3.3}$$

whenever $w_N = (n_1, n_2)$ and $w_N' = (n_1', n_2')$ satisfy

$$n_t + n_t' = N_t \text{ for } t = 1, 2. \tag{3.4}$$

minimax rule. See Guggenberger, Mehta, and Pavlov (2024) for an example of a minimax rule that is not part of a Nash equilibrium.

Comments. 1. Note that when δ is symmetric then trivially also $\delta_1(w_N) + \delta_1(w'_N) = 1$ for all $w_N = (n_1, n_2)$ and $w'_N = (n'_1, n'_2)$ that satisfy (3.4). Further note, that when both N_1 and N_2 are even, then necessarily $\delta_2(w_N) = 1/2$ for $w_N = (N_1/2, N_2/2)$. Therefore, symmetric rules sometimes need to randomize.

2. Symmetry conditions are used in the analytical derivations of minimax rules in Stoye (2009, Proposition 1) and Chen and Guggenberger (2024, Proposition 1) but the symmetry condition defined here in the particular context of unbalanced samples appears to be new.

3. The main purpose of the symmetry condition employed here is to gain computational improvements. Note that if δ is symmetric then knowing $\delta_2(w_N)$ automatically fixes the value for $\delta_2(w'_N)$. E.g. if $N_1 = 1$ and $N_2 = 5$ then knowing the value of $\delta_2(0, n_2)$ for $n_2 = 0, 1, ..., 5$ pins down the value of $\delta_2(1, 5-n_2)$. Therefore rather than searching over all $2 \times 6 = 12$ values of $\delta_2(n_1, n_2)$ for $n_1 = 0, 1$ and $n_2 = 0, 1, ..., 5$ one only needs to consider values for the 6 values of $\delta_2(0, n_2)$. If for instance one were to employ an algorithm whose computation time increases exponentially with $(n_1 + 1)(n_2 + 1)$ then searching in the class of symmetric rules instead would reduce the computation time by a factor $\exp((n_1+1)(n_2+1))/\exp(.5(n_1+1)(n_2+1))$.

We next show why symmetry is an important condition. The next proposition establishes that one can find an action pair (δ, s) with $\delta \in \mathbb{D}$ and $s \in \Delta\mathbb{S}$ for the policymaker and nature, respectively, that constitutes a Nash equilibrium and where δ is symmetric.

Proposition 5 *(i) Suppose nature plays a mixed strategy that picks the mean vectors* $\overline{\mu}_m = (\mu_{1m}, \mu_{2m})$ *with probability* $p_m > 0$ *for* $m = 1, ..., 2\overline{m}$ *for some* $\overline{m} \geq 1$, *where* $p_1 + ... + p_{2\overline{m}} = 1$ *and for all* $m = 1, ..., \overline{m}$ *we have*

$$(\mu_{1m}, \mu_{2m}) = (1 - \mu_{1m+\overline{m}}, 1 - \mu_{2m+\overline{m}}) \text{ and } p_m = p_{m+\overline{m}}. \tag{3.5}$$

Then if there exists a treatment rule δ *that is a best response to that strategy (defined by* $\overline{\mu}_m$ *and* p_m *for* $m = 1, ..., 2\overline{m}$*) then it can be chosen to satisfy (3.3).*

(ii) Assume the policymaker chooses a symmetric treatment rule δ *and* (μ_1, μ_2) *is a best response by nature. Then also* $(1 - \mu_1, 1 - \mu_2)$ *is a best response.*

Comments. 1. In the setup of Proposition 5(i) nature randomizes over mean vectors that come in pairs. For every mean vector (μ_{1m}, μ_{2m}) that nature picks with positive probability the mean vector $(1 - \mu_{1m}, 1 - \mu_{2m})$ is also picked with the same probability. Together with Proposition 5(ii) it follows that a Nash equilibrium of the unrestricted game exists in which the policymaker plays a symmetric treatment rule. The proof of Proposition 5(ii) establishes that if δ is symmetric then $R(\delta, (1 - \mu_1, 1 - \mu_2)) = R(\delta, (\mu_1, \mu_2))$.

2. Proposition 5 implies that if there exists a minimax regret rule that is part of a Nash equilibrium then

$$\inf_{\delta\in\mathbb{D}}\max_{\mu\in[0,1]^2} R(\delta,\mu) = \inf_{\substack{\delta\in\mathbb{D}\\ \delta \text{ is symmetric}}} \max_{\substack{s\in\Delta\mathbb{S}\\ s \text{ satisfies (3.5)}}} R(\delta,\mu). \tag{3.6}$$

A mixed strategy $s \in \Delta\mathbb{S}$ by nature is a random vector with outcomes in $[0,1]^2$ that specifies which mean vector nature chooses. The condition (3.5) restricts this random variable to only pick a certain number of pairs (μ_1, μ_2) and $(1-\mu_1, 1-\mu_2)$ whose two elements have to be chosen with equal probabilities. In the next subsection we are applying Algorithm 1 to the formulation of the problem on the right side of the equality in (3.6) using a discretized version of nature's parameter space.

3. The symmetry condition is relevant in the case where M satisfies the following restriction. If

$$(\mu_1, \mu_2) \in M \text{ then } (1-\mu_1, 1-\mu_2) \in M. \tag{3.7}$$

That restriction is satisfied in the lead example where $M = [0,1]^2$.

4. An important question for current research concerns the development of techniques that allow one to systematically determine the set of all symmetry conditions in this and any other application. We came across the above symmetry condition after we applied the algorithm for small sample sizes and noted a pattern in the obtained approximation of minimax rules. We don't know whether there are other symmetry conditions that we have not yet discovered.

3.2 Algorithm

To justify the discretization of nature's action space below, we first establish that in the example considered here, the assumption of Lemma 1 about the regret function $R(\delta,\cdot)$ being continuous in $\mu \in M$ uniformly in $\delta \in \mathbb{D}$ is satisfied.

Lemma 2 $\forall\lambda > 0$ *there exists* $\eta > 0$ *such that when* $\|\mu - \mu'\| < \eta$ *for* $\mu, \mu' \in M$ *then for all* $\delta \in \mathbb{D}$ *we have* $|R(\delta,\mu) - R(\delta,\mu')| < \lambda$.

For any $\varepsilon = p^{-1}$ for $p \in \mathbb{N}$ we will use the set

$$M_\varepsilon = \{(\mu_1, \mu_2); \mu_j = p_j/p \text{ for some } p_j \in \{0, 1, ..., p\} \text{ for } j = 1, 2\} \tag{3.8}$$

as the ε-discretization of the set $(\mu_1, \mu_2) \in M = [0,1]^2$. That guarantees that if $(\mu_1, \mu_2) \in M_\varepsilon$ then also $(1-\mu_1, 1-\mu_2) \in M_\varepsilon$.

Next, we apply the general Algorithm 1 from Subsection 2.2 to the problem considered here, assuming that (3.7) holds. Using the insights from Proposition 5 as expressed in (3.6) we exploit the particular symmetrical structure of the current application to simplify the complexity of the problem. Namely, we take D as the finite set of symmetric rules with outcomes in $\{0,1\}$ except for samples $w_N = (n_1, n_2)$ for which $n_1 = N_1/2$ and $n_2 = N_2/2$ (which can only happen when both sample sizes are even). For nature, the finite set of actions, M_ε^{Na} say, consists of those mixed strategies that uniformly randomize between a pair of (μ_1, μ_2) and $(1-\mu_1, 1-\mu_2)$ with $(\mu_1, \mu_2) \in M_\varepsilon$. The algorithm then follows the exact same steps as in the Algorithm 1 but with the particular choices of D and finite set of actions for nature as just described. For clarity, we write down the entire algorithm again.

Algorithm 2 ***Initialization:***

i) For δ^1 use the following "empirical success rule," $\delta_2^1(n_1, n_2) = 1$ if $n_1/N_1 < n_2/N_2$, $\delta_2^1(n_1, n_2) = 0$ if $n_1/N_1 > n_2/N_2$, $\delta_2^1(n_1, n_2) = 1$ if $n_1/N_1 = n_2/N_2 > 1/2$, $\delta_2^1(n_1, n_2) = 0$ if $n_1/N_1 = n_2/N_2 < 1/2$, and $\delta_2^1(n_1, n_2) = 1/2$ otherwise (that is, when $n_1/N_1 = n_2/N_2 = 1/2$).

ii) Initialize $\nu^0 = 0$ and $R^0 = 0$.

***Iteration: For** $n = 1, 2, 3...$ **DO**:*

i) Find a best response $\mu_{BR}^n \in M_\varepsilon^{Na}$ to δ^n for nature, that is, a mixed strategy with equal weights for μ^n and $(1,1) - \mu^n$ for a $\mu^n \in M_\varepsilon$. Let

$$R(\delta^n, \mu_{BR}^n) = \max_{\mu \in M_\varepsilon^{Na}} R(\delta^n, \mu). \tag{3.9}$$

ii) If

$$R(\delta^n, \mu_{BR}^n) - R^{n-1} < \xi \tag{3.10}$$

*then **break.** Use the rule δ^n.*

iii) Update nature's mixed strategy to be the following weighted average

$$\nu^n = (1-\alpha_n)\nu^{n-1} + (\alpha_n/2)(I(\mu^n) + I((1,1) - \mu^n)), \tag{3.11}$$

where $I(x)$ denotes a point mass of size 1 at the point $x \in [0,1]^2$.

iv) Compute a best response $\delta_{BR}^n \in D$ by the policymaker to ν^n, that is,

$$R^n := R(\delta_{BR}^n, \nu^n) = \min_{\delta \in D} R(\delta, \nu^n). \tag{3.12}$$

v) Update the treatment rule by

$$\delta^{n+1} = (1-\alpha_{n+1})\delta^n + \alpha_{n+1}\delta_{BR}^n. \tag{3.13}$$

Comment. 1. To implement the algorithm, one has to chose $p = 1/\varepsilon \in \mathbb{N}$. That choice, obviously, is a trade-off between computational effort and closeness of the minimax regret value in the original model and the ε-discretized one. By increasing p nature becomes "more powerful" and thus the minimax regret value in the ε-discretized model increases in p and (as shown in Lemma 1) converges to the minimax regret value of the original model as $p \to \infty$. Each iteration step i) is solved by a grid search over all $(p+1)^2$ possible choices of mean vectors.

2. Note that δ^1 satisfies the symmetry condition. It is randomized when both N_1 and N_2 are even for samples w_N such that $n_1/N_1 = n_2/N_2 = 1/2$. Alternatively, one could initiate the algorithm with any other symmetric treatment rule. We suggest picking the empirical success rule because it is a good starting point as it is known that its maximal regret converges to zero at the optimal rate, see Kitagawa and Tetenov (2018). Note that ν^n satisfies the condition given in the Proposition 5(i).

3. The best response of the policymaker is calculated via Bayes rule as in (2.15) and uses the formula

$$P(w_N|\text{nature picks } \overline{\mu}_m) = \prod_{t=1}^{2} \binom{N_t}{n_t} \mu_{tm}^{n_t}(1-\mu_{tm})^{N_t - n_t} \tag{3.14}$$

when $w_N = (n_1, n_2)$.

4. Given this is a special case of the general setup, Proposition 1 about convergence of the maximal regret of the sequence of treatment rules to the minimax regret value of the ε-discretized model continues to hold.

5. When potential outcomes are elements of $[0,1]$ rather than $\{0,1\}$ one applies the coarsening approach to the obtained treatment rule from Algorithm 2.

6. If there are discrete covariates X in the model then an important result in Stoye (2009) states that one obtains a minimax regret rule by simply solving the conditional-on-X problems.

3.3 Results

In this subsection, we investigate the performance of Algorithm 2 for the example of unbalanced samples with $T = 2$.

We first consider the case of **balanced samples** where $N_1 = N_2$. The reason is that in that case Proposition 3 provides the minimax regret rule in analytical form and Stoye (2009, Corollary 1) provides a formula to calculate the minimax regret value. The latter allows us to directly assess the performance of the algorithm. We also use the case of balanced samples to experiment with different weighting schemes α_n and initializations of the treatment rule.

Specifically, we consider $N_1 = N_2 \in \{5, 10, 20, 40, 60, 80, 100, 200\}$. In the discretization of the parameter space for nature in (3.8) we use $p = 1000$ throughout.[13]

First we experiment with different choices η and C in the weighting scheme $\alpha_n = (C+n)^{-\eta}$ discussed in Leslie and Collins (2006) when $N_1 = N_2 = 40$. Specifically, we consider a grid of values for $\eta \in [.1, 1]$ and for $C \in [0, 10]$. Note that the particular choice $C = 0$ and $\eta = 1$ leads to the weights proposed in Robinson (1951).[14] This experiment suggests that $\eta \in [.5, .7]$ and $C \in [5, 10]$ work "best." We will only report results for $\alpha_n = 1/n$ and $\alpha_n = (5 + n)^{-.7}$ below. Those results will document the importance that the weighting scheme has on the convergence speed of maximal regret to the minimax regret value.

We also experiment with two different initializations δ^1 of the algorithm, namely, we initialize with the particular empirical success (ES) rule suggested in Algorithm 2. However, given that this initialization is very close to the actual minimax regret rule in the case of balanced samples, we also consider one that is very different from the actual minimax regret rule to challenge the algorithm. Namely we take the symmetric δ^1 defined by $\delta^1_2(n_1, n_2) = 1$ if $n_1/N_1 < 1/2$ or if ($n_1/N_1 = 1/2$ and $n_2/N_2 < 1/2$). (Consequently, by symmetry, $\delta^1_2(n_1, n_2) = 0$ if $n_1/N_1 > 1/2$ or ($n_1/N_1 = 1/2$ and $n_2/N_2 > 1/2$)). For easy reference, we refer to this initialization by SO (for "suboptimal").

In TABLE I, we report the minimax regret value using the formula in Stoye (2009, Corollary 1) for the original model (before discretization) for the subset $N_1 = N_2 \in \{5, 60, 100, 200\}$ of the sample sizes we considered. In addition, for each initialization (ES and SO) and each of the two weighting schemes α_n we report the maximal regret of δ^n for $n \in \{1, 150, 500, 2000\}$ (over all actions by nature with $p = 1000$). All digits of maximal regret of δ^n that coincide with the ones of the actual minimax regret value are displayed in bold, starting at the tenths place, then the hundredths place, etc. until there is a discrepancy.[15]

The **key findings** from the results in TABLE I are as follows. The results reflect the convergence results from Proposition 1. As the number of iterations of the algorithm increases the maximal regret of δ^n approaches the minimax regret value. As suggested by the theory, this holds for both choices of α_n and initializations δ^1 in all cases considered. We

[13]We also do some robustness checks by comparing the results to the case where we only take $p = 500$ and the results vary only very slightly. In Stoye (2009, Corollary 1) we use stepsize .000001 in a gridsearch over different values of a (that appears in Stoye (2009, Corollary 1) to evaluate the formula for minimax regret given there.

[14]Because $\delta^{n+1} = (1 - \alpha_{n+1})\delta^n + \alpha_{n+1}\delta^n_{BR}$.one obtains $\delta^1 = \delta^1_{BR}$ with $\alpha_n = 1/n$ and the impact of the initialization is only transferred through its impact on μ^n_{BR}.

[15]The additional results for other (N_1, N_2) values appear in the Supplementary Appendix. As discussed in Comment 1 after Algorithm 2 minimax regret of the discretized model is smaller than the minimax regret value (but converges to the latter as $p \to \infty$). That explains why in very few cases (like e.g. when $N_1 = N_2 = 5$) the reported maximal regret is (very slightly, i.e. in the 10^{-7} digit) smaller than the minimax regret value of the original model.

find that $\alpha_n = (5+n)^{-.7}$ leads to faster convergence than $\alpha_n = 1/n$. Namely, maximal regret of δ^{500} and δ^{2000} is identical to minimax regret up to at least four digits after the comma (and sometimes up to seven digits after the comma) for the former weighting scheme. Quite remarkably, maximal regret of δ^{150} already typically approximates minimax regret up to four digits (in very few cases only up to three or even up to five digits). Instead, for the latter weighting, maximal regret of δ^{500} (and δ^{2000}) coincides with minimax regret only up to two or three (two to four) digits. Based on these results, we strongly suggest using $\alpha_n = (5+n)^{-.7}$ over the original weighting $1/n$ proposed by Robinson (1951). These results also reinforce an earlier remark about the importance of using a good weighting scheme and potentially obtaining theoretical insights on the convergence speed as a function of the weighting scheme.

Regarding the initialization, the values reported for maximal regret of δ^1 illustrate that SO is indeed a very suboptimal treatment rule. E.g. for $(N_1, N_2) = 60$ maximal regret under SO equals .3851399 while under ES it equals .0170815. Somewhat surprisingly however, Algorithm 2 recovers quite quickly from the suboptimal initialization and displays quite similar convergence patterns when started with SO versus ES. When considering maximal regret of δ^{150} often the algorithm started with SO leads to at least equal if not smaller values! A reason for why the initialization may not matter that much is that no matter what the initialization is, given the iterative structure of the algorithm, the least favorable distribution of nature needs to build up from scratch. One might want to consider modifications of the algorithm that also start off with an informed guess on a least favorable distribution. Overall, the finding that the initialization is not crucial for fast convergence is good news.

Substantial improvements in terms of maximal regret are obtained by the algorithm after only quite few iterations. E.g. when $(N_1, N_2) = 5$ and $\alpha_n = (5+n)^{-.7}$, the maximal regret of ES equals .0703386 while minimax regret equals .054308. After 150 iterations the algorithm generates a treatment rule whose maximal regret equals .054309.

As mentioned already, $R(\delta^n, \mu^n_{BR})$ is not monotonically decreasing in n. Likewise, R^n is not monotonically increasing in n. In practice therefore, after a certain number of iterations, N say, one should report the interval

$$I_N := [\max_{n \leq N} R^n, \min_{n \leq N} R(\delta^n, \mu^n_{BR})] \tag{3.15}$$

that contains the minimax regret value (of the discretized game) and use the treatment rule δ^n for the n that minimizes $R(\delta^n, \mu^n_{BR})$ over all $n \leq N$. For example, for ES and $\alpha_n = (5+n)^{-.7}$, I_{2000} equals [.0542240,.0543086], [.0155017,.0155298], [.0119668,.0120253], [.0083640,.0085001] for $(N_1, N_2) = 5$, 60, 100, 200, respectively. Note again here that the upper endpoint of I_N maybe smaller than the minimax regret value of the original (not discretized) model. E.g. that happens for $(N_1, N_2) = 200$ but only in the 10^{-7} digit.

In all examples, not only this one, as one would expect, maximal regret decreases in the sample size.

Regarding **computation time** (using $p = 1000$, ES, $\alpha_n = (5+n)^{-.7}$, and 2000 iterations in all cases) for $(N_1, N_2) = (10, 10)$ it takes less than 5 minutes while for $(N_1, N_2) = (200, 200)$ it takes about 24 minutes. In the latter case, the precision for maximal regret of δ^{2000} is about $3 \cdot 10^{-6}$. This is quite remarkable; namely, note that there are $2^{N_1(N_2+1)/2+N_2/2}$ symmetric treatment rules that take values $\delta_2(n_1, n_2) \in \{0, 1\}$ except when $n_1/N_1 = n_2/N_2 = 1/2$ (in which case $\delta_2(n_1, n_2) = 1/2$). With $(N_1, N_2) = (200, 200)$ that number equals $2^{20200} \approx 10^{6080.8}$!

Note that even larger sample sizes are possible. E.g. we also considered $(N_1, N_2) = (250, 250)$, where maximal regret for δ^n equals .008007, .007665, .007615, and .007603 for $n = 1, 150, 500$, and 2000, respectively, and $R(\delta^n, \mu^n_{BR}) - R^n$ equals .007668, .001310, .000472, and .000223 for these choices of n and computation time being about 39 minutes. In that case $2^{N_1(N_2+1)/2+N_2/2} = 2^{31500} \approx 10^{9482.4}$! The minimax regret value of the original model equals .007602. When $N_1 = N_2 = 300$ there are about $10^{13636.7}$ symmetric "nonrandomized" rules. The theoretical minimax regret value equals .006940. Maximal regret of δ^n equals .007279, .007053, .006950, and .006939 for $n = 1, 150, 500$, and 2000, respectively, and $R(\delta^n, \mu^n_{BR}) - R^n$ equals .006998, .001222, .000352, and .000271 for these choices of n. Computation time is 51 minutes. With $N_1 = N_2 = 400$ computation time increases to 1 hour 27 minutes.

To give some idea about the complexity of least favorable distributions, when $(N_1, N_2) = 20$, ES, and $\alpha_n = (5+n)^{-.7}$ the best response by nature against δ^{2000} is a mixed strategy with $\overline{m} = 142$ in Proposition 5.

TABLE I: Maximal regret of δ^n for various choices of n, (N_1, N_2), weighting schemes and initializations.

Weighting choice α_n	n^{-1}	n^{-1}	$(5+n)^{-.7}$	$(5+n)^{-.7}$
Initialization δ^1	SO	ES	SO	ES
$(N_1, N_2) = 200$; minimax regret value=.00850086				
Maximal regret of δ^1	.4233904	**.00**90009	.4233904	**.00**90009
Maximal regret of δ^{150}	**.0**108846	**.0**110614	**.0085**676	**.008**6036
Maximal regret of δ^{500}	**.00**90704	**.00**91066	**.0085**934	**.0085**027
Maximal regret of δ^{2000}	**.008**6211	**.008**6256	**.0085008**	**.0085**038
$(N_1, N_2) = 100$; minimax regret value=.01202529				

Maximal regret of δ^1	.4022899	**.012**9888	.4022899	**.012**9888
Maximal regret of δ^{150}	**.01**31981	**.01**44591	**.0121**011	**.0120**440
Maximal regret of δ^{500}	**.012**3099	**.012**5261	**.012025**8	**.0120**691
Maximal regret of δ^{2000}	**.0120**876	**.012**1344	**.012025**8	**.01202**64
$(N_1, N_2) = 60$; minimax regret value=.01553018				
Maximal regret of δ^1	.3851399	**.01**70815	.3851399	**.01**70815
Maximal regret of δ^{150}	**.01**62264	**.01**72229	**.0155**947	**.01553**77
Maximal regret of δ^{500}	**.015**6859	**.015**9248	**.01553**18	**.0155**883
Maximal regret of δ^{2000}	**.0155**676	**.015**6287	**.0155301**	**.015530**6
$(N_1, N_2) = 5$; minimax regret value=.05430889				
Maximal regret of δ^1	.2730320	**.0**703386	.2730320	**.0**703386
Maximal regret of δ^{150}	**.054**5794	**.054**7494	.**0543**194	**.0543**309
Maximal regret of δ^{500}	**.054**4255	**.054**4402	.**0543**246	.**0543**287
Maximal regret of δ^{2000}	**.0543**327	**.0543**418	.**0543**110	.**0543087**

Next, we apply Algorithm 2 to **unbalanced samples** for which no analytical formula is known for the minimax regret rule. Namely, we consider all combinations of $N_1 \in \{10, 50, 100\}$ and $N_2 - N_1 \in \{10, 50, 100\}$ using the initialization as in Algorithm 2. Based on the insights from the simulations in the balanced case we use $\alpha_n = (5+n)^{-.7}$. All reported results use $p = 1000$ again. Given we do not have an analytical formula for the minimax regret value, TABLE II reports results on the maximal regret of δ^n together with the bound $R(\delta^n, \mu^n_{BR}) - R^n$ (for how far the maximal regret of δ^n differs at most from the minimax regret value of the discretized model) for $n \in \{1, 150, 500, 2000\}$ and six of the nine choices of (N_1, N_2). The remaining results appear in the Supplementary Appendix. In addition, we report I_{2000} introduced in (3.15).

TABLE II: Maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for several n and I_{2000} for several (N_1, N_2).

$n \backslash (N_1, N_2)$	$(10, 20)$	$(10, 60)$	$(10, 110)$
1	.037601;.034347	.035049;.035049	.035049;.035049
150	.032765;.000074	.028602;.001713	.027716;.000452
500	.032730;.000033	.028580;.000003	.027593;.000624
2000	.032728;.000003	.028580;.000003	.027583;.000081
I_{2000}	[.032726,.032727]	[.028578,.028579]	[.027501,.027579]

$n\backslash(N_1,N_2)$	$(100,110)$	$(100,150)$	$(100,200)$
1	.011793;.003706	.011269;.010099	.010921;.010379
150	.011812;.000918	.011112;.001011	.010527;.001344
500	.011766;.000104	.010973;.000439	.010404;.000422
2000	.011739;.000058	.010963;.000097	.010403;.000144
I_{2000}	[.011705,.011736]	[.010874,.010962]	[.010311,.010402]

The **key findings** from TABLE II are as follows. As expected from the theory, maximal regret of δ^n converges (to the minimax regret value of the discretized model) and $R(\delta^n, \mu^n_{BR}) - R^n$ goes to zero as the number of iterations n increases. In all cases considered $R(\delta^n, \mu^n_{BR}) - R^n$ for $n = 2000$ ranges between .000003 and .000208. Furthermore, in all cases considered the width of I_{2000} falls into the interval $[10^{-6}, 9 \cdot 10^{-5}]$.

We also considered Robinson's (1951) weighting, $\alpha_n = n^{-1}$. Consistent with the findings above from balanced samples, in all cases $\alpha_n = (5+n)^{-.7}$ produced smaller maximal regret for δ^{2000}. For brevity, we do not report those results.

To give some idea about the structure of the minimax regret rule, in the Supplementary Appendix we report δ^{5000} for $N_1 = 10$, $N_2 = 20$, ES, and $\alpha_n = (5+n)^{-.7}$.

What are the gains of incorporating symmetry as in Algorithm 2 versus using Algorithm 1 in terms of computation time and maximal regret after a given number of iterations? Consider e.g. the case $p = 1000$, ES, and $\alpha_n = (5+n)^{-.7}$. When $(N_1, N_2) = (200, 200)$ it takes 42 minutes to run 2000 simulations (versus 24 minutes when symmetry is exploited) and maximal regret of δ^{2000} equals .008505 (versus .008503 when symmetry is exploited). For $(N_1, N_2) = (10, 60)$ it takes 11 minutes to run 2000 simulations with maximal regret of δ^{2000} equal to .028595 while the algorithm that exploits symmetry only takes 5.5 minutes and maximal regret of δ^{2000} equals .028580. The advantage of Algorithm 2 for these sample sizes is therefore mostly in terms of computation speed.

Next, we provide some comparisons with the "multiplicative weights" algorithm of Aradillas Fernández, Blanchet, Montiel Olea, Qiu, Stoye, and Tan (2024, ABMQS from now on).[16] Given the gigantic number of nonrandomized treatment rules even for moderate numbers of N_1 and N_2 it is not possible to implement that algorithm when the policymaker is the "outer" player. However, using the identity $\inf_{\delta \in \mathbb{D}} \sup_{s \in S} R(\delta, s) = \sup_{s \in S} \inf_{\delta \in \mathbb{D}} R(\delta, s)$ discussed in the Appendix B of ABMQS, one could let the policymaker play the role of the "inner" player and use comparisons of conditional means and Bayes' rule to solve $\inf_{\delta \in \mathbb{D}} R(\delta, s)$ for given s as we suggest in (2.13)-(2.15). Using that approach, the subgradient and nature's strategy are p^T-dimensional vectors. Calculating the former, requires p^T many regret cal-

[16] We would like to thank Pepe Montiel for sharing very helpful insights related to this discussion.

culations (as does grid search used in fictitious play). Storing the latter, while likely still possible when $T = 2$, is likely impossible when $T > 2$. Furthermore, by construction, "multiplicative weights" assigns positive probability to *each* of the p^T actions by nature. Thus, calculation of the sums appearing in (2.15) requires summing $\overline{m} = p^T$ terms for each of the $(N_1+1)(N_2+1)$ arguments of the treatment rule. Instead, using fictitious play we find that typically $\overline{m}$ is of the order of 100 even after 2000 iterations which leads to considerably less computational effort.[17] Thirdly, note that the maximal number of iterations $\lceil 2\ln(p^T)/\epsilon^2 \rceil$ given in Theorem 1 of ABMQS that guarantees an ϵ minimax regret rule equals 2.76×10^9 when $T = 2, p = 1000$, and $\epsilon = 10^{-4}$. Instead, note that in all cases reported in TABLE II after just 2000 iterations, a minimax regret rule is found whose maximal regret is at most 9×10^{-5} larger than the minimax regret value of the discretized model. Conversely, if for $T = 2$ and $p = 1000$ one sets $\lceil 2\ln(p^T)/\epsilon^2 \rceil$ equal to 2000 one obtains $\epsilon = 0.11$.

In general, we believe that a comparison of two algorithms and the notion of optimality of an algorithm, should take into account the number of iterations *and* the number of operations performed in each iteration. The upper bound $\lceil 2\ln(p^T)/\epsilon^2 \rceil$ given in Theorem 1 of ABMQS is a powerful result that applies to a broad class of models. But it does not seem to be binding in the specific application considered here.

3.3.1 Restricted strategy space for nature

We now consider the case where the set M of mean vectors no longer equals the unrestricted set $[0,1]^2$. Instead, we examine a scenario where the policymaker uses a priori information about certain parametric restriction on (μ_1, μ_2). Being able to accommodate such restrictions is highly empirically relevant. For example, average wage after job training should be expected to be nonsmaller. If one also factors in a fixed cost for job training into the analysis one ends up with a set of particular restrictions on the mean vectors (μ_1, μ_2). For the purpose of illustration, we take

$$M = \{(\mu_1, \mu_2) \in [0,1]^2 : 0.9\mu_1 \leq \mu_2 \leq 1.2\mu_1\}. \tag{3.16}$$

For restricted M, Proposition 5 that relied heavily on symmetry, no longer applies. Therefore, we do not restrict the search to symmetric rules and use Algorithm 1 to approximate the minimax regret rule.

We consider sample sizes $N_1 = N_2 \in \{5, 10, 20, 50\}$, $(N_1, N_2) \in \{(5,10), (10,5), (10,50), (50,10)\}$, $\alpha_n = (5+n)^{-.7}$, $p = 1000$, and use 2,000 iterations. We initialize with δ^1 equal to the ES

[17]Potentially, using some thresholding approach could alleviate the computational burden for "multiplicative weights."

rule defined by $\delta_2^1(n_1, n_2) = 1$ if $n_1/N_1 \leq n_2/N_2$ and $\delta_2^1(n_1, n_2) = 0$ otherwise and take D to be the set of all nonrandomized rules. We take

$$M_\varepsilon = \{(\mu_1, \mu_2) \in M; \mu_j = p_j/p \text{ for some } p_j \in \{0, 1, ..., p\} \text{ for } j = 1, 2\}. \tag{3.17}$$

TABLE III gives results on the maximal regret of δ^n and the bound $R(\delta^n, \mu_{BR}^n) - R^n$ for $n \in \{1, 150, 500, 2000\}$.

TABLE III: Maximal regret of δ^n and $R(\delta^n, \mu_{BR}^n) - R^n$ for $n \in \{1, 150, 500, 2000\}$.

$n \backslash (N_1, N_2)$	$(5,5)$	$(10,10)$	$(20,20)$	$(50,50)$
1	.059049;.059049	.036090;.036090	.027256;.027256	.018518;.018518
150	.036396;.000623	.030029;.000665	.023947;.000492	.016758;.000809
500	.036159;.000275	.029816;.000541	.023775;.000182	.016601;.000743
2000	.035986;.000173	.029789;.000100	.023788;.000119	.016595;.000131
$n \backslash (N_1, N_2)$	$(5,10)$	$(10,5)$	$(10,50)$	$(50,10)$
1	.035049;.035049	.059049;.059049	.027084;.027084	.035049;.035049
150	.033472;.000542	.032834;.001307	.025716;.000195	.024615;.000683
500	.033415;.000170	.032579;.000217	.025690;.000342	.024567;.000422
2000	.033388;.000153	.032552;.000147	.025574;.000116	.024529;.000195

As the **key findings** TABLE III illustrates again the convergence properties of maximal regret of δ^n and convergence to zero of $R(\delta^n, \mu_{BR}^n) - R^n$ as n increases. After 2000 iterations the algorithm has reached treatment rules δ^{2000} whose maximal regret is at most $1.95 \cdot (10^{-4})$ away from the minimax regret value. Compared to TABLE I for equal sample sizes we find that here maximal regret of δ^n is always smaller given that nature is less powerful here given the restrictions on its parameter space.

Furthermore, the results from TABLE III illustrate that when there is a priori information the ES rule may be a very suboptimal choice. For example when $(N_1, N_2) = (5, 5)$ or $(N_1, N_2) = (10, 5)$ its maximal regret equals .059049, more than 1.6 times or 1.8 times as much as the maximal regret of δ^{2000} which equals .035986 or .032552. With M being restricted it is very hard to guess a reasonable treatment rule. This provides strong motivation for the algorithm provided in this paper whose convergence properties do not seem to be affected much by the initialization.

We describe δ_2^{2000} in the case where $N_1 = N_2 = 5$. We find that $\delta_2^{2000}(n_1, n_2) = 1$ if $n_1 < n_2$ and $\delta_2^{2000}(n_1, n_2) = 0$ if $n_1 > n_2$. Finally, $\delta_2^{2000}(n_1, n_1)$ equals 1, .948767, .906576, .844678, .704494, and .533296 when $n_1 = 0, 1, ..., 5$, respectively. Next we (partially) describe

δ_2^{2000} in the case where $N_1 = 5$ and $N_2 = 10$. In that case, $n_1/N_1 < n_2/N_2$ still implies $\delta_2^{2000}(n_1, n_2) = 1$ but it is no longer true that for all 30 cases where $n_1/N_1 > n_2/N_2$, always $\delta_2^{2000}(n_1, n_2) = 0$ holds. In five of those cases $\delta_2^{2000}(n_1, n_2) > 0$.[18]

Even though we cannot exploit any symmetry conditions for this scenario, the algorithm is still running very quickly.[19] E.g. it takes less than 4 and 4.5 minutes to obtain 2000 iterations when $N_1 = N_2 = 5$ and 50, respectively; virtually the same computation time in both cases.

4 Treatment assignment when testing innovations

Suppose the policymaker knows the mean outcome μ_T of a status quo treatment. She needs to consider whether to switch from the status quo to one of $T - 1$ alternative treatments whose mean outcomes are unknown. Stoye (2009) considers the case $T = 2$ and works out a minimax regret rule. In this section, we apply Algorithm 1 to $T > 2$ where a sample of equal sizes on $T - 1$ alternative treatments is observed. Potential outcomes are elements of $\{0, 1\}$; the general case where outcomes are in $[0, 1]$ is then dealt with via the coarsening approach of Proposition 2. A priori information on restrictions of the mean vector $\{\mu_1, ..., \mu_{T-1}\}$ can be incorporated. Without further restrictions $M = [0, 1]^{T-1} \times \mu_T$. As in the previous example, to reduce the computational effort, we derive certain symmetry conditions first.

As before, we can identify the DGP by the vector of means. The observed sample w_N can be summarized by the number of observed successes n_t for treatments $t = 1, ..., T - 1$. For a rule δ and sample w_N we denote by $\delta_t(w_N)$ the probability that treatment t (for $t = 1, ..., T - 1$) will be chosen and consequently by $1 - \sum_{t=1}^{T-1} \delta_t(w_N)$ the probability that the status quo persists. We suppress the dependence on μ_T in the notation when deriving a minimax regret rule for a particular μ_T.

4.1 Symmetry restrictions

There are T treatments to choose from and treatment T is the known status quo. Data of equal sample size $\overline{N}$ is observed for each of the $T - 1$ innovations. We take $S = \{0, 1\}^{T-1}$. Denote by σ^{-1} the inverse of a permutation σ in the group of permutations. We make the following definition.

[18] We ran the algorithm for 4000 iterations and now only two such cases (up to 10^{-6} precision) occured. Namely $\delta_2^{4000}(1, 0) = 1$ and $\delta_2^{4000}(1, 1) = .000280$. So, when 1 success for treatment 1 and zero successes for treatment 2 are observed, treatment 2 should be chosen with probability 1 - that seems quite counterintuitive.

[19] What significantly reduces the running time is that the grid search over mean vectors when updating nature's strategy requires less time given the imposed restrictions on M.

Definition (symmetric treatment rule). *We call a treatment rule δ symmetric*[20] *if whenever the vectors $w_N = (n_1, ..., n_{T-1})$ and $w'_N = (n'_1, ..., n'_{T-1})$ are permutations of each other, i.e. $n_{\sigma(t)} = n'_t$ for some permutation σ on $\{1, ..., T-1\}$ then*

$$\delta_t(w_N) = \delta_{\sigma^{-1}(t)}(w'_N) \text{ for } t \in \{1, ..., T-1\}. \tag{4.1}$$

Comments. 1. As in Section 3, the symmetry property allows for computational gains in approximating minimax regret rules. Note that (4.1) implies that $\delta_T(w_N) = \delta_T(w'_N)$. In the case $T = 3$, symmetry requires that $\delta_1((n_1, n_2)) = \delta_2((n_2, n_1))$ for all $n_1, n_2 \in \{0, 1, ..., \overline{N}\}$. Thus, for a symmetric treatment rule δ, given values $\delta_2(w_N)$ and $\delta_3(w_N)$ for $w_N = (n_1, n_2)$ one can determine $\delta_1(w_N)$ and $\delta_t((n_2, n_1))$ for $t \in \{1, 2, 3\}$. Therefore, knowing $\delta_t((n_1, n_2))$ for $t \in \{2, 3\}$ and all $n_2 \geq n_1$ one automatically knows the entire treatment rule.

The next proposition establishes that one can find an action pair (δ, s) with $\delta \in \mathbb{D}$ and $s \in \Delta\mathbb{S}$ for the policymaker and nature, respectively, that constitutes a Nash equilibrium and where δ is symmetric according to the definition just given.

Proposition 6 (i) *Suppose nature plays a mixed strategy that picks m mixed strategies μ_m^{MS} with probabilities $p_m > 0$ for $m = 1, ..., \overline{m}$ for some $\overline{m} \geq 1$, where $p_1 + ... + p_{\overline{m}} = 1$ and for all $m = 1, ..., \overline{m}$, μ_m^{MS} mixes uniformly over all of the $(T-1)!$ permutations $(\mu_{\sigma(1)m}, ..., \mu_{\sigma(T-1)m}, \mu_T)$ of a vector $(\mu_{1m}, ..., \mu_{T-1m}, \mu_T)$, where σ denotes a permutation on $\{1, ..., T-1\}$. Then if there exists a treatment rule δ that is a best response to that strategy (defined by the μ_m^{MS} and p_m for $m = 1, ..., \overline{m}$) then it can be chosen to satisfy (4.1).*

(ii) *Assume the policymaker chooses a symmetric treatment rule δ and $(\mu_1, ..., \mu_T)$ is a best response by nature. Then for any permutation σ on $\{1, ..., T-1\}$ the mean vector $(\mu_{\sigma(1)}, ..., \mu_{\sigma(T-1)}, \mu_T)$ is also a best response.*

Comments. 1. Proposition 6 establishes that a Nash equilibrium of the unrestricted game exists in which the policymaker plays a symmetric treatment rule. Therefore, if there exists a minimax regret rule that is part of a Nash equilibrium then

$$\inf_{\delta \in \mathbb{D}} \max_{\mu \in [0,1]^{T-1} \times \mu_T} R(\delta, \mu) = \inf_{\substack{\delta \in \mathbb{D} \\ \delta \text{ is symmetric}}} \max_{\substack{s \in \Delta\mathbb{S} \\ s \text{ is as in Proposition 6(i)}}} R(\delta, \mu). \tag{4.2}$$

Using the formulation of the problem as on the right side of the equation in (4.2), one can pick as D the finite set of symmetric nonrandomized rules. The proof of part (ii) follows from the fact that if δ is symmetric then $R(\delta, (\mu_1, ..., \mu_T)) = R(\delta, (\mu_{\sigma(1)}, ..., \mu_{\sigma(T-1)}, \mu_T))$ for any permutation σ on $\{1, ..., T-1\}$.

[20] It should not be confusing that the notion of symmetry changes in different applications.

4.2 Algorithm

We establish that the assumption of Lemma 1 about the regret function $R(\delta, \cdot)$ being continuous in $\mu \in [0,1]^{T-1} \times \mu_T$ uniformly in $\delta \in \mathbb{D}$ is satisfied.

Lemma 3 *$\forall \lambda > 0$ there exists $\eta > 0$ such that when $\|\mu - \mu'\| < \eta$ for $\mu, \mu' \in [0,1]^{T-1} \times \mu_T$ then for all $\delta \in \mathbb{D}$ we have $|R(\delta, \mu) - R(\delta, \mu')| < \lambda$.*

We will employ the analogous discretization strategy given in (3.8). For given $\mu_T \in [0,1]$, some $p \in \mathbb{N}$ and $\varepsilon = 1/p$ define

$$M_\varepsilon = \{(\mu_1, ..., \mu_{T-1}, \mu_T); \mu_t = p_t/p \text{ for some } p_t \in \{0,1,...,p\} \text{ and } t = 1,...,T-1\}. \quad (4.3)$$

We modify the general algorithm to reflect the symmetry property under the setup of testing $T-1$ innovations. Pick again a tolerance level $\xi > 0$.

Algorithm 3 ***Initialization:***

i) Define δ^1 as the empirical success rule. More precisely, let $\overline{n} = \max\{n_1/\overline{N}, ..., n_{T-1}/\overline{N}, \mu_T\}$ and define W by letting $t \in W \subset \{1,...,T\}$ if $n_t/\overline{N} = \overline{n}$ for $t \in \{1,...,T-1\}$ and $\mu_T = \overline{n}$ for $t = T$. Then

$$\delta_t^1(w_N) = 1/|W| \text{ if } t \in W \text{ and } \delta_t^1(w_N) = 0 \text{ otherwise.} \quad (4.4)$$

ii) Initialize $\nu^0 = 0$ and $R^0 = 0$.

Iteration:

For $n = 1, 2, 3...$ ***DO***:

i) Find a best response μ_{BR}^n by nature in response to δ^n, where μ_{BR}^n equals a mixed strategy that randomizes uniformly over all $(T-1)!$ permutations $(\mu_{\sigma(1)}^n, ..., \mu_{\sigma(T-1)}^n, \mu_T)$ of the first $T-1$ components of some vector $(\mu_1^n, ..., \mu_{T-1}^n, \mu_T)$ in M_ε.

ii) If

$$R(\delta^n, \mu_{BR}^n) - R^{n-1} < \xi \quad (4.5)$$

*then **break.** Use the rule δ^n.*

iii) Update nature's mixed strategy to be the following weighted average

$$\nu^n = (1-\alpha_n)\nu^{n-1} + (\alpha_n/(T-1)!) \textstyle\sum_\sigma I((\mu_{\sigma(1)}^n, ..., \mu_{\sigma(T-1)}^n, \mu_T^n)), \quad (4.6)$$

where $I(x)$ denotes a point mass of size 1 at the point $x \in [0,1]^T$ and the sum is over all permutations of $(1,...,T-1)$.

iv) Compute a best response δ^n_{BR} in nonrandomized symmetric strategies by the player to ν^n, that is,

$$R^n := R(\delta^n_{BR}, \nu^n) = \min_{\substack{\delta \in \mathbb{D}, \delta \text{ is nonrandomized} \\ \text{and symmetric}}} R(\delta, \nu^n). \tag{4.7}$$

v) Update the treatment rule by

$$\delta^{n+1} = (1 - \alpha_{n+1})\delta^n + \alpha_{n+1}\delta^n_{BR}. \tag{4.8}$$

Comments. 1. In each iteration, the best response of the policymaker can be calculated via Bayes rule using (3.14). The best response by nature is solved again by grid search over all $(p+1)^{T-1}$ possible choices of mean vectors.

2. If M is a strict subset of $[0,1]^{T-1} \times \mu_T$ the symmetry property can in general not be exploited and Algorithm 1 is used instead.

3. Note that the existence of best responses of the particular type in iteration steps i) and iv) is guaranteed by Proposition 6.

We implement Algorithm 3 for the case $T = 3$, sample sizes $\overline{N} = N_1 = N_2 \in \{5, 10, 20, 30, 40, 50, 100, 200\}$ and known mean of the status quo treatment equal to $\mu_3 \in \{.2, .5, .8\}$. The initialization rule is the empirical success rule described in the beginning of Algorithm 3. We choose $p = 1000$ and $\alpha_n = (5 + n)^{-.7}$. TABLE IV reports maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for various choices of n and I_{2000} for a subset of the sample sizes.[21]

TABLE IV: Maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for several n and I_{2000} for several $\overline{N}$ and μ_3.

μ_3	0.2	0.5	0.8
$\overline{N} = 200$			
$n = 1$	.008615;.001698	.010715;.010715	.0090221;.009022
$n = 150$	.008500;.000046	.011725;.002386	.008726;.001326
$n = 500$	.008500;.000042	.010495;.000564	.008397;.000888
$n = 2000$	.008500;.000043	.010492;.000159	.008374;.000171
I_{2000}	[.008499;.008500]	[.010417;.010488]	[.008241;.008372]

[21] To save space we report the remaining results in the Supplementary Appendix.

$\overline{N} = 100$			
$n = 1$	.012333;.002467	.015157;.015157	.013036;.013036
$n = 150$	.012025;.000046	.014872;.000681	.011889;.000784
$n = 500$	.012025;.000000	.014860;.000252	.011841;.000180
$n = 2000$	.012025;.000002	.014831;.000081	.011839;.000166
I_{2000}	[.012025;.012025]	[.014771;.014830]	[.011822.011835]
$\overline{N} = 50$			
$n = 1$	.017768;.003592	.021440;.021440	.019001;.019001
$n = 150$	.017057;.000111	.021054;.000321	.016804;.000155
$n = 500$	.017015;.000040	.020978;.000045	.016775;.000448
$n = 2000$	.017015;.000016	.020967;.000073	.016736;.000166
I_{2000}	[.017010;.017015]	[.020963;.020966]	[.016725;.016729]
$\overline{N} = 10$			
$n = 1$	.043262;.009139	.047978;.047978	.047926;.047926
$n = 150$	.038815;.000303	.047061;.000290	.037808;.000808
$n = 500$	.038832;.000110	.047365;.000498	.037655;.000151
$n = 2000$	.038819;.000046	.046945;.000132	.037631;.000109
I_{2000}	[.038810;.038810]	[.046936;.046936]	[.037537;.037549]

As the **key findings** TABLE IV illustrates again the convergence properties of maximal regret of δ^n and convergence to zero of $R(\delta^n, \mu_{BR}^n) - R^n$ as n increases. After 2000 iterations the algorithm has reached treatment rules δ^{2000} whose maximal regret is at most $1.7 \cdot (10^{-4})$ away from the minimax regret value. TABLE IV also illustrates that it is beneficial to keep track of the maximal regret of the treatment rules along all iterations n as the right boundary of I_{2000} maybe strictly smaller than maximal regret of δ^{2000} indicating that a treatment rule with smaller maximal regret occurred in an earlier iteration. What is the impact of μ_3 on the results? The minimax regret value when $\mu_3 = 0.5$ is the largest (amongst those considered) followed by the minimax regret value when $\mu_2 = 0.2$ and $\mu_1 = 0.8$, although the difference is very small among the latter two.

5 Conclusion

We propose an algorithm to numerically approximate minimax regret rules and prove that the maximal regret of the treatment rules generated by the algorithm converges to the minimax regret of the discretized model. The latter converges to the minimax regret value of the original model as the discretization becomes finer. We illustrate in several examples that the algorithm works very well in practice. Importantly, our framework allows incorporation of a priori information that the policymaker can use regarding the set of DGPs that nature can choose from. Also, the framework allows for a restricted set of policy rules that the policymaker can choose from, reflecting the important case where policy restrictions prevent the policymaker from using certain rules. However, in the latter case one would in general not be able to use the approach based on Bayes' rule.

There are several important open questions that go beyond the scope of the current paper and require future research. In particular, we found that the choice of weights α_n in the updating step of the algorithm significantly impacted the convergence speed. Theoretical results should be developed to work out optimal choices for the weights. Second, we found that certain symmetry conditions can be exploited to significantly reduce the complexity of the algorithm. Techniques need to be developed to detect all such symmetries. Third, implementation details of the algorithm under policy restrictions are quite important in cases where the approach via Bayes' rule is no longer applicable. Fourth, progress has to be made in allowing for covariates in the model.

There are many additional examples and related applications that the algorithm or variants of it could be applied to. In particular, the proposed algorithm may be helpful in approximating optimal tests in terms of weighted average power when the null hypothesis is composite.

6 Appendix

To simplify the presentation, we typically do not index expectations and probabilities by the DGP s unless it is needed for clarity.

Proof of Lemma 1. First consider part (i). Clearly, $V \geq V_m$. Pick a $\lambda > 0$. We need to show that there exist an $m_\lambda \in \mathbb{N}$ such that for all $m \geq m_\lambda$ we have $V - V_m \leq \lambda$. For every $\delta \in \mathbb{D}$ let $\mu_\delta \in M$ be such that $\max_{\mu \in M} R(\delta, \mu) = R(\delta, \mu_\delta)$. By the assumed continuity of $R(\delta, \cdot)$ there exists $\eta_\lambda > 0$ such that $|R(\delta, \mu) - R(\delta, \mu')| < \lambda$ whenever $\|\mu - \mu'\| < \eta_\lambda$. Pick $m_\lambda \in \mathbb{N}$ such that for all $m \geq m_\lambda$, $\varepsilon_m < \eta_\lambda$. Because M_ε is an ε-discretization of M we can find $\mu_{\delta,\varepsilon_m} \in M_{\varepsilon_m}$ such that $\|\mu_\delta - \mu_{\delta,\varepsilon_m}\| < \varepsilon_m < \eta_\lambda$ and therefore

$$\max_{\mu \in M_{\varepsilon_m}} R(\delta, \mu) \geq R(\delta, \mu_{\delta,\varepsilon_m}) \geq R(\delta, \mu_\delta) - \lambda = \max_{\mu \in M} R(\delta, \mu) - \lambda. \tag{6.1}$$

That implies that $V_m \geq V - \lambda$ as desired. For part (ii) let δ_m be any rule that satisfies

$\max_{\mu \in M} R(\delta_m, \mu) \to V$ as $m \to \infty$. Note that for all $m \geq \overline{m}_\lambda$ for some $\overline{m}_\lambda \in \mathbb{N}$

$$\max_{\mu \in M} R(\delta_{\varepsilon_m}, \mu) \leq \max_{\mu \in M_{\varepsilon_m}} R(\delta_{\varepsilon_m}, \mu) + \lambda \leq \max_{\mu \in M_{\varepsilon_m}} R(\delta_m, \mu) + 2\lambda, \tag{6.2}$$

where the first inequality holds by (6.1) for all $m \geq m_\lambda$ with δ_{ε_m} playing the role of δ. The second inequality in (6.2) holds because by definition δ_{ε_m} is approximately minimax regret for the case where the parameter space for nature equals M_{ε_m}. It follows that $\max_{\mu \in M} R(\delta_{\varepsilon_m}, \mu)$ $\leq \max_{\mu \in M} R(\delta_m, \mu) + 2\lambda$ for all $m \geq \overline{m}_\lambda$ which implies the desired result. $\square$

Proof of Proposition 2. We first consider the case of fixed assignment. **i)** The distribution of $\widetilde{Y}_{t_i,i}$ is Bernoulli and by the law of iterated expectations the success probability equals μ_{t_i}. Therefore, the distribution of the sample $\widetilde{w}_N$ depends on the state of nature s only via $(\mu_1, ..., \mu_T)$. Denote by $\widetilde{\delta}$ the "coarsened version" of δ for an arbitrary $\delta \in \mathbb{D}$. Given

$$R(\widetilde{\delta}, s) = \max_{t \in \mathbb{T}} \{\mu_t\} - \textstyle\sum_{t=1}^{T} \mu_t E\delta_t(\widetilde{w}_N), \tag{6.3}$$

also $R(\widetilde{\delta}, s)$ depends on s only via $\mu = (\mu_1, ..., \mu_T)$. Denote by s' the distribution obtained from $s \in \mathbb{S}$ by replacing each marginal Y_t for $t \in \mathbb{T}$ with a Bernoulli distribution with the same expectation and let $\mathbb{S}'$ be the space of all such distributions s'. Given $\mathbb{S}$ is unrestricted (except for $\mu \in M$), it follows that $\mathbb{S}' \subset \mathbb{S}$. Because s and s' have the same mean vector, $R(\widetilde{\delta}, s) = R(\widetilde{\delta}, s')$. Because under s' it follows that $\widetilde{w}_N = w_N$ with probability 1 we obtain $R(\widetilde{\delta}, s') = R(\delta, s')$ and thus

$$R(\widetilde{\delta}, s) = R(\delta, s'). \tag{6.4}$$

Clearly,

$$\min_{\delta \in \mathbb{D}} \max_{s' \in \mathbb{S}'} R(\delta, s') \leq \min_{\delta \in \mathbb{D}} \max_{s \in \mathbb{S}} R(\delta, s) \leq \min_{\widetilde{\delta}: \delta \in \mathbb{D}} \max_{s \in \mathbb{S}} R(\widetilde{\delta}, s). \tag{6.5}$$

But because of $R(\widetilde{\delta}, s) = R(\delta, s')$ the inequalities in (6.5) actually hold as equalities and $\delta \in \arg\min_{\delta \in \mathbb{D}} \max_{s' \in \mathbb{S}'} R(\delta, s')$ implies $\widetilde{\delta} \in \arg\min_{\delta \in \mathbb{D}} \max_{s \in \mathbb{S}} R(\delta, s)$. Therefore, δ^C, the coarsened version of a minimax regret rule in the binomial case, is minimax regret when outcomes live on the unit interval.

ii) As in (6.4), $R(\delta_\varepsilon^C, s) = R(\delta_\varepsilon, s')$, where s' is again the distribution obtained from $s \in \mathbb{S}$ by replacing each marginal Y_t with a Bernoulli distribution with the same expectation. Therefore,

$$\max_{s \in \mathbb{S}} R(\delta_\varepsilon^C, s) = \max_{s', s \in \mathbb{S}} R(\delta_\varepsilon, s') \leq \min_{\delta \in \mathbb{D}} \max_{s' \in \mathbb{S}'} R(\delta, s') + \varepsilon = \min_{\delta \in \mathbb{D}} \max_{s \in \mathbb{S}} R(\delta, s) + \varepsilon, \tag{6.6}$$

where the inequality holds by assumption on δ_ε and the second equality follows because (6.5) holds with equalities.

Next, we consider the case of random assignment, $P(t_i = t) = p_t$ for all $i = 1, ..., N$ and $t \in \mathbb{T}$. The proof of i) and ii) is identical to the one just given. The distribution of $\widetilde{Y}_{t_i,i}$ is Bernoulli and by the law of iterated expectations the success probability equals $\sum_{t\in\mathbb{T}} p_t \mu_t$ which equals the expectation of $Y_{t_i,i}$ when we consider randomness both in treatment assignment and outcomes. In particular, again, given the particular sampling scheme, the distribution of the sample $\widetilde{w}_N$ depends on the state of nature s only via $(\mu_1, ..., \mu_T)$. The proof for testing innovations is also identical. $\square$

Proof of Proposition 4. Assume nature mixes evenly between the two vectors of means $(0, 1/2)$ and $(1, 1/2)$ and the decision maker employs δ_2 as described in Proposition 4. It is enough to establish that the two actions by nature and the policymaker are best responses to each other. Note that δ as defined is a symmetric treatment rule as defined in (3.3) and therefore Proposition 5(ii) applies.

We now show that the vector of means $(0, 1/2)$ maximizes regret given the treatment rule δ. Without loss of generality, we focus on the case where $\mu_1 < \mu_2$. Note that as below (6.18)

$$R(\delta, (\mu_1, \mu_2)) = (\mu_2 - \mu_1) \sum_{n_2=0}^{N_2} (1 - \delta_2(0, n_2)) \binom{N_2}{n_2} \mu_2^{n_2} (1 - \mu_2)^{N_2 - n_2}. \tag{6.7}$$

In particular, $R(\delta, (\mu_1, \mu_2))$ is a linear and strictly decreasing function of μ_1. Therefore, in order to maximize regret, nature must choose $\mu_1 = 0$. Plugging in $\mu_1 = 0$ and $\delta_2(0, n_2) = n_2/N_2$, $R(\delta, (\mu_1, \mu_2))$ becomes

$$\begin{aligned}
R(\delta, (0, \mu_2)) &= \mu_2 \sum_{n_2=0}^{N_2-1} (1 - n_2/N_2) \binom{N_2}{n_2} \mu_2^{n_2} (1 - \mu_2)^{N_2 - n_2} \\
&= \mu_2 (1 - \mu_2) \sum_{n_2=0}^{N_2-1} \binom{N_2 - 1}{n_2} \mu_2^{n_2} (1 - \mu_2)^{N_2 - n_2 - 1} \\
&= \mu_2 (1 - \mu_2)(\mu_2 + (1 - \mu_2))^{N_2 - 1} \\
&= \mu_2 (1 - \mu_2),
\end{aligned} \tag{6.8}$$

where in the second equation we use $(1-n_2/N_2)\binom{N_2}{n_2} = \binom{N_2-1}{n_2}$ and the third equation uses the expansion of $(\mu_2+(1-\mu_2))^{N_2-1}$ exploiting the binomial coefficients. Taking FOC establishes that $\mu_2(1 - \mu_2)$ is maximized for $\mu_2 = 1/2$.

Next we show that δ is a best response to nature's strategy of mixing uniformly over $(0, 1/2)$ and $(1, 1/2)$. Note that in both DGPs the value for μ_2 is the same. Because there are no observations for treatment 1 (and the marginals Y_1 and Y_2 are independent), the sample conveys no information about μ_1. It follows that the conditional means for outcomes under each treatment after observing any sample $w_N = (0, n_2)$ are the same as the unconditional ones and both equal 1/2. Any treatment rule, in particular the proposed rule, is therefore a best response. $\square$

Proof of Proposition 5. (i) Given the mixed strategy $s \in \Delta\mathbb{S}$ by nature, iff a rule δ° by the policymaker satisfies the following condition

$$E(Y_t|w_N) > E(Y_{t'}|w_N) \text{ then } \delta^\circ_{t'}(w_N) = 0 \tag{6.9}$$

for any $t, t' \in \{1, 2\}$ and all samples w_N with $P(w_N) > 0$, then δ is a best response to s. We show below that for $s \in \Delta\mathbb{S}$ as in (i) the following condition holds:

$$E(Y_1|w_N) - E(Y_2|w_N) = E(Y_2|w'_N) - E(Y_1|w'_N) \tag{6.10}$$

for any $w_N = (n_1, n_2)$ and $w'_N = (n'_1, n'_2)$ that satisfy (3.4). Assume δ° is a best response to s. Then, by (6.9), if $E(Y_t|w_N) > E(Y_{t'}|w_N)$ for a sample w_N with $P(w_N) > 0$ it must be that $\delta^\circ_{t'}(w_N) = 0$. But then, by (6.10) $E(Y_{t'}|w'_N) > E(Y_t|w'_N)$ and $\delta^\circ_{t'}(w'_N) = 1$ and therefore (3.3) holds. Lastly, if for a sample w_N with $P(w_N) > 0$, $E(Y_t|w_N) = E(Y_{t'}|w_N)$ holds then, by (6.10) also $E(Y_t|w'_N) = E(Y_{t'}|w'_N)$ holds and optimality of a treatment rule is not affected by what value the treatment rule assigns to the samples w_N and w'_N, in particular, these values can be picked to satisfy (3.3).

We are therefore left to show (6.10). We first provide some preliminary derivations. Note first that for $t \in \{1, 2\}$ we have

$$\begin{aligned}
&E(Y_t|w_N)\\
&= P(Y_t = 1|w_N)\\
&= \textstyle\sum_{m=1}^{2\overline{m}} P(Y_t = 1 \text{ \& nature picks } \overline{\mu}_m|w_N)\\
&= \textstyle\sum_{m=1}^{2\overline{m}} P(Y_t = 1|\text{nature picks } \overline{\mu}_m \text{ \& } w_N)P(\text{nature picks } \overline{\mu}_m|w_N)\\
&= \textstyle\sum_{m=1}^{2\overline{m}} P(Y_t = 1|\text{nature picks } \overline{\mu}_m)P(w_N|\text{nature picks } \overline{\mu}_m)p_m/P(w_N)\\
&= \textstyle\sum_{m=1}^{2\overline{m}} \mu_{tm}P(w_N|\text{nature picks } \overline{\mu}_m)p_m/P(w_N),
\end{aligned} \tag{6.11}$$

where the first equality holds because Y_t is a Bernoulli random variable, the second equality holds by the law of total probability, the third one uses the definition of conditional probability, the fourth one holds by Bayes' theorem and exploits $P(Y_t = 1|\text{nature picks } \overline{\mu}_m \text{ \& } w_N) = P(Y_t = 1|\text{nature picks } \overline{\mu}_m)$, $P(w_N) > 0$, and $P(\text{nature picks } \overline{\mu}_m) = p_m$ for all m, and the fifth one uses $P(Y_t = 1|\text{nature picks } \overline{\mu}_m) = \mu_{tm}$.

Next, note that for $w_N = (n_1, n_2)$ and $w'_N = (n'_1, n'_2)$ that satisfy (3.4) we have for all $m \leq \overline{m}$

$$
\begin{aligned}
& P(w_N| \text{ nature picks } \overline{\mu}_m) \\
& = \prod_{t=1}^{2} \binom{N_t}{n_t} \mu_{tm}^{n_t} (1-\mu_{tm})^{N_t - n_t} \\
& = \prod_{t=1}^{2} \binom{N_t}{n_t} (1-\mu_{tm+\overline{m}})^{N_t - n'_t} \mu_{tm+\overline{m}}{}^{n'_t} \\
& = P(w'_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}),
\end{aligned} \tag{6.12}
$$

where the second equality uses (3.4) and (3.5). Likewise, we have

$$
P(w_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) = P(w'_N | \text{nature picks } \overline{\mu}_m). \tag{6.13}
$$

Note that then by (3.5)

$$
\begin{aligned}
& P(w_N) \\
& = \textstyle\sum_{m=1}^{\overline{m}} P(w_N | \text{nature picks } \overline{\mu}_m) p_m + \sum_{m=1}^{\overline{m}} P(w_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) p_{m+\overline{m}} \\
& = \textstyle\sum_{m=1}^{\overline{m}} P(w'_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) p_{m+\overline{m}} + \sum_{m=1}^{\overline{m}} P(w'_N | \text{nature picks } \overline{\mu}_m) p_m \\
& = P(w'_N).
\end{aligned} \tag{6.14}
$$

Combining (6.12), (3.5), and (6.14) we obtain for all $m \leq \overline{m}$

$$
\begin{aligned}
P(w_N | \text{nature picks } \overline{\mu}_m) p_m / P(w_N) &= P(w'_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) p_{m+\overline{m}} / P(w'_N), \\
P(w_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) p_{m+\overline{m}} / P(w_N) &= P(w'_N | \text{nature picks } \overline{\mu}_m) p_m / P(w'_N).
\end{aligned} \tag{6.15}
$$

With these preliminaries we can now easily establish (6.10). Namely,

$$
\begin{aligned}
& E(Y_1 | w_N) - E(Y_2 | w_N) \\
& = \textstyle\sum_{m=1}^{2\overline{m}} (\mu_{1m} - \mu_{2m}) P(w_N | \text{nature picks } \overline{\mu}_m) p_m / P(w_N) \\
& = \textstyle\sum_{m=1}^{\overline{m}} (\mu_{1m} - \mu_{2m}) P(w'_N | \text{nature picks } \overline{\mu}_{m+\overline{m}}) p_{m+\overline{m}} / P(w'_N) + \\
& \textstyle\sum_{m=1}^{\overline{m}} (\mu_{1m+\overline{m}} - \mu_{2m+\overline{m}}) P(w'_N | \text{nature picks } \overline{\mu}_m) p_m / P(w'_N) \\
& = \textstyle\sum_{m=1}^{2\overline{m}} (\mu_{2m} - \mu_{1m}) P(w'_N | \text{nature picks } \overline{\mu}_m) p_m / P(w'_N) \\
& = E(Y_2 | w'_N) - E(Y_1 | w'_N),
\end{aligned} \tag{6.16}
$$

where the first and last equalities hold by (6.11), the second equality holds by (6.15), and the third one by (3.5).

(ii) By $\delta_2(n_1, n_2)$ we denote the probability with which δ chooses treatment 2 when n_t successes are observed for $t = 1, 2$. Note that

$$\begin{aligned}
&E_{(\mu_1,\mu_2)}(\delta_2(w_N)) \\
&= \sum_{n_1=0}^{N_1} \sum_{n_2=0}^{N_2} \delta_2(n_1, n_2) \binom{N_1}{n_1} \mu_1^{n_1} (1-\mu_1)^{N_1-n_1} \binom{N_2}{n_2} \mu_2^{n_2} (1-\mu_2)^{N_2-n_2} \\
&= \sum_{k=0}^{N_1} \sum_{l=0}^{N_2} \delta_2(N_1 - k, N_2 - l) \binom{N_1}{N_1 - k} \mu_1^{N_1-k} (1-\mu_1)^k \binom{N_2}{N_2 - l} \mu_2^{N_2-l} (1-\mu_2)^l \\
&= \sum_{k=0}^{N_1} \sum_{l=0}^{N_2} (1 - \delta_2(k, l)) \binom{N_1}{N_1 - k} \mu_1^{N_1-k} (1-\mu_1)^k \binom{N_2}{N_2 - l} \mu_2^{N_2-l} (1-\mu_2)^l \\
&= 1 - E_{(1-\mu_1,1-\mu_2)}(\delta_2(w_N)), \qquad (6.17)
\end{aligned}$$

where the second equation uses a change of the summation indices from $N_1 - n_1$ to k and $N_2 - n_2$ to l and the third equality uses the symmetry property of δ from (3.3) and a well-known property of the binomial coefficient. Note that

$$\begin{aligned}
&R(\delta, (\mu_1, \mu_2)) \\
&= \max\{\mu_1, \mu_2\} - \mu_1 E_{(\mu_1,\mu_2)}(\delta_1(w_N)) - \mu_2 E_{(\mu_1,\mu_2)}(\delta_2(w_N)) \\
&= \max\{\mu_1, \mu_2\} - \mu_1(1 - E_{(\mu_1,\mu_2)}(\delta_2(w_N))) - \mu_2 E_{(\mu_1,\mu_2)}(\delta_2(w_N)). \qquad (6.18)
\end{aligned}$$

Wlog assume that $\mu_1 < \mu_2$, in which case we obtain $R(\delta, (\mu_1, \mu_2)) = (\mu_2 - \mu_1)(1 - E_{(\mu_1,\mu_2)}(\delta_2(w_N)))$. Furthermore,

$$\begin{aligned}
&R(\delta, (1 - \mu_1, 1 - \mu_2)) \\
&= \max\{1 - \mu_1, 1 - \mu_2\} - (1 - \mu_1)(1 - E_{(1-\mu_1,1-\mu_2)}(\delta_2(w_N))) - (1 - \mu_2) E_{(1-\mu_1,1-\mu_2)}(\delta_2(w_N)) \\
&= (1 - \mu_1) - (1 - \mu_1) E_{(\mu_1,\mu_2)}(\delta_2(w_N))) - (1 - \mu_2)(1 - E_{(\mu_1,\mu_2)}(\delta_2(w_N))) \\
&= (\mu_2 - \mu_1)(1 - E_{(\mu_1,\mu_2)}(\delta_2(w_N))), \qquad (6.19)
\end{aligned}$$

where the second equality uses (6.17). It follows that $R(\delta, (1 - \mu_1, 1 - \mu_2)) = R(\delta, (\mu_1, \mu_2))$.It follows that when (μ_1, μ_2) maximizes regret then so does $(1 - \mu_1, 1 - \mu_2)$ if δ is chosen to satisfy (3.3). □

Proof of Lemma 2. In general,

$$\begin{aligned}&R(\delta,(\mu_1,\mu_2))\\&=\max\{\mu_1,\mu_2\}-\mu_1E_{(\mu_1,\mu_2)}(\delta_1(w_N))-\mu_2E_{(\mu_1,\mu_2)}(\delta_2(w_N))\end{aligned}\tag{6.20}$$

and, under the particular sampling design considered here,

$$\begin{aligned}&E_{(\mu_1,\mu_2)}(\delta_2(w_N))\\&=\sum_{n_1=0}^{N_1}\sum_{n_2=0}^{N_2}\delta_2(n_1,n_2)\binom{N_1}{n_1}\mu_1^{n_1}(1-\mu_1)^{N_1-n_1}\binom{N_2}{n_2}\mu_2^{n_2}(1-\mu_2)^{N_2-n_2}.\end{aligned}\tag{6.21}$$

Given that $\delta_2(n_1,n_2)\in[0,1]$ uniform continuity clearly holds. □

Proof of Proposition 6. (i) We will show below that for $s\in\Delta\mathbb{S}$ as in Proposition 6(i),

$$E(Y_t|w_N)=E(Y_{\sigma^{-1}(t)}|w'_N)\text{ for }t\in\{1,...,T-1\}\tag{6.22}$$

holds for all samples w'_N as in (4.1) and permutations σ on $\{1,...,T-1\}$, where σ^{-1} is the inverse of permutation σ defined on the group of permutations.

We first establish that (6.22) implies that a symmetric best response δ to nature's mixed strategy exists. Recall a treatment rule δ° is a best response to nature's strategy iff for all samples w_N with $P(w_N)>0$

$$E(Y_t|w_N)>E(Y_{t'}|w_N)\text{ implies }\delta^{\circ}_{t'}(w_N)=0\tag{6.23}$$

for $t,t'\in\{1,...,T\}$. Note that $E(Y_T|w_N)=\mu_T$ because the mean of the status quo treatment is known.

Without loss of generality, assume that under sample w_N we can order the posterior mean as (with possible relabeling)

$$E(Y_1|w_N)=\cdots=E(Y_{\bar{t}}|w_N)>E(Y_{\bar{t}+1}|w_N)\geq\cdots\geq E(Y_{T-1}|w_N)\tag{6.24}$$

for some $\bar{t}<T-1$. Otherwise, the decision maker will be indifferent with all $T-1$ treatments under both w_N and w'_N in view of (6.22), so he can choose a treatment rule that satisfies the symmetry condition (4.1).

Let E^{max} be the value of $E(Y_1|w_N)$. Using (6.22), we have

$$E^{max} = E(Y_{\sigma^{-1}(1)}|w'_N) = \cdots = E(Y_{\sigma^{-1}(\bar{t})}|w'_N) > E(Y_{\sigma^{-1}(\bar{t}+1)}|w'_N) \geq \cdots \geq E(Y_{\sigma^{-1}(T-1)}|w'_N) \tag{6.25}$$

If $\mu_T > E^{max}$, then under both w_N and w'_N, define δ to assign zero probability to the treatments $1, \cdots, T-1$ and probability 1 to treatment T. Then δ satisfies the symmetry restriction.

If $\mu_T < E^{max}$, then the policymaker is indifferent between treatments $1, \cdots, \bar{t}$ under w_N and indifferent between $\sigma^{-1}(1), \cdots, \sigma^{-1}(\bar{t})$ under w'_N, we can choose δ as a treatment rule that respects the symmetry condition and puts zero probability on treatments $\bar{t}+1, \cdots, T-1$ under sample w_N and zero probability on treatments $\sigma^{-1}(\bar{t}+1), \cdots, \sigma^{-1}(T-1)$ under sample w'_N.

If $\mu_T = E^{max}$, then the policy maker is indifferent between treatments $1, \cdots, \bar{t}, T$ under w_N and indifferent between $\sigma^{-1}(1), \cdots, \sigma^{-1}(\bar{t}), T$ under w'_N. Using the same reasoning as above, one can choose δ such that $\delta_T(w_N) = \delta_T(w'_N)$ and the remaining treatment rule satisfies the symmetry restrictions in (4.1).

We now verify (6.22). Take w_N and w'_N such that $n_{\sigma(t)} = n'_t$ for some permutation σ on $\{1, ..., T-1\}$ as above (4.1). We will hold σ fixed for the rest of the proof. Define the shorthand notation $\mu_m = (\mu_{1m}, ..., \mu_{T-1m}, \mu_T)$, $\mu_{\sigma m} = (\mu_{\sigma(1)m}, ..., \mu_{\sigma(T-1)m}, \mu_T)$, and

$$B(n, \mu_{tm}) = \binom{\overline{N}}{n} \mu_{tm}^n (1 - \mu_{tm})^{\overline{N}-n}. \tag{6.26}$$

We have for any permutation σ'

$$P(w_N|\text{nature picks } \mu_{\sigma' m}) = \prod_{t=1}^{T-1} B(n_t, \mu_{\sigma' m}) = \prod_{t=1}^{T-1} B(n_{\sigma(t)}, \mu_{\sigma\sigma' m}) = P(w'_N|\text{nature picks } \mu_{\sigma\sigma' m}), \tag{6.27}$$

where the second equality uses a change in the order of multiplication, and the last equality uses the definition of w'_N.

Also,

$$\begin{aligned} & P(w_N) \\ &= \textstyle\sum_{m=1}^{\overline{m}} p_m \sum_{\sigma'} ((T-1)!)^{-1} P(w_N|\text{nature picks } \mu_{\sigma' m}) \\ &= \textstyle\sum_{m=1}^{\overline{m}} p_m \sum_{\sigma'} ((T-1)!)^{-1} P(w'_N|\text{nature picks } \mu_{\sigma\sigma' m}) \\ &= P(w'_N), \end{aligned} \tag{6.28}$$

where the summation is over all permutations σ' on $\{1, ..., T-1\}$ and where we use (6.27) and the fact that $\sigma\sigma'$ loops over all possible permutations as we vary σ' in the second equality. The set of all permutations form a group (with the obvious operation), so for each permutation σ, there exists a unique inverse σ^{-1}. We obtain

$$\begin{aligned}
&E(Y_t|w_N)\\
&= \textstyle\sum_{m=1}^{\overline{m}}\sum_{\sigma'}\mu_{\sigma'(t)m}P(w_N|\text{nature picks }\mu_{\sigma'm})p_m/((T-1)!P(w_N))\\
&= \textstyle\sum_{m=1}^{\overline{m}}[\sum_{\sigma'}\mu_{\sigma'(t)m}P(w'_N|\text{nature picks }\mu_{\sigma\sigma'm})]p_m/((T-1)!P(w'_N))\\
&= \textstyle\sum_{m=1}^{\overline{m}}[\sum_{\sigma'}\mu_{\sigma^{-1}\sigma\sigma'(t)m}P(w'_N|\text{nature picks }\mu_{\sigma\sigma'm})]p_m/((T-1)!P(w'_N))\\
&= \sum_{m=1}^{\overline{m}}\sum_{\sigma\sigma'}\mu_{\sigma^{-1}\sigma\sigma'(t)m}P(w'_N|\text{nature picks }\mu_{\sigma\sigma'm})]p_m/((T-1)!P(w'_N))\\
&= E(Y_{\sigma^{-1}(t)}|w'_N),
\end{aligned} \tag{6.29}$$

where the second equality uses (6.27) and (6.28) and the third equality uses the identity permutation can be written as multiplication of σ and σ^{-1}. The fourth equality uses $\sigma\sigma'$ loops over all permutations as we vary σ' and hold fixed σ.

(ii). Let $\delta_t(n_1,...,n_{T-1})$ be the probability that treatment t is chosen when $w_N = (n_1,...,n_{T-1})$ is observed. Let σ be a permutation on $\{1,...,T-1\}$. We use the shorthand notation $\mu = (\mu_1,...,\mu_{T-1},\mu_T)$ and $\mu_\sigma = (\mu_{\sigma(1)},...,\mu_{\sigma(T-1)},\mu_T)$. Note that for $t = 1,...,T-1$

$$\begin{aligned}
&E_\mu(\delta_t(w_N))\\
&= \sum_{n_1=0}^{\overline{N}}\cdots\sum_{n_{T-1}=0}^{\overline{N}}\delta_t(n_1,...,n_{T-1})B(n_1,\mu_1)...B(n_{T-1},\mu_{T-1})\\
&= \sum_{n_1=0}^{\overline{N}}\cdots\sum_{n_{T-1}=0}^{\overline{N}}\delta_{\sigma^{-1}(t)}(n_{\sigma(1)},...,n_{\sigma(T-1)})B(n_1,\mu_1)...B(n_{T-1},\mu_{T-1})\\
&= \sum_{n_{\sigma(1)}=0}^{\overline{N}}\cdots\sum_{n_{\sigma(T-1)}=0}^{\overline{N}}\delta_{\sigma^{-1}(t)}(n_{\sigma(1)},...,n_{\sigma(T-1)})B(n_{\sigma(1)},\mu_{\sigma(1)})...B(n_{\sigma(T-1)},\mu_{\sigma(T-1)})\\
&= \sum_{n_1=0}^{\overline{N}}\cdots\sum_{n_{T-1}=0}^{\overline{N}}\delta_{\sigma^{-1}(t)}(n_1,...,n_{T-1})B(n_1,\mu_{\sigma(1)})...B(n_{T-1},\mu_{\sigma(T-1)})\\
&= E_{\mu_\sigma}(\delta_{\sigma^{-1}(t)}(w_N)),
\end{aligned} \tag{6.30}$$

where (4.1) is used in the second equality, the third equality simply changes the order of summation, and the fourth equality follows from the change in variables $n_{\sigma(t)} \to n_t$ for $t = 1,...,T-1$. The previous derivation implies that also $E_\mu(\delta_T(w_N)) = E_{\mu_\sigma}(\delta_T(w_N))$. Note that

$$
\begin{aligned}
&R(\delta, \mu_\sigma) \\
&= \max\{\mu_{\sigma(1)}, ..., \mu_{\sigma(T-1)}, \mu_T\} - \textstyle\sum_{t=1}^{T-1} \mu_{\sigma(t)} E_{\mu_\sigma}(\delta_t(w_N)) - \mu_T E_{\mu_\sigma}(\delta_T(w_N)) \\
&= \max\{\mu_{\sigma(1)}, ..., \mu_{\sigma(T-1)}, \mu_T\} - \sum_{t=1}^{T-1} \mu_{\sigma(t)} E_{\mu_\sigma}(\delta_{\sigma^{-1}\sigma(t)}(w_N)) - \mu_T E_{\mu_\sigma}(\delta_T(w_N)) \\
&= \max\{\mu_1, ..., \mu_T\} - \textstyle\sum_{t=1}^{T-1} \mu_{\sigma(t)} E_{\mu}(\delta_{\sigma(t)}(w_N)) - \mu_T E_{\mu}(\delta_T(w_N)) \\
&= \max\{\mu_1, ..., \mu_T\} - \textstyle\sum_{t=1}^{T-1} \mu_t E_{\mu}(\delta_t(w_N)) - \mu_T E_{\mu}(\delta_T(w_N)) \\
&= R(\delta, \mu),
\end{aligned}
\tag{6.31}
$$

where we used (6.30) in the third equality and a change in the order of summation in the fourth equality. Consequently, if μ maximizes regret given the symmetric δ, then μ_σ also maximizes the regret for δ. □

Proof of Lemma 3. Note that

$$
R(\delta, \mu) = \max_{t=1,...,T}\{\mu_t\} - \textstyle\sum_{t=1}^{T-1} \mu_t E_\mu(\delta_t(w_N)) - \mu_T(1 - \sum_{t=1}^{T-1} E_\mu(\delta_t(w_N))) \tag{6.32}
$$

and, under the particular sampling design considered here, for $t \in \{1, ..., T-1\}$

$$
E_\mu(\delta_t(w_N)) = \sum_{n_1=0}^{\overline{N}} ... \sum_{n_{T-1}=0}^{\overline{N}} \delta_t(n_1, ..., n_{T-1}) B(n_1, \mu_1) ... B(n_{T-1}, \mu_{T-1}). \tag{6.33}
$$

Given that $\delta_t(n_1, ..., n_{T-1}) \in [0, 1]$ uniform continuity clearly holds. □

References

Aradillas Fernández, A., J. Blanchet, J.L. Montiel Olea, C. Qiu, J. Stoye, and L. Tan (2024), "ϵ-Minimax Solutions of Statistical Decision Problems," unpublished working paper, Cornell University.

Aradillas Fernández, A., J.L. Montiel Olea, C. Qiu, J. Stoye, and S. Tinda (2024), "Robust Bayes Treatment Choice with Partial Identification," arXiv:2408.11621.

Ben-Tal, A., T. Margalit, and A. Nemirovski (2001), "The ordered subsets mirror descent optimization method with applications to tomography," *SIAM Journal on Optimization*, 12, 79–108.

Berger, J. (1985), Statistical Decision Theory and Bayesian Analysis, Second Edition, New York: SpringerVerlag.

Bubeck, S. (2015), Convex Optimization: Algorithms and Complexity. Foundations and Trends in Machine Learning, vol. 8, no. 3-4, 231–358.

Chamberlain, G. (2000), "Econometric Applications of Maxmin Expected Utility," *Journal of Applied Econometrics*, 15, 625–644.

Chen, H. and P. Guggenberger (2024), "A note on minimax regret rules with multiple treatments in finite samples", forthcoming in *Econometric Theory*.

Cucconi, O. (1968), "Contributi all'Analisi Sequenziale nel Controllo di Accettazione per Variabili. *Atti dell' Associazione Italiana per il Controllo della Qualità* 6, 171–186.

Daskalakis, C. and Q. Pan (2014), "A Counter-Example to Karlin's Strong Conjecture for Fictitious Play," 55th IEEE Symposium on Foundations of Computer Science, FOCS.

Daskalakis, C., A. Deckelbaum, and A. Kim (2024), "Near-Optimal No-Regret Algorithms for Zero-Sum Games," unpublished working paper, MIT.

Dominitz, J. and C. Manski (2024), "Comprehensive OOS Evaluation of Predictive Algorithms with Statistical Decision Theory", unpublished working paper, Northwestern University.

Elliott, G., U. Mueller, and M.W.Watson (2015), "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Fudenberg, D. and D.K. Levine (1998). The theory of learning in games. Volume 2. MIT press.

Giacomini, R. and T. Kitagawa (2022), "Robust Bayesian inference for set-identified models", *Econometrica* 89, 1519–1556.

Guggenberger, P., N. Mehta, and N. Pavlov (2024), "Minimax regret treatment rules with finite samples when a quantile is the object of interest," unpublished working paper, Pennsylvania State University.

Gupta, S. and S. Hande (1992), "On some nonparametric selection procedures. In: Saleh, A.K.Md.E. (Ed.), Nonparametric Statistics and Related Topics. Elsevier.

Hirano, K. and J. Porter (2009), "Asymptotics for Statistical Treatment Rules," *Econometrica* 77, 1683–1701.

Ishihara, T. and T. Kitagawa (2021), "Evidence Aggregation for Treatment Choice," unpublished working paper, Brown University. ArXiv: 2108.06473.

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

Kitagawa, T. and A. Tetenov (2018), "Who Should be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

Kitagawa, T., S. Lee, and C. Qiu (2024), "Treatment Choice with Nonlinear Regret," arXiv:2205.08586.

Leslie, D.S and E.J. Collins (2006), "Generalised weakened fictitious play," *Games and Economic Behaviour*, 56, 285–298.

Manski, C. (1988), "Ordinal Utility Models of Decision Making Under Uncertainty," *Theory and Decision*, 25, 79–104.

______ (2004), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 221–246.

______(2007), "Minimax-regret treatment choice with missing outcome data," *Journal of Econometrics,* 139, 105–115.

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

______ and A. Tetenov (2016), "Sufficient Trial Size to Inform Clinical Practice," *Proceedings of the National Academy of Sciences,* 113, 10518–10523.

____________(2019), "Trial Size for Near-Optimal Choice Between Surveillance and Aggressive Treatment: Reconsidering MSLT-II," *The American Statistician* 73, 305–311.

____________(2021), "Statistical Decision Properties of Imprecise Trials Assessing Coronavirus Disease 2019 (COVID-19) Drugs, *Value in Health,* 24, 641–647.

____________(2023), "Statistical Decision Theory Respecting Stochastic Dominance," *The Japanese Economic Review*, 74, 447–469.

Masten, M. (2023), "Minimax-regret treatment rules with many treatments," *The Japanese Economic Review*, 74, 501–537.

Montiel Olea, J.L, C. Qiu, and J. Stoye (2023), "Decision Theory for Treatment Choice with Partial Identification," ArXiv: 2312.17623.

Robinson, J. (1951), "An iterative method of solving a game," *Annals of Mathematics*, 54(2):296–301.

Savage, L. (1954), The Foundations of Statistics, New York: Wiley.

Schlag, K. (2003), "How to minimize maximum regret under repeated decision-making," EUI working paper.

______ (2006), "ELEVEN - Tests needed for a recommendation," EUI working paper, ECO No. 2006/2.

Stoye, J. (2009), "Minimax Regret Treatment Choice with Finite Samples," *Journal of Econometrics*, 151, 70–81.

______ (2012), "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments," *Journal of Econometrics*, 166, 138–156.

Tetenov, A. (2012), "Statistical Treatment Choice Based on Asymmetric Minimax Regret Criteria," *Journal of Econometrics*, 166, 157–165.

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

Yata, Kohei (2023), "Optimal Decision Rules Under Partial Identification," ArXiv:2111.04926.

7 Supplementary Appendix

The Supplementary Appendix provides some additional results not reported in the main body of the paper. Namely, we report additional results from the section on treatment assignment with unbalanced samples and all results from the section on testing the status quo against two innovations.

7.1 Additional Results for Treatment assignment with unbalanced samples

In this subsection we report i) results for additional sample sizes (N_1, N_2) not covered in the main body of the paper in TABLE I, ii) results for additional sample sizes (N_1, N_2) not covered in the main body of the paper in TABLE II, and iii) an example of δ^n.

i) **TABLE I (continued):** Maximal regret of δ^n for different sample sizes, number of iterations, weighting schemes and initializations.

Weighting choice α_n	n^{-1}	n^{-1}	$(5+n)^{-.7}$	$(5+n)^{-.7}$
Initialization δ^1	SO	ES	SO	ES
$(N_1, N_2) = 80$; minimax regret value=.01344651				
Maximal regret of δ^1	.3949134	**.01**46335	.3949134	**.01**46335
Maximal regret of δ^{150}	**.01**45548	**.01**57714	**.0134**618	**.0134**659
Maximal regret of δ^{500}	**.0137**163	**.013**9940	**.0134473**	**.0134**778
Maximal regret of δ^{2000}	**.013**5119	**.013**5638	**.0134467**	**.0134**617
$(N_1, N_2) = 40$; minimax regret value=.01902905				
Maximal regret of δ^1	.3712294	**.0**212825	.3712294	**.0**212825
Maximal regret of δ^{150}	**.019**5494	**.0**203048	**.0190**365	**.0190**399
Maximal regret of δ^{500}	**.019**1818	**.019**3110	**.0190**334	**.0190**980
Maximal regret of δ^{2000}	**.0190**664	**.019**1008	**.019029**6	**.019029**2
$(N_1, N_2) = 20$; minimax regret value=.02694711				
Maximal regret of δ^1	.3493432	**.0**311654	.3493432	**.0**311654
Maximal regret of δ^{150}	**.02**75462	**.02**77504	**.0269**735	**.0269**522
Maximal regret of δ^{500}	**.02**71285	**.02**71833	**.0269**663	**.02694**82
Maximal regret of δ^{2000}	**.0269**900	**.02**70054	**.0269471**	**.026947**7
$(N_1, N_2) = 10$; minimax regret value=.03820907				

Maximal regret of δ^1	**.3**353402	**.0**460039	.3353402	**.0**460039
Maximal regret of δ^{150}	**.038**6786	**.038**7004	**.0382**168	**.0382**227
Maximal regret of δ^{500}	**.038**3319	**.038**3552	**.0382**167	**.038209**1
Maximal regret of δ^{2000}	**.0382**395	**.0382**450	**.0382090**	**.0382090**

ii) **TABLE II (continued):** Maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for various choices of n and (N_1, N_2)

$n \backslash (N_1, N_2)$	$(50, 60)$	$(50, 100)$	$(50, 150)$
1	.016376;.012255	.015723;.014738	.014585;.013844
150	.016272;.001074	.014721;.000518	.013886;.000424
500	.016250;.000557	.014706;.000545	.013894;.001413
2000	.016241;.000030	.014705;.000208	.013850;.000174
I_{2000}	[.016225,.016241]	[.014647,.014703]	[.013781,.013849]

iii) Next, we report δ^{5000} for $N_1 = 10$ and $N_2 = 20$, ES, and $\alpha_n = (5 + n)^{-.7}$. Namely, we tabulate $\delta_2^{5000}(n_1, n_2)$ in a matrix with $N_1 + 1 = 11$ rows and $N_2 + 1 = 21$ columns. The (i, j)-element in the following matrix represents $\delta_2^{5000}(i - 1, j - 1)$, $i = 1, ..., 11$ and $j = 1, ..., 21$.

.0013 1 1.000 1 1.000 1 1.000 1 1.000 1 1.0 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0006 0 .9998 1 1.000 1 1.000 1 1.000 1 1.0 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .4291 1 1.000 1 1.000 1 1.0 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .4111 1 1.000 1 1.0 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .4112 1 1.0 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .50 1 1.000 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .00 0 .5889 1 1.000 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .00 0 .0000 0 .5889 1 1.000 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .00 0 .0000 0 .0000 0 .5709 1 1.000 1 1.000
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .00 0 .0000 0 .0000 0 .0000 0 .0002 1 .9994
.0000 0 .0000 0 .0000 0 .0000 0 .0000 0 .00 0 .0000 0 .0000 0 .0000 0 .0000 0 .9987

7.2 Additional results for treatment assignment when testing innovations

We implement Algorithm 3 for the case $T = 3$, sample sizes $\overline{N} = N_1 = N_2 \in \{5, 10, 20, 30, 40, 50, 100, 200\}$ and known mean of the status quo treatment $\mu_3 \in \{.2, .5, .8\}$. The initialization rule is the empirical success rule described in the beginning of Algorithm 3. We choose

$p = 1000$ and $\alpha_n = (5+n)^{-.7}$. TABLE IV (continued) reports maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for various choices of n and I_{2000} for $\overline{N} = N_1 = N_2 \in \{5, 20, 30, 40\}$.

Note that if $\mu_T = 0$ then it is always (at least weakly) better for the policymaker to switch to one of the alternative treatment. Therefore, in that case the problem is reduced to choosing between $T-1$ treatments which (for the case of equal sample sizes) has been dealt with in Chen and Guggenberger (2024). On the other hand, if $\mu_T = 1$ it is always (at least weakly) better for the policymaker to remain with the status quo no matter what the data says. Therefore, it is enough to consider mean values $0 < \mu_T < 1$.

TABLE IV (continued): Maximal regret of δ^n and $R(\delta^n, \mu^n_{BR}) - R^n$ for several n and I_{2000} for several $\overline{N}$ and μ_3.

μ_3	0.2	0.5	0.8
$\overline{N} = 40$			
$n = 1$	.020020;.004062	.023973;.023973	.021501;.021501
$n = 150$	.019029;.000012	.023511;.000744	.018796;.000620
$n = 500$	.019029;.000021	.023491;.000058	.018931;.001456
$n = 2000$	.019029;.000000	.023443;.000161	.018725;.000125
I_{2000}	[.019028;.019029]	[.023434;.023438]	[.018673;.018707]
$\overline{N} = 30$			
$n = 1$	.023387;.004777	.027684;.027684	.025266;.025266
$n = 150$	.021991;.000051	.027122;.000160	.021773;.000514
$n = 500$	.021987;.000042	.027169;.000359	.021721;.001021
$n = 2000$	.021982;.000009	.027075;.000223	.021625;.000078
I_{2000}	[.021979;.021982]	[.027060;.027061]	[.021593;.021603]
$\overline{N} = 20$			
$n = 1$	.029221;.006025	.033911;.033911	.031851;.031851
$n = 150$	.027047;.000061	.033271;.001901	.026575;.000348
$n = 500$	.027025;.000027	.033155;.000083	.026521;.000120
$n = 2000$	.027009;.000011	.033185;.000080	.026492;.000256
I_{2000}	[.026999;.027000]	[.033133;.033143]	[.026470;.026476]

$\overline{N}=5$			
$n=1$	.065019;.014349	.064435;.010440	.073224;.073224
$n=150$	.057589;.000655	.064419;.000924	.053662;.000179
$n=500$	.057316;.000276	.064210;.000140	.053707;.000225
$n=2000$	.057185;.000167	.064235;.000047	.053740;.000163
I_{2000}	[.057086;.057086]	[.064190;.064190]	[.053593;.053593]