Eleven – Designing Randomized Experiments under Minimax Regret[1]

Karl H. Schlag[2]

March 12, 2007

[1] The author would like to thank all those that provided support in this adventure, in particular Dean Foster, Charles Manski, Thomas Ronde, Avner Shaked and Richard Spady (in alphabetical order).

[2] Economics Department, European University Institute, Via della Piazzuola 43, 50133 Florence, Italy, Tel: 0039-055-4685951, email: schlag@eui.eu

Abstract

Assume that there are two alternative actions or treatments, each generating a random payoff in [0,1] with unknown distribution. Consider the objective to design a randomized experiment involving a given number of tests and then to select the action with the higher mean.

We present the binomial average rule and prove that it minimizes maximal regret as axiomatized by Milnor (1954). 11 tests are needed to guarantee regret to be below 5%. Neither conditioning later tests on earlier outcomes nor availability of counterfactual evidence can be used to reduce the value of minmax regret.

Keywords: statistical decision making, binomial average rule, empirical success, experimental design, sequential testing, clinical trials, clinical equipoise.

JEL classification numbers: C44, D81, C72, C90.

1 Introduction

Consider the following statistical decision making problem faced by a so-called decision maker. There are two alternative actions (or treatments), each action is associated to a real random variable with range $[0, 1]$ and unknown mean. The decision maker wishes to choose (or recommend) the action that has the higher mean and gathers information about each action by conducting a *randomized experiment*. A randomized experiment consists of a given number of N tests or trials. In each test the decision maker selects one of the two actions and then observes a random payoff generated by the random variable associated to the action selected. Tests are assumed to be conducted sequentially so that the decision about which action to test next may be conditioned on previous outcomes during testing. Examples presented in more detail in Eozenou et al. (2006) include choosing whether or not to introduce computer assisted learning in Indian schools and investigating in which country offers in the Ultimatum Game are highest.

How many tests are needed to be able to make a choice that yields an expected payoff within ρ% of the true maximal expected payoff regardless of the true distribution? Do tests have to be conducted sequentially or is it enough if the decision maker decides at the outset how many tests to conduct on each action? Could counterfactual evidence improve performance? We solve the underlying problem exact for any error percentage ρ.[1] The title of this paper stems from the fact that only 11 tests are needed when $\rho = 5\%$. If more generally the associated random variables are known to have range $[w_1, w_2]$ then the 5% are measured in terms of $w_2 - w_1$. For the first time in such a nonparametric environment only small samples are needed to derive valuable conclusions. Small sample sizes are typical in many applications, e.g. in the four examples in Eozenou et al. (2006) sample sizes are between 50 and 60.

The underlying choice is typically randomized and we show how to obtain a nonrandomized method that only requires 47 observations when $\rho = 5\%$.

We start by presenting the decision making criterion and then provide an overview of the results and how they relate to the existing literature.

1.1 Minimax Regret

Before running the first test the two actions are considered indistinguishable. The decision maker has no strong belief regarding which is better. In the context of clinical trials, collective uncertainty in this respect is described by “clinical equipoise” (cf. Freedman, 1987). The decision criterion underlying our objective is distribution free, the given bound ρ has to be met for all distributions as any one of them could be the true one. This criterion should only depend on knowledge, we want to consider

[1] The term ‘exact’ refers to results obtained for finite sample sizes.

learning or inference in its most basic sense. In particular we do not allow for a prior. However, our results put a tight upper bound (across the set of priors) on how many tests are needed if the decision maker instead uses a prior to solve this problem. At most 11 tests are needed if $\rho = 5\%$.

The underlying decision criterion is called *minimax regret. Regret* introduced by Savage (1951) is a particular *loss function* (Wald, 1950) that depends on the choice and the underlying true distributions and is defined as the difference between the expected payoff of the better action and the expected payoff underlying the recommendation of the decision maker. For illustration, assume that Y_j is the random variable associated to choice of action j and that p_j is the probability that the decision maker chooses action j $(j = 1, 2)$. Then regret r is given by

$$r = \max\{E_P Y_1, E_P Y_2\} - (p_1 E_P Y_1 + p_2 E_P Y_2)$$

where regret is defined in terms of taking expectations based on the true underlying distribution P. Thus,

$$r = \max\{(E_P Y_1 - E_P Y_2)\, p_2, (E_P Y_2 - E_P Y_1)\, p_1\}$$

which highlights the difference to classical hypothesis testing. Instead of being interested in type I and type II error, the probability of making the wrong choice is adjusted for the value of making the correct choice. Given this definition of regret, the choice of the decision maker is evaluated in terms of its maximal regret across all possible distributions P. Under the minimax regret criterion the decision maker then makes a recommendation that yields the lowest maximal regret. Note that the recommendation will depend on the outcome of the testing and on how the randomized experiment was designed.

It is important to stress that our concern for maximal regret is not motivated behaviorally nor does it require that the decision maker learns the true underlying distribution ex-post. It is the axioms that found minimax regret (Milnor, 1954, Stoye, 2006) that we choose as basis to justify decision making under minimax regret. We briefly highlight why we find these axioms appropriate for this setting, using the axiomatization of Stoye (2006) to connect minimax regret to subjective expected utility theory (SEU, Anscombe and Aumann, 1963).

We are interested in decision making under complete uncertainty, wish to base the decision only on (verifiable) knowledge. Thus the Symmetry axiom is postulated. It assesses that recommendations may not depend on how actions or states of the world are labelled. Non-verifiable elements such as subjective priors are ruled out. The Symmetry axiom is not compatible with all axioms underlying SEU and hence they cannot all be postulated. Independence of Irrelevant Alternatives (IIA), sometimes mistakenly associated to rationality per se, does not allow for decisions to depend on the set of actions available. It precludes to incorporate all verifiable information and will be relaxed as follows. Independence of the most preferred choice is only imposed when adding actions that would not change payoffs achievable in the perfect information setting (see the INA axiom in Stoye, 2006). The resulting minimax regret criterion is called menu or opportunity dependent as choice depends on the set of actions available (for some examples where opportunity dependence yields valuable insights see Hayashi, 2006).

An alternative way (see Milnor, 1954, Stoye, 2006) to adapt the axioms underlying SEU after postulating the Symmetry axiom is to maintain IIA but to weaken the Independence Axiom, an axiom that can be associated to time consistency in decision making. This results in the maximin criterion of Wald (1950). However, as already pointed out by Manski (2005) in a closely related setting, maximin criterion cannot capture learning in its common sense in the setting of this paper. The rule to choose action one regardless of the observations in the sample attains maximin regardless of how large the sample is.

Minimax regret has been used to investigate bilateral bargaining (Linhart and Radner, 1989), monopoly pricing (Bergemann and Schlag, 2006, ch. 3), general equilibrium with overlapping generations (Cozzi and Giordani, 2006), regression (Droge, 1998, Eldar et al., 2004) and learning in two-armed bandits (Berry and Fristedt, 1985, Lai and Robbins, 1985, Schlag, 2003) among other topics. The main references on minimax regret in statistical decision making for this paper are Canner (1970), Manski (2004) and Stoye (2005).

1.2 Binary Valued Payoffs and Balanced Samples

We begin our overview of the model and the related literature by first considering a less general setting in which payoffs are binary valued (so payoffs are either 0 or 1) and where the recommendation has to be made based on a *balanced sample* (each action is tested $N/2$ times where N is even).

Following Canner (1970), minimax regret is attained when using the *simple success rule*: select the action that yielded the higher average payoff during testing, randomize equally likely whenever there is a tie. Stoye (2006) provides an alternative proof using a saddle point characterization and provides a formula for deriving minimax regret. We show how this formula can be obtained directly using results from the literature on selection procedures.

Sobel and Huyett (1957) are interested in the minimal probability of correct selection (i.e. choosing the action with the highest mean) for a given lower bound $d > 0$ on the difference between the two means, so $|E_P Y_1 - E_P Y_2| \geq d$. They suggest without formal justification to use the simple success rule. The minimal probability of correct selection under this rule is investigated by using simulations and by identifying asymptotically the least favorable distribution where the probability of correct selection is lowest. Hoel (1972) analytically derives the least favorable distribution for an alternative selection procedure which, given later results of Bechhofer and Kulkarni (1982), is also a least favorable distribution for the rule of Huyett and Sobel (1957). We demonstrate how the formula for deriving minimax regret presented in Stoye (2006) follows immediately from these results and the relationship illustrated above between regret and the probability of correct selection. $N = 12$ observations (6 of each action) is sufficient to obtain a value of maximal regret below 5%. For large samples we show that the value of minimax regret is approximately equal to $0.17 * N^{-1/2}$.

1.3 Nearly Balanced Samples

Assume now that the recommendation has to be made based on an odd number of tests and a sample that is *nearly balanced* in the sense that each action is equally likely tested once more than the other action. We build again on findings in the literature on selection procedures. Bechhofer and Kulkarni (1982) show how one can stop testing early when implementing the simple success rule by anticipating the final recommendation. In particular they show that one will never stop after an even number of tests. Using their insights we derive a recommendation rule for nearly balanced samples (that we call the *adjusted success rule*) that yields the same probability of correct selection as the simple success rule but requires one less test. In particular, the adjusted success rule attains minimax regret when the sample is nearly balanced. Consequently, only $N = 11$ observations are needed to guarantee that regret is below 5%.

1.4 Endogenous Testing

Now allow the decision maker to design the randomized experiment by choosing which actions to test. Two different scenarios can be imagined. Under *simultaneous testing* the decision maker simultaneously determines the number of tests of each action, constrained by a given total number of tests N. Under *sequential testing* the decision maker runs one test after the other, is able to observe the outcome after each test and is allowed to condition future tests on the outcomes of all previous tests. While there has been a large interest in the sequential design of randomized experiments (e.g. Jennison and Turnbull, 2000), we know of no nonparametric exact results, neither for sequential selection procedures nor for hypothesis testing, that show how to design these optimally. We prove how to design the randomized experiment in view of attaining minimax regret.

For the simultaneous testing setting, we prove that it is best to test each action as equally often as possible. While this result is intuitive it does not apply as such when there are three actions, a case we discuss in Section 5.6. Turning to sequential testing one would imagine that the decision maker can reduce maximal regret by conditioning on previous test outcomes. For instance, consider a decision maker who encounters after running m tests on each action $(2m < N)$ that action 1 yielded outcome $2/3$ in all tests while action 2 yielded outcomes 0 and 1 equally often. Can the decision maker ignore this information and go on testing each action equally often? The answer is yes although counter intuitive. We prove that there is no added value under minimax regret to design the randomized experiment sequentially.

Our proof involves game theory. Following von Neumann and Morgenstern (1944, see also Savage, 1954) minimax regret problems can be solved by finding a Nash equilibrium of a zero sum game between the decision maker and nature where the decision maker aims to minimize regret while nature's goal is to maximize regret. The equilibrium strategy of the decision maker then attains minimax regret.[2] The proof that there is no value added to sequential testing in our setting follows almost directly from a property of the least favorable distributions, those distributions that maximize regret. Building on the connection to Hoel (1972) and Bechhofer and Kulkarni (1982) mentioned above it follows that there is a distribution that has support in $\{(0,1),(1,0)\}$ that maximizes regret under the simple success rule. Consider nature putting equal weight on two such distributions, one derived from the other by interchanging the labels of the two actions. For a given balanced sample, the simple success rule is a best response and hence attains minimax regret (a similar line of proof can be found in Stoye, 2006). The additional implications for sequential testing in this paper are novel but simple. When facing this strategy of nature, the same information is gathered regardless of which action is tested or even whether the payoffs to both actions can be observed simultaneously. Consequently, maximal regret cannot be lowered by some elaborate sequential testing scheme or if we additional assume that counterfactual evidence is available. Concern for the worst case yields a simple solution.

We hasten to add that sequential testing can of course be used to economize on the number of tests without increasing maximal regret. In fact we present such a rule and call it the *truncated success rule.* However, economizing on tests is not of immediate concern for this paper as testing is assumed costless. The full analysis of a model with costly testing is too intricate and we only present some initial insights in a later section.

1.5 Bounded Payoffs

Finally, turn to the actual setting of this paper where payoffs are not necessarily binary valued but instead are only known to be contained in $[0,1]$. Manski (2004) investigates performance of the *empirical success rule*, a rule that compares average payoffs in a balanced sample. Manski (2004, eq. 23) derives a loose upper bound

[2] Schlag (2003) and Stoye (2006) similarly use this method to derive strategies that attain minimax regret.

to show that the empirical success rule requires at most 148 tests to guarantee that regret is below 5%. Stoye (2005) runs some numerical simulations that indicate that at least 20 tests are needed for $\rho = 5\%$.

We extend all results mentioned above for the setting of binary valued payoffs to the nonparametric case where payoffs belong to $[0,1]$. In particular, only $N = 11$ tests are needed, but not less, in order to guarantee that regret is below 5%. It is this finding that motivated the title of this paper.

This extension from $\{0,1\}$ to $[0,1]$ is possible due to a simple trick used in Schlag (2003). Minimax regret can be attained for the case where payoffs belong to $[0,1]$ by first randomly transforming the sample into a binary valued sample and then applying the rule that attains minimax regret when payoffs are binary valued. The particular random transformation used will be referred to as the *binomial transformation.* Accordingly, each payoff in the sample that is contained in the interval $(0,1)$ is independently randomly transformed into an outcome in $\{0,1\}$, using the payoff observed as the probability that the data point is transformed into outcome 1. The decision maker strategically throws away information.[3] As this transformation is mean preserving, it is as if the decision maker faces a binary valued distribution which weakly lowers maximal regret as compared to facing any distribution with support in $[0,1]^2$.

We thus obtain a rule that attains minimax regret for all N for the setting where payoffs are known to be contained in $[0,1]$. The rule is called the *binomial average rule* and can be described as follows. Start with testing either action equally likely and continue testing the two actions alternatingly until N tests have been run. Binomially transform the sample into a binary valued sample and then make a recommendation based on the simple or the adjusted success rule, depending on whether N is even or odd.

Schlag (2007) presents this randomization trick for more general settings including nonparametric estimation and hypothesis testing. Recently we discovered that this method was also used by Cucconi (1968) to derive an exact nonparametric sequential probability ratio test. We have also discovered meanwhile that Gupta and Hande (1992), without citing Canner (1970) or Hoel (1972), have shown for the case of a balanced sample with bounded payoffs that behavior as prescribed by the binomial average solves the related objective, to maximize the minimal probability of choosing the action with the highest mean.

To complete this overview we clarify the connection to other recent papers on minimax regret in chronological order. Schlag (2003) considers minimax regret in a two-armed bandit setting with a sufficiently impatient decision maker where, unlike the present paper, there is a genuine trade-off between exploitation and exploration. The related results by Stoye (2005) that existed before the first version of this paper

[3] The decision maker is only interested in the means. Using terminology from statistics, all aspects of the distribution apart from the mean are 'nuisance parameters'.

are confined to binary valued payoffs and exogenously given balanced samples (so N is even). Eozenou et al. (2006) present an alternative rule that reduces variance in the recommendation of the binomial average rule, numerical simulations reveal that it also attains minimax regret. Schlag (2006b) builds on the present paper and further investigates nonparametric learning for the case where each test reveals the payoff of both actions, counterfactual evidence thus being available.

In Section 2 the decision problem is introduced. In Section 3 the necessary criteria for making decisions and measuring performance are introduced. Section 4 presents important examples and derives some properties. In Section 5 we investigate minimax regret, proving minimax regret property of the binomial average rule and then discussing a variety of extensions. In Section 6 we briefly consider the alternative criterion of maximin. The appendix contains the specific analysis of minimax regret for three actions and three tests as well as all tables.

2 Setting

Consider a decision maker who has to recommend (or choose) one of two possible actions (or treatments) labelled 1 and 2. Choice of action $j \in \{1, 2\}$ generates a random outcome belonging to a set $\mathcal{Y}$ where $\mathcal{Y}$ contains at least two elements. Outcomes are drawn from an unknown joint distribution (or *environment*) $P \in \Delta\left(Y^2\right)$. Specifically, the random outcome generated from choosing action j is drawn from the marginal P_j of P with respect to the j-th component.[4] Depending on the specification of P, actions can, but need not generate independent outcomes.

Even if outcomes of each action are only measured in terms of success and failure, $\mathcal{Y}$ will contain more than two elements unless treatment specific side effects are absent. The case in which $\mathcal{Y}$ contains only two elements will play a special role for the analysis.

Before making a recommendation the decision maker is allowed to perform a *randomized experiment* which consists of N independent tests (or trials). In each such test the decision maker chooses an action and then observes a random payoff generated by this action. This testing procedure is also called the *test phase*, the choice thereafter is also called the *recommendation* or *final choice.* So a *strategy* of the decision maker consists of two parts: (i) which actions to test in the test phase and (ii) which action to recommend based on the observations in the test phase. We consider two alternative informational settings for modelling the test phase. We say that testing is *sequential* if tests are run sequentially with feedback on the outcome of each test being available before the next test and if the choice of which action to test next may be conditioned on all previous outcomes. Testing is called *simultaneous* if

[4] ΔA denotes the set of distributions over the set A. Any element $a \in A$ is identified with the distribution that places probabilty 1 on a, hence $A \subset \Delta A$.

the decision maker has to pre-commit to the total number tests run on each action before starting the test phase.

We now develop notation for describing what the decision maker does. In the next subsection we specify the objectives for how to determine what the decision maker should do.

Consider first sequential testing. In the spirit of game theory it is important to use formal notation to describe behavior when deriving exact results. A *strategy* (or rule) formally describes how the decision maker first designs tests and then makes a recommendation based on the outcome of testing. After running $m \in \{1, .., N\}$ tests, the *history h of length m* is given by $h = ((j_1, y_1), (j_2, y_2), ..., (j_m, y_m))$ where j_k is the action chosen in the k-th round of the test phase and y_k is resulting outcome. So y_k is an outcome randomly drawn from the distribution P_{j_k} independently of any earlier event. The strategy f of the decision maker for running sequential tests is to assign to each history h of length m with $m < N$ the action to test next and after testing is over to decide based on the history of length N which action to recommend. We allow for the decision maker to randomize over actions both during the test phase and when making the recommendation.[5] Formally,[6]

$$f : \cup_{m=0}^{N} (\{1, 2\} \times \mathcal{Y})^m \to \Delta \{1, 2\}$$

where $f(h)_j$ denotes the probability of testing action j after history h of length $m < N$ and denotes the probability of recommending action j if the history is of length N. The recommendation only refers to how the strategy f evaluates histories of length N.

The recommendation is called *deterministic* if $f(h)_1 \in \{0, 1\}$, it is called *randomized* if $f(h)_1 \in (0, 1)$, where h represents here the history of events during the test phase. A strategy f is called *symmetric* if it does not depend on how actions are labelled. Formally this means that $f(\emptyset)_1 = 1/2$ and that $f(h)_j = f(h^s)_{3-j}$ where $h_k^s = ((3 - j_k, y_2))$ for $k = 1, .., m$ given $h \in (\{1, 2\} \times \mathcal{Y})^m$ with $1 \leq m \leq N$.

While most of this paper considers costless testing there is no need to run all tests. When given the possibility of running less than N tests we need to define a stopping rule that is associated to the strategy. Briefly, let s be the stopping rule where, given history h, $s(h) = 1$ specifies to continue testing while $s(h) = 0$ specifies to stop testing and to make a recommendation based on the gathered sample.

Consider now the more restricted setting of simultaneous testing. The decision maker simultaneously assigns the number N_j of tests to be run on each action j, $j = 1, 2$, so $N_1, N_2 \in \mathbb{N}_0$ with $N_1 + N_2 = N$. The tuple (N_1, N_2) is also called an *assignment*, $N_{\{1,2\}} := \{(N_1, N_2) \in \mathbb{N}_0^2 \text{ s.t. } N_1 + N_2 = N\}$ denotes the set of all possible assignments. The history h of observations generated by the set of observations

[5] To keep notation simple, we describe strategies below in terms of behavioral strategies.

[6] The convention $(\{1, .., T\} \times [0, 1])^0 = \{\emptyset\}$ is used.

during the test phase is similar to the one under sequential testing except that it is now unordered, so $h = \{(j_1, y_1), (j_2, y_2), ..., (j_m, y_m)\}$. Formally, the strategy f is defined by a mapping

$$\begin{aligned} f &: \emptyset \to \Delta N_{\{1,2\}} \\ f &: (\{1,2\} \times \mathcal{Y})^N \to \Delta \{1,2\} \end{aligned}$$

where $f(\emptyset)_{(N_1,N_2)}$ denotes the probability of selecting assignment $(N_1, N_2) \in N_{\{1,2\}}$. Of course any rule for simultaneous sampling can be embedded in the formal framework of sequential sampling. Thus we sometimes consider rules for simultaneous sampling as a subset of the set of rules for sequential sampling. If not mentioned otherwise we will be considering the sequential testing scenario.

3 Minimax Regret, Priors and Learning

We consider a decision maker who searches for a strategy that attains minimax regret. First we show how to implement this decision making criterion then add a brief discussion.

In order to implement minimax regret we have to specify the *alternatives*, the *states of the world* and the *consequences*. The strategies defined above are the alternatives (or acts). The underlying distribution P is the unknown state of the world. We assume that the decision maker is only interested in the recommendation and does not care about the outcomes during testing. Hence the distribution of outcomes generated by the recommendation is the consequence. So if action j is recommended and the true distribution is given by P then P_j is the associated consequence.

It can be natural to ignore outcomes during testing when those tested constitute only a vanishing minority among the population of those needing the treatment (see Manski, 2004). The separation of testing and recommendation parallels classical hypothesis testing where the major concern is for size and power with outcomes arising during tests being only secondary. Of course an analysis of costly testing as considered by Canner (1970) is also an interesting research agenda and we come back to this later.

The next step is to describe the rational preferences of the decision maker over the set of consequences. These preferences are assumed to satisfy the axioms of von Neumann and Morgenstern (1947) and hence have an expected utility representation. Thus there exists a utility function $u : \mathcal{Y} \to \mathbb{R}$ such that, if the decision maker would know P, then he or she would recommend the action that maximizes expected utility. More specifically, let $u_j(P)$ denote the expected payoff generated by choosing action j so

$$u_j(P) := \int_{y \in \mathcal{Y}^2} u(y_j)\, dP(y) = \int_{y_j \in \mathcal{Y}} u(y_j)\, dP_j(y_j).$$

If the decision maker would know P then he or she would choose j such that $u_j(P) \geq u_k(P)$. Actions contained in $\arg\max_{j\in\{1,2\}} u_j(P)$ are called *best* (given P).

For simplicity we sometimes work with a reduced representation in which outcomes are identified with their utility, $\mathcal{Y}$ is identified with $\{u(y), y \in \mathcal{Y}\}$ with $u(y)$ also called a *payoff*, and then associate choice of action j with the random variable that realizes the payoff $u(y_j)$ with y_j drawn from P_j. In this representation, best actions are the ones that yield the highest mean.

We add an assumption on the set of outcomes $\mathcal{Y}$ and assume that there is a least preferred element y_L and a most preferred element y_H of $\mathcal{Y}$ where these two elements are not identical. Then we can normalize u by a positive affine transformation so that $u(y_L) = 0$ and $u(y_H) = 1$. Thus all payoffs are contained in $[0,1]$. If $\mathcal{Y} = \{0,1\}$ then we also refer to payoff 0 as a *failure* and to payoff 1 as a *success*. If $P \in \Delta\{0,1\}^2$ then we say that *payoffs are binary valued.*

The axioms leading to minimax regret criterion specify how the decision maker deals with not knowing the true distribution P. The following notation will be needed. Let $p_j(f,P)$ denote the probability of recommending action j when using strategy f and facing distribution P where this probability is calculated ex-ante before running any test. $p_j(f,P)$ will also be called the *recommendation probability* under f given P.[7] Let $u(f,P)$ denote the *expected* payoff of the recommendation induced by using strategy f when facing distribution P where expectations are calculated based on the distribution P ex-ante before starting the test phase, so

$$u(f,P) = \sum_{j=1}^{2} p_j(f,P)\, u_j(P).$$

Regret $r(f,P)$ is defined as the difference between the expected payoff of the action that maximizes utility given P and the expected payoff realized by using strategy f when facing by P. Formally

$$\begin{aligned} r(f,P) &= \max_{j\in\{1,2\}} u_j(P) - u(f,P) \\ &= \max\{(u_1(P) - u_2(P))\, p_2(f,P), (u_2(P) - u_1(P))\, p_1(f,P)\}. \end{aligned}$$

We are now able to formulate how to make a recommendation without knowing P in the minimax regret framework. Under minimax regret, the decision maker who chooses strategy f is concerned with the maximal regret attained by f, accordingly $\hat{P} \in \arg\max_{P\in\Delta\mathcal{Y}^2} r(f,P)$ is called a *least favorable distribution* under f. Consequently, the decision maker minimizes this maximum regret. Formally, the strategy f^* attains *minimax regret* if

$$f^* \in \arg\min_f \sup_{P\in\Delta\mathcal{Y}^2} r(f,P)\ .$$

[7] We refrain from presenting a formal expression for $p_t(f,P)$ as it is too intricate to be insightful. Explicit calculations in later examples will demonstrate better how $p_t(f,P)$ can be derived.

$r_N^* := \inf_f \sup_{P\in\Delta\mathcal{Y}^2} r(f,P)$ is called the *value of minimax regret.*

The minimax regret criterion is due to Savage (1951). Milnor (1954) was the first to present an axiomatic characterization. Recently Stoye (2006) has presented an axiomatization that can be directly compared to the axioms underlying subjective expected utility maximization. It is the axiomatic foundation that leads us to choose this criterion. The Symmetry axiom, that does not allow choice to depend on labels, plays a central role in ensuring that non-verifiable elements such as subjective priors do not play a role (for more details see Stoye, 2006). The alternative of maximin is discussed in Section 6 below.

If the decision maker instead would be a subjective expected utility maximizer (Anscombe and Aumann, 1963) then there would exist some *prior* $Q \in \Delta\Delta\mathcal{Y}^2$ such that the strategy f would be a *best response* to Q in the sense that it would be chosen to maximize $u(f,Q) := \int u(f,P)\,dQ(P)$ (provided a maximizer exists). The value of minimax regret can be interpreted as an upper bound on how difficult it is for a subjective expected utility maximizer to learn. For this interpretation, the difficulty of learning under prior Q is measured as the difference between the expected payoff of the decision maker with perfect and imperfect information, assuming that nature draws P according to Q. The observation follows when verifying

$$\begin{aligned} & \int \max_{j\in\{1,2\}} u_j(P)\,dQ(P) - \sup_f \int u(f,P)\,dQ(P) \\ = \; & \inf_f r(f,Q) \le \sup_{Q'} \inf_f r(f,Q') \le \min_f \sup_{Q'} r(f,Q') = r_N^* \end{aligned}$$

where $r(f,Q) := \int r(f,P)\,dQ(P)$.[8]

Typically the upper bound $\sup_{Q'} \inf_f r(f,Q')$ on the difficulty of learning with a prior is equal to r_N^*. This equality emerges when the minimax regret strategy satisfies a saddle point condition. This saddle point condition (see (1) below) is now presented as it will be used to derive minimax regret. Assume that the strategy f^* and prior Q^* satisfy

$$r(f^*,Q) \le r(f^*,Q^*) \le r(f,Q^*) \;\; \forall f,Q. \tag{1}$$

In other words, (f^*,Q^*) is a Nash equilibrium of the fictitious zero sum game between the decision maker and nature where the decision maker aims to minimize regret while nature aims to maximize regret. Note that the support of Q^* only contains least favorable distributions under f^*. Following the minimax theorem of von Neumann and Morgenstern (1947), f^* attains minimax regret and

$$\sup_Q \inf_f r(f,Q) = r(f^*,Q^*) = r_N^*.$$

As $Q^* \in \arg\max_Q \inf_f r(f,Q)$, Q^* can considered as a prior under which learning is most difficult and is called a *least favorable prior.*

[8] We replace max by sup and min by inf to ensure existence.

As we consider decision making based on information gathered while testing one may wish to qualify whether our approach to choice using minimax regret is associated to some form of learning. As mentioned above, the value of minimax regret can be directly interpreted as a measure for learning as it presents an upper bound on the impact of imperfect information in the more traditional subjective expected utility model. We mention some alternative ways to measure learning. Consider a sequence of rules $\left(\hat{f}_N\right)_{N\in\mathbb{N}}$ such that $\hat{f}_N$ attains minimax regret for sample size N. One could consider learning in terms of asymptotic performance as the sample size tends to infinity and require that $\left(\hat{f}_N\right)_N$ is a *pointwise consistent estimator* of a best action, namely that $\limsup_{N\to\infty} p_j\left(\hat{f}_N, P\right) > 0$ implies $u_j(P) \geq u_{3-j}(P)$. Given that one never knows whether the sample is sufficiently large it is more natural to qualify learning in terms of sample size and not conditional on the underlying distribution. To do this one introduces a wedge between the expected utility of the two actions and considers the set of distributions $\mathcal{P}_\delta$ that contain only those distributions in which the means of the two actions differ by at least δ so $\mathcal{P}_\delta = \{P \in \Delta\mathcal{Y}^2 \text{ s.t. } |u_1(P) - u_2(P)| \geq \delta\}$. Then $\left(\hat{f}_N\right)_N$ is a *uniformly consistent estimator* of the best action (within the class $\mathcal{P}_\delta$) if for every $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $u_j(P) \geq u_{3-j}(P) + \delta$ and $N \geq N_\varepsilon$ then $p_j\left(\hat{f}_N, P\right) \geq 1 - \varepsilon$. One may wish to qualify learning in a dynamic context by investigating how choice changes with the sample size. $\left(\hat{f}_N\right)_N$ is called *ex-ante improving* (Schlag, 2002, cf. Börgers et al., 2004) if $u\left(\hat{f}_{N+1}, P\right) \geq u\left(\hat{f}_N, P\right)$, it is called *ex-ante strictly improving* if $u_j(P) > u_{3-j}(P)$ and $p_j\left(\hat{f}_N, P\right) < 1$ then $u\left(\hat{f}_{N+1}, P\right) > u\left(\hat{f}_N, P\right)$. Since there are only two actions, changes in expected utility can also be translated into monotonicity of the probability of choosing a best action.[9] In this paper we will obtain rules that attain minimax regret that are associated to the forms of learning described above.

4 Some Examples

Before we derive a strategy that attains minimax regret we present some examples. Both historically and for this paper is important to first consider strategies designed for the case where payoffs are binary valued. So first we consider $P \in \Delta\{0,1\}^2$.

[9] Related concepts that are conditional on the history are absolute expediency (Lakshimivarahan and Thathachar, 1973) and monotonicity (Börgers et al, 2004).

4.1 Binary-Valued Payoffs

Consider the most natural candidate for learning which action is better under binary valued payoffs, called the *simple success rule* and defined for simultaneous testing and even N. This strategy specifies to test each action equally often and then to recommend the action that realized more successes during testing, if there is a tie to recommend each action equally likely. This rule plays a central role in the literature on selection procedures under binary valued payoffs starting with Sobel and Huyett (1957) and also emerges from the analysis of Canner (1970).[10]

The *adjusted success rule* is defined for simultaneous testing and odd N. Accordingly, create a so-called *nearly balanced sample* by testing each action equally likely once more than the other action, so

$$f(\emptyset)_{\left(\frac{N-1}{2},\frac{N+1}{2}\right)} = f(\emptyset)_{\left(\frac{N+1}{2},\frac{N-1}{2}\right)} = \frac{1}{2}.$$

Recommendation follows the selection rule of Hoel (1972). Let S_j be the number of successes and F_j be the number of failures observed in the previous tests of action j. Then recommend action j if $S_j + F_{3-j} \geq (N+1)/2$. Note that there is a unique recommendation as N is odd. The role of including failures is explained below.

The *truncated success rule* is defined for sequential testing and any N. It specifies to select each action equally likely in the first test and then to test the two actions alternatingly until testing is stopped using both the stopping and recommendation rule of Hoel (1972). Let S_j be the number of successes and F_j be the number of failures observed in the previous tests of action j. Then stop if $S_j + F_{3-j} \geq N/2$ for some $j \in \{1,2\}$ and recommend action j. Note that the recommendation is unique.

To gain some intuition why failures of the alternative action are incorporated, rewrite the stopping rule as $S_j \geq S_{3-j} + [N/2 - (S_{3-j} + F_{3-j})]$ and consider N even. This condition implies that action j will yield at least as many successes as the alternative action if testing is not stopped but instead both actions are tested $N/2$ times (the term in squared brackets is the number of remaining tests of the alternative action). The idea is that the truncated success rule stops early as sufficient information has been gathered.

It turns out that all three rules presented above generate identical recommendation probabilities. This follows almost directly from a result of Bechhofer and Kulkarni (1982, Th. 5.1, see also Kulkarni and Jennison, 1986, Th. 2.1). For the argument one needs to use the fact that both the adjusted and the truncated success rule are symmetric and do not test either action more than $N/2$ times. The difficulty of the proof is that knowing action j yields at least as many successes as the alternative action is not enough to prove that recommendation probabilities are identical. This

[10] In the literature on selection procedures, selection refers to what we call the recommendation, to test each action equally often is called vector-at-a-time sampling.

is because the two actions can be equally successful during testing in which case the simple success rule uses a particular tie breaking rule. Below we present an alternative proof of the identical performance.

4.2 Bounded Payoffs

Now consider richer outcome spaces that induce more than two different possible payoffs. The natural approach is to extend the simple success rule by recommending the action that yielded higher average payoffs during testing while maintaining the same tie-breaking rule. This defines the *empirical success rule* (Manski, 2004, Stoye, 2006).[11]

A more sophisticated approach first used by Schlag (2003) in this framework is to first randomly transform each payoff yielded during the testing phase into a binary valued outcome and then to apply a rule designed for binary valued payoffs. With the appropriate random transformation, the properties of interest that a rule has for binary valued payoffs will carry over to the more general setting with payoffs in $[0,1]$. This approach is useful as it is difficult to establish exact results when working directly with the general case where payoffs are in $[0,1]$.

We present the random transformation and start by explaining how a payoff $y_j \in [0,1]$ observed in some test is transformed. Transform $y_j \in [0,1]$ into payoff 1 with probability y_j and into payoff 0 with probability $1 - y_j$. Notice that this random transformation is mean preserving and that payoffs 0 and 1 are the fixed points. Perform this transformation independently transform for each payoff observed during testing. The result is a binary valued sample, it is as if only payoffs in $\{0,1\}$ were realized in each round of the test phase. This is called the *binomial transformation.*

With this transformation, rules defined for binary valued payoffs turn into rules defined for payoffs belonging to $[0,1]$. The rule that emerges when evaluating the simple success rule for even N and the adjusted success rule for odd N to the transformed sample will be called the *binomial average rule.* Thus the binomial average rule is a simultaneous sampling rule. To apply instead the truncated success rule to the transformed sample will be called the *truncated binomial average rule.*

The binomial transformation is mean preserving. When looking at the binary valued sample after the transformation it is as if the decision maker is facing a binary valued distribution with the same means as the true underlying distribution. This is why properties for binary valued payoffs that can be described in terms of means carry over to the more general setting with payoffs belonging to $[0,1]$ when using the binomial transformation.

Proposition 1 *If a rule attains minimax regret, is uniformly consistent or is ex-ante improving when $\mathcal{Y} = \{0,1\}^2$ then applying this rule to the binomially transformed*

[11] Note that Manski (2004) uses an alternative tie-breaking rule.

sample generates a rule that attains minimax regret, is uniformly consistent or is ex-ante improving, respectively, when $\mathcal{Y} = [0,1]^2$.

The part of the statement concerning minimax regret is proven in Schlag (2006c). The statement in terms of uniform consistency and ex-ante improvingness follows by the same argument and is left to the reader. All proofs are built on the simple observation that $u(f,P) = u(f,P^0)$ if $P^0 \in \Delta\{0,1\}^2$ is such that $P_j^0(1) = \int y_j dP_j(y_j)$ for $j = 1,2$ and f can be described as a rule that first binomially transforms the sample and then makes the recommendation based on the transformed sample.

4.3 Learning

To prepare for later results we present some learning properties of the rules described above. Add a superscript to recommendation probabilities to index the sample size.

Proposition 2 *(i) The binomial average rule and the truncated binomial average rule yield identical recommendation probabilities and are uniformly consistent. Both rules are ex-ante improving, specifically, $p_j^{2k-2}(\cdot,P) < p_j^{2k-1}(\cdot,P) = p_j^{2k}(\cdot,P)$ holds if $P \in \Delta\mathcal{Y}^2$, $u_j(P) > u_{3-j}(P)$ and $k \in \mathbb{N}$.*

(ii) The empirical success rule is uniformly consistent but not ex-ante improving.

We recently discovered that the statement and proof of the fact that ex-ante expected probability of choosing the best action increases in k for $N = 2k$ has previously been shown by Gupta and Hande (1992, Theorem 2.2).

Proof. Consider first part (i) except for the statement about uniform consistency. Given Proposition 1 it is enough to consider only binary valued payoffs. Let N be even.

We first show that the simple success rule, the adjusted success rule and the truncated success rule all yield the same recommendation probabilities. We do this in three simple steps. Say that a rule has property "$*$" if it makes the same recommendation as the simple success rule whenever the simple success rule recommends some action with probability one.

Step 1. We first show that all three rules have property "$*$". We only need to focus on the event that action j yields strictly more successes than the alternative action after N tests. The arguments at the end of Section 4.1 show that the adjusted and the truncated success rule make the same recommendation as the simple success rule, namely they recommend action j.

Step 2. Now we show that any rule that satisfies property "$*$" is a best response against any prior Q_P that puts equal weight on P and P^s where P^s is the distribution that is derived from P by interchanging the labels of the two actions. This is a slightly more general claim than Lemma 1 in Canner (1970) and follows immediately for instance when looking at the likelihood ratios. The important insight is that there is no condition on the best response behavior on how to recommend when both actions have been equally successful after $N/2$ tests of each.

Step 3. Finally we prove that any rule that is symmetric and that satisfies property "$*$" has the same recommendation probabilities as the simple success rule. A consequence of step 2 is that $u(f, Q_P) = u(f', Q_P)$ holds if both f and f' satisfy property "$*$".[12] The proof then follows from the fact that $u(f, Q_P) = u(f, P)$ if f is symmetric.

The fact that the adjusted success rule makes the same recommendations as the simple success rule implies that $p_j^{2k-1}(\cdot, P) = p_j^{2k}(\cdot, P)$.

Assume that $u_j(P) > u_{3-j}(P)$. An immediate consequence of step 2 and the fact that more information cannot make a rational decision maker worse off is that $p_j^n(\cdot, P) \leq p_j^{n+1}(\cdot, P)$. As it is possible that $S_1^{2k-2} > S_2^{2k-2}$ while $S_1^{2k} < S_2^{2k}$, where the superscript refers to the number of tests conducted, the decision maker is strictly better off when running two more tests. Thus, $p_j^{2k-2}(\cdot, P) < p_j^{2k}(\cdot, P)$.

The statements on uniform consistency in parts (i) and (ii) follow directly from Tschebyscheff's inequality using the fact that variances of P_1 and P_2 are bounded above by 1/4.

Finally we present an example showing that the empirical success rule is not ex-ante improving. Assume that outcomes are identified with utilities and set $P_1(1) = 1 - P_1(0) = \mu$ and $P_2(x) = 1$ where $0 < x < \mu < 1/2$. Then $p_1^0 = 1/2 > p_1^2 = \mu$ but the mean of action 1 is strictly larger than that of action 2. ■

In Table 1 we present the probability of recommending action 1 under the binomial average rule for $N \in \{0, 1, 3\}$.

5 Minimax Regret

We now investigate minimax regret. Previous results exist for binary valued payoffs when the sample is constrained to be balanced. The findings of Canner (1970) show that the simple success rule attains minimax regret. Stoye (2005, 2006) contains an alternative proof of this result using the saddle point condition and also contains a formula for deriving the value of minimax regret ((2) below for the case of binary valued payoffs).

We present two rules that attain minimax regret when payoffs belong to $[0, 1]$ and when the choice of which actions are tested is part of the design. Let $B(j, m, z) = \binom{m}{j} z^j (1-z)^{m-j}$ be the probability of drawing j successes among m independent samples of a Bernoulli distribution with success probability z where $j, m \in \mathbb{N}_0$ with

[12] The proof up to here can be replaced by citing a finding of Bechhofer and Kulkarni (1982, Th. 5.1, see also Kulkarni and Jennison, 1986, Th. 2.1) proven using combinatorics that shows that $u(*, Q_P)$ is identical for the three rules simple, adjusted and truncated success.

$0 \leq j \leq m$ and $z \in [0,1]$. Define $N_{even} \in \mathbb{N}$ such that $N_{even} = N$ if N is even and $N_{even} = N+1$ if N is odd.

Proposition 3 *Assume sequential testing. Both the binomial average rule and the truncated binomial average rule attain minimax regret. In particular, minimax regret can be attained by a rule - the binomial average rule - that requires only simultaneous testing. The value* r_N^* *of minimax regret is given by*[13]

$$r_N^* = \max_{d \in [0,1]} \left(d \cdot \sum_{k < N_{even}/2} B\left(k, N_{even} - 1, \frac{1}{2}(1+d)\right)\right). \qquad (2)$$

Proof. We first show that the binomial average rule, denoted by f^*, attains minimax regret. For binary valued distributions P, regret $r(f^*, P)$ is a continuous function of $\mu(P)$ where $\mu(P)$ is contained in the compact set $[0,1]^2$. Thus there exists $P^* \in \arg\max_{P \in \mathcal{P}^B : \mu_1(P) > \mu_2(P)} r(f^*, P)$. Following the proof of Proposition 2 or (Canner, 1970, Lemma 1), f^* is a best response to Q_{P^*}. Consequently, (f^*, Q_{P^*}) satisfies (1) which implies that the binomial average rule f^* attains minimax regret.

While (2) can be derived explicitly (see Stoye, 2006 or Schlag, 2006a) we derive (2) here using existing results on selection procedures. Given $d \in [0,1]$ let $\mathcal{P}_d$ be the set of all binary valued distributions P such that $\mu_1(P) = \mu_2(P) + d$. Let P^d the particular one that has support $\{(1,0),(0,1)\}$ where $P^d(1,0) = \frac{1}{2}(1+d)$ and $P^d(0,1) = \frac{1}{2}(1-d)$. Hoel (1972) suggested a rule f_H for binary valued payoffs and derived that $P^d \in \arg\max_{P \in \mathcal{P}_d} p_2(f_H, P)$ (Hoel, 1972, p. 149, first paragraph). Bechhofer and Kulkarni (1982) present a class of rules that generate the identical recommendation probabilities as the simple success rule when N is even and payoffs are binary valued. Kulkarni and Jennison (1986, Remark 2.2) mention that the rule f_H of Hoel (1972) belongs to the class of Bechhofer and Kulkarni (1982), hence that $p_2(f_H, P) = p_2(f^*, P)$ holds for all $P \in \mathcal{P}^B$. Combining these results yields that $P^d \in \arg\max_{P \in \mathcal{P}_d} p_2(f^*, P)$. We now use this to derive r_N^* as

$$\begin{aligned} r_N^* &= \max_{P \in \mathcal{P}^B} r(f^*, P) = \max_{P \in \mathcal{P}^B} r(f^*, P) = \max_{d \in [0,1]} \max_{P \in \mathcal{P}^d} r(f^*, P) \\ &= \max_{d \in [0,1]} \max_{P \in \mathcal{P}^d} \{d \cdot p_2(f^*, P) : u_1(P) > u_2(P)\} = \max_{d \in [0,1]} \left(d \cdot p_2\left(f^*, P^d\right)\right) \end{aligned}$$

where $d \cdot p_2\left(f^*, P^d\right) = d \cdot \sum_{k < N_{even}/2} B\left(k, N_{even} - 1, \frac{1}{2}(1+d)\right)$. This completes the proof. ■

[13] For numerical evaluation of the specific value one best uses the first order conditions which are

$$\sum_{n=0}^{N_{even}/2-1} \left(\binom{N_{even}-1}{n} (1+d)^n (1-d)^{N_{even}-1-n} \right) = d \cdot (N_{even}-1) \binom{N_{even}-2}{\frac{N_{even}}{2}-1} \left(1-d^2\right)^{\frac{N_{even}}{2}-1}$$

The binomial average rule has been constructed in way that the design of the next test does not depend on the total number of tests remaining. This immediately yields the following corollary.

Corollary 1 *The binomial average rule attains minimax regret if the sample size N is a priori not known to the decision maker who only finds out that no more tests can be run once the N-th test has been conducted.*

5.1 The Role of Sample Size

We present some properties of the value of minimax regret r_N^*.

Proposition 4 $r_{2k+1}^* < r_{2k}^* = r_{2k-1}^*$ *for* $k \in \mathbb{N}$ *with* $r_N^* \to 0$ *as* $N \to \infty$ *such that*

$$\lim_{N\to\infty}\left(\sqrt{N}\cdot r_N^*\right) = \sqrt{\frac{1}{2\pi}}z^2 e^{-z^2/2}$$

where z is the unique solution to

$$\int_{-\infty}^{-z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx = \sqrt{\frac{1}{2\pi}}ze^{-z^2/2}\ .$$

Proof. The statements on the change in the value of minimax regret as the sample increases follow directly from Proposition 2.

Consider the binary valued distribution P^d from the proof of Proposition 2. Let N be even and let $\bar{y}_N$ be the average number of successes when N independent payoffs of action 1 are drawn from P^d. Then $p_2\left(f^*, P^d\right) = \Pr\left(\bar{y}_N < N/2\right)$. Using the central limit theory we obtain that

$$\Pr\left(\bar{y}_N < N/2\right) \approx_N \phi\left(\frac{\frac{1}{2}-u}{\sqrt{\frac{u(1-u)}{N}}}\right)$$

where ϕ is the cdf of the standardized normal distribution, $u = \frac{1}{2}(1+d)$ and "$\approx_N$" is used throughout this proof to identify formulae that have the same convergence rates as N tends to infinity. Let $g(u) := \sqrt{N}\left(u-\frac{1}{2}\right)/\sqrt{u(1-u)}$ and $h_N(u) := (2u-1)\phi(-g)$ then $r_N^* \approx_N \max_{u\in(1/2,1)} h_N(u)$.

We derive

$$\frac{d}{du}h_N = 2\phi(-g) + (2u-1)\frac{1}{\sqrt{2\pi}}e^{-w^2/2}\cdot g'(u) \qquad (3)$$

where

$$g'(u) = -\frac{1}{4}\frac{N+2u-1}{u(1-u)\sqrt{u(1-u)N}} \approx_N -\frac{1}{4}\frac{\sqrt{N}}{(u(1-u))^{3/2}}\ .$$

Let u_N^* be the maximizer of (2). Since $\lim_{N\to\infty} g'(u) = \infty$ we obtain from $\frac{d}{du}h_N(u_N^*) = 0$ and (3) that $\lim_{N\to\infty} u_N^* = 1/2$. Thus $g'(u_N^*) \approx_N -2\sqrt{N}$ and with the definition of g we obtain $2u_N^* - 1 \approx_N \frac{g}{\sqrt{N}}$. So

$$\frac{d}{du}h_N \approx_N 2\phi(-g) - g\sqrt{\frac{2}{\pi}}e^{-g^2/2} \text{ if } u = u_N^*.$$

Comparing this to the definition of z in the statement of the proposition we obtain that $\lim_{N\to\infty} g(u_N^*) = z$. So

$$r_N^* \approx_N (2u_N^* - 1)\phi(-z) \approx_N \frac{1}{\sqrt{N}}z^2\sqrt{\frac{1}{2\pi}}e^{-z^2/2} \ .$$

■

Table 2 contains some numerical values of r_N^* for small N where we include the first value of N for which minimax regret is below 5%, 4%, 3%, 2.5%, 2% and 1%. In particular we obtain the following result that motivated the title of this paper.

Corollary 2 *The value of minimax regret is below* 5% *when* $N = 11$ *but above* 5% *when* $N \leq 10$.

We verify numerically that $\sqrt{\frac{1}{2\pi}}z^2e^{-z^2/2} \approx 0.17$ up to four digits behind the comma. So for large N this means that $r_N^* \approx 0.17/\sqrt{N}$. As shown in Table 2, a slightly adjusted denominator yields a good approximation for $N \geq 3$:

$$r_N^* \approx \frac{0.17}{\sqrt{N_{even} - 0.2}} \ .$$

5.2 Deterministic Strategies

Manski (2004) uses the empirical success rule to investigate maximal regret and the role of covariate information. For the setting of this paper without covariate information, $e^{-\frac{1}{2}}/\sqrt{N} \approx 0.61/\sqrt{N}$ is presented as an upper bound on regret for even N under the empirical success rule (Manski, 2004, eq. 23). In particular, $N \geq 148$ is needed to ensure that regret is below 5%. For large N this means that the empirical success rule requires 12 times more observations than the binomial average rule to guarantee the same maximal regret. A tight upper bound for regret under the empirical success rule is not known.

Lower bounds on maximal regret of the empirical success rule are easily computed analytically for small N. These computations reveal why the empirical success, denoted by f^e, does not attain minimax regret. Consider for instance the distribution $\hat{P}$ such that $\hat{P}(1, z) = \lambda$ and $\hat{P}(0, z) = 1 - \lambda$ and focus on $N = 2$. Then $p_2^3\left(f^e, \hat{P}\right) = 1 - \lambda$. So if $\lambda > z > 0$ then $r\left(f^e, \hat{P}\right) = (\lambda - z)(1 - \lambda) \geq \frac{1}{4}$ while

$r_3^* = 0.087$. In particular, the empirical success rule does not attain minimax regret. The empirical success rule uses only the fact that $z > 0$ and does not account for how large z is. The binomial transformation on the other hand takes the size of z into account. Accordingly, z and 0 are treated identically with probability $1 - z$ as the observation $(0, z)$ is transformed into $(0, 0)$ with probability $1 - z$. The particular parametrized distribution $\hat{P}$ can be used to analytically show that the empirical success rule does not attain minimax regret when $N \leq 18$. For larger N we revert to numerical simulations and cite some results of Stoye (2006, Table 3). Maximal regret under the empirical success rule is at least equal to 0.0495, 0.0326 and 0.0193 (the value of minimax regret equals 0.0382, 0.0269 and 0.017) for values of N equal to 20, 40 and 100 respectively.

Gupta and Hande (1992) show how to derive an alternative deterministic strategy using a 'rounding' trick. We show that this yields a rule f^d that outperforms the empirical success rule when comparing the respective upper bounds on regret. Assume N even, test each action equally often and then recommend (with probability one) the action that is recommended most likely under the binomial average rule f^*, recommending action 1 under f^d whenever both actions are recommended equally likely under f^*.[14] More specifically, let $f^d(h) = f^*(h)$ for histories of length strictly smaller than N and let $\Pr\left(f^d(h) = 2\right) = 1$ if and only if $\Pr(f^*(h) = 2) > 0.5$ for any history h of length N. We call f^d the *rounded binomial average rule.*

Note first that the existence of a tie breaking rule precludes that f^d attains minimax regret. This is because Q_{P^*} defined in the proof of Proposition 3 only maximizes regret under f^d if both actions are chosen equally likely whenever data is binary valued and each action has yielded the same number of successes.

Gupta and Hande (1992) show that $\Pr\left(f^d(h) = j\right) \geq 2\Pr(f^*(h) = j) - 1$ when j is the best action and $\mu_1 \neq \mu_2$. Thus, $\Pr\left(f^d(h) \neq j\right) \leq 2\Pr(f^*(h) \neq j)$ and hence $r\left(f^d, P\right) \leq 2r(f^*, P)$. In particular this means that the rounded binomial average rule has regret bounded above by twice the value of minimax regret. Following Table 2, maximal regret of the binomial average rule is below 2.5% for the first time when $N = 47$ and hence $N = 47$ observations are necessary to ensure that the rounded binomial average rule has regret below 5%. For large N this means that this deterministic rule requires 4 times more observations than the binomial average rule and hence 1/3 of the observations of the empirical success rule.

For more on the use of the rounding trick see Section 5.7 and Schlag (2007).

5.3 Foregone Payoffs and Counterfactual Evidence

Consider briefly a decision maker that has more information to base the recommendation on. Assume after each test that the outcome of the action not chosen, the

[14] Notice that Gupta and Hande (1992) do not consider an exogenous tie breaking rule and hence their rule is formally not deterministic.

so-called *foregone payoff*, is also observed. It is as if *counterfactual evidence* is available. Formally the binomial average rule is defined as above, only incorporating payoffs of those actions tested.

Our proofs above reveal that counterfactual evidence cannot be used to reduce maximal regret.

Corollary 3 *The value of minimax regret when counterfactual evidence is available is equal to the value of minimax regret derived in Proposition 3 for the case where only outcomes of actions tested are observed.*

Proof. In the proof of Proposition 3 we find that there is a least favorable prior that puts weight only on distributions that have support in $\{(1,0),(0,1)\}$. When faced with such a least favorable prior, it is as if the decision maker observes in each round of the test phase the payoff that the action not tested would have achieved. There is no advantage to receiving explicit information about foregone payoffs when facing such a prior. Thus the saddle point derived without foregone payoffs remains a saddle point with foregone payoff information. Consequently, any strategy that attains minimax regret without foregone payoffs attains minimax regret with foregone payoffs. Moreover, the value of minimax regret remains unchanged when foregone payoffs are included. ■

The above finding builds on the fact that foregone payoffs are not needed when facing a distribution in the support of the least favorable prior. On the other hand, it is intuitive that counterfactual evidence can be used to reduce maximal regret when facing some other distributions that are not least favorable. A deeper analysis of the setting with foregone payoffs is outside the scope of this paper, for initial results see Schlag (2006b).

5.4 Costly Testing

Return to the original setup but now assume that testing is costly where we follow Canner (1970). There are two types of cost, an explicit cost c of a test and the implicit cost on a subject pool when testing the inferior action. Assume that there are $M \in \mathbb{N}$ subjects that have to be treated where each test costs $c \geq 0$. The utility of the decision maker of recommending action j after testing action i in N_i tests with $N = N_1 + N_2$ facing distribution P is given by $N_1 u_1(P) + N_2 u_2(P) + (M-N) u_j(P) - Nc$. Regret $\hat{r}(f,P)$ is then given by

$$\begin{aligned}\hat{r}(f,P) &= M \max\{u_1(P), u_2(P)\} - [(M-N) p_1(f,P) + N_1] u_1(P) \\ &\quad - [(M-N) p_2(f,P) + N_2] u_2(P) + Nc \\ &= \max\left\{ \begin{array}{l} (u_1(P) - u_2(P)) [(M-N) p_2(f,P) + N_2], \\ (u_2(P) - u_1(P)) [(M-N) p_1(f,P) + N_1] \end{array} \right\} + Nc\ .\end{aligned}$$

Note that the term Nc appears as no tests have to be run if the true underlying distribution is known.

Canner (1970) proves for distributions that generate binary valued payoffs that the simple success rule attains minimax regret among the rules that test each action equally often (so N is even). Numerical simulations are used to conjecture a property of a least favorable distribution: support is in $\{(1,0),(0,1)\}$. We solve for minimax regret among all rules based on simultaneous testing that determine the total number of tests deterministically. A least favorable distribution is derived, showing that the property of a distribution conjectured by Canner (1970) holds for a much more general setting.

Proposition 5 *Consider simultaneous testing. There exists an odd number N^* such that the binomial average rule based on sample size N^* attains minimax regret among the rules that choose the total number of tests deterministically. There is a least favorable distribution of this rule that has support in* $\{(1,0),(0,1)\}$.

As the support of a least favorable distribution is known, the value of minimax regret can be derived as in the case with costless testing (see (2)).

Our proof follows the arguments used in the proof of Proposition 3.

Proof. We could extend the result of Canner (1970) to the case where payoffs are in $[0,1]$ by using a binomial transformation as in Proposition 1. However this approach would not yield formal results on the least favorable distribution. Hence we expand instead on the line of reasoning established in the proof of Proposition 3.

We first restrict attention to rules that run a given number of N tests. Assume N even. Following Hoel (1972) and Bechhofer and Kulkarni (1982) there exists d^* such that P^{d^*} is a least favorable distribution under the binomial average rule. In order to prove that $(f^*, Q_{P^{d^*}})$ is a saddle point we have to show that the binomial average rule f^* is a best response against $Q_{P^{d^*}}$. Given Proposition 3 we only have to deal with the choice of N_1 and N_2. The associated terms equal $\frac{1}{2}N_1d^* + \frac{1}{2}N_2d^* = \frac{1}{2}Nd^*$ for all N_1 and N_2 with $N_1 + N_2 = N$ and as N is fixed the statement is proven. The case of N odd then also follows.

As we restrict attention to rules that choose the number of tests deterministically the solution follows by finding N^* that minimizes the maximal regret of the binomial average rule f^* across all N. Index regret by the sample size and let N be even. Then

$$\begin{aligned}\hat{r}_N\left(f^*, P^d\right) &= d\cdot\left[(M-N)\,p_2(f,P) + N/2\right] + Nc \\ \hat{r}_{N-1}\left(f^*, P^d\right) &= d\cdot\left[(M-N-1)\,p_2(f,P) + (N-1)/2\right] + (N-1)\,c\end{aligned}$$

so

$$\hat{r}_N\left(f^*, P^d\right) - \hat{r}_{N-1}\left(f^*, P^d\right) = d\cdot\left[1/2 - p_2(f,P)\right] + c > 0\ .$$

Hence N^* is odd. ■

We solve minimax regret by restricting attention to rules that assign the total number of tests deterministically. The description of minimax regret behavior when N can be chosen randomly is more intricate and is not presented here. One can show that the decision maker will randomize over odd values of N and then implement the binomial average rule for the selected sample size.

An analysis of sequential costly testing is too complex and hence outside the scope of this paper. Of course regret under the truncated binomial average rule will be strictly lower than that under the binomial average rule.

5.5 Learning which action is better

In this section we review the connection between attaining minimax regret and recommending the better action. We recently found out that the case of N even has already been solved in Gupta and Hande (1992) who refer for $N \leq 20$ to tables in Sobel and Huyett (1957). In the introduction we demonstrated why regret is equal to the value of not recommending the better action weighted by how much better the better action is. Consider briefly a decision maker that only cares about recommending the better action. So we change the objectives and maintain the rest of the basic setting with costless testing. Following the literature on selection procedures, e.g. Sobel and Huyett (1957), one needs to introduce a so-called indifference zone in order to get sensible results.[15] We assume that the decision maker only cares about the recommendation when $|EY_1 - EY_2| \geq d$ where $d \in (0,1)$ is given and that the decision maker aims to maximize the minimal probability of correct selection. $p_i(f, P)$ is called the *probability of correct selection* if $EY_i > EY_{3-i}$. Formally the objective is to solve $\max_f \min \{p_i(f, P) : EY_i - EY_{3-i} \geq d\}$. Notice that it is as if we are testing the null hypothesis that $EY_1 \leq EY_2 - d$ against the alternative hypothesis that $EY_1 \geq EY_2 + d$ with the constraint that the type I error equals the type II error.

Let R be the *risk function* (Wald, 1950) defined by $R(f, P) = 0$ if $|EY_1 - EY_2| < d$ and $R(f, P) = p_j(f, P)$ if $EY_j \leq EY_{3-j} - d$. So we will search for a rule that attains minimax risk in the sense that it solves $\min_f \max_P R(f, P)$. No new proofs are needed. The proof of Proposition 3 extends immediately, the binomial average rule attains minimax risk. That proof also reveals that the value of minimax risk is given by

$$\sum_{k<N_{even}/2} B\left(k, N_{even} - 1, \frac{1}{2}(1+d)\right) . \tag{4}$$

The goal to attain minimax regret is compatible with the goal of maximizing the minimal probability of correct selection for any $d \in (0,1)$.

[15] Without this restriction the minimal probability of correct selection is equal to 1/2 regardless of how large N is.

Given the above we can gain some insight on how difficult or easy it is to find the better action and thus on the performance of the binomial average rule. Using (4) we present values of N and d that yields a value of minimax risk equal to 5% in Table 3. In particular, we find that $N \geq 267$ is needed in order for minimax risk to fall below 5% when $d = 0.1$.

5.6 Three or More Actions

A natural extension of the basic setting is to consider a decision involving more than two actions. Assume that there are $I \geq 3$ actions. We simplify our presentation by restricting attention to a sample size N that is a multiple of I. The binomial average rule is hence easily extended. We leave formalities to the reader and directly move to the findings. Note that there is no previous characterization of minimax regret in any setting with more than two actions, even for the more limited case where payoffs are binary valued. However, Gupta and Hande (1992) did show that the behavior as under the binomial average rule will maximize the minimum probability of correct selection whenever there is a unique best action.

Proposition 6 *The binomial average rule attains minimax regret among all rules that test each action N/I times.*

Proof. The key to the proof is that one can assume that the I random variables associated to each of the actions are independent. The rest is a simple adaptation of the proof of Proposition 3. Let $P^* \in \Delta \{0,1\}^I$ maximize regret of the binomial average rule and have the property that the marginals are independent. Let Q^* put equal weight on the $I!$ copies of P^* that emerge when permuting the labels of the I actions. The binomial average rule f^* clearly is a best response against Q^* as the distributions associated to each action are independent and hence only the successes matter. Thus (f^*, Q^*) is a saddle point and thus f^* attains minimax regret. ■

Note that the above statement is much more restricted than the analogous one in Proposition 3 for the case of two actions. (i) The value of minimax regret is unknown as we do not know enough about a least favorable distribution of the binomial average rule when $N \geq 3$.[16] All we know is that there is a least favorable distribution that has binary valued payoffs. (ii) The statement in Proposition 6 is not true if we allow for all simultaneous testing rules. In the appendix we verify for $I = N = 3$ that minimax regret cannot be attained by almost surely testing each action equally often.

5.7 Other Extensions

We round up our investigation of minimax regret by mentioning some additional extensions that are of interest but where the analysis is outside the scope of the

[16] We refer to Schlag (2006a) for an upper bound on minimax regret.

present paper. Each of these extensions refers to a variation of the main setting of this paper in which the decision maker knows more about the underlying environment.

1. Assume that the set of possible outcomes is action dependent (cf. Puppe and Schlag, 2006). For instance it could be that treatment one is more expensive than treatment two. It is clear from the insights gained in Schlag (2006c) that the problem can be simplified by randomly transforming each outcome into either the most (a success) or the least (a failure) preferred outcome of the respective action. As a success or a failure will now typically depend on the action it no longer makes sense to compare the average number of successes of each action. In particular, we do not expect that the binomial average rule attains minimax regret. This is easily verified in simple settings such as when $N = 1$. Further analysis is needed. Of course one can obtain an upper bound on minimax regret by applying the present results to the union of the two action dependent outcome spaces.

2. For the next scenario assume that there is some information about the set of possible underlying means. For example one may be concerned in predicting one of two events. Payoffs are either 1 if prediction is correct or 0 if incorrect. Then one knows that $\mu_1(P) + \mu_2(P) \geq 1$ where the inequality is due to the fact that the two events are not known to be disjoint. Such restrictions on the means, as long as these do not depend on how actions are labelled, can be dealt with the present framework. Given this symmetry the proof of Proposition 3 can be easily extended. Consequently the binomial average rule still attains minimax regret. Of course the value of minimax regret can be different. In our particular example this is however not the case as the support of the least favorable prior satisfies this restriction.

3. Assume that there is only one action with unknown mean. Manski (2005) presents a rule that attains minimax regret when payoffs are binary valued. The insights of Schlag (2006c) and Proposition 1 show how to generalize this rule to one that attains minimax regret for general payoffs by first applying the binomial transformation. One might also imagine a setting in which there are $I - 1$ actions with unknown means and one action with known mean. The mean of the one action could be known because it is associated to some outside option. Again the binomial transformation allows to focus on binary value payoffs. However here minimax regret behavior is yet not known for the case binary value payoffs.

4. Consider the case where each test yields additional information about a covariate (such as the sex of a patient) where recommendations can be conditioned on the covariate values. Foundations for gaining an understanding for how to attain

minimax regret under covariate information have been established by Stoye (2006). Results derived by Stoye (2005) prior to the first version of this paper show how to generally deal with covariate information. However an explicit value of minimax regret was not available for the case analyzed by Manski (2004) where payoffs are constrained to $[0,1]$ prior to this paper. Adding the insights from Propositions 3 and 4 it is easy to see for instance that $11 * |X|$ observations are needed to ensure that regret is below 5% where $|X|$ is the finite number of different values of the covariate. The first results in the literature on minimax regret behavior under endogenous sampling and covariate information were contained in an earlier version of this paper (Schlag, 2006a) but are left out of the present version due to space constraints.

Note that whenever there are two actions then the rounding trick of Gupta and Hande (1992) explained in Section 5.2 shows how to create a deterministic strategy from a randomized minimax regret strategy which at most doubles the value of maximal regret. In particular this can be used for all settings described above including the cases where there is one unknown action and where there is information on covariates.

6 Maximin

There is an alternative popular method in some fields for selecting choices without priors: maximin. We briefly demonstrate why this alternative does not make sense in our setting.

According to maximin, the performance of a strategy is measured by the minimal payoff it achieves among all feasible environments, and then the strategy that achieves the largest minimum is selected. This criterion was introduced by Wald (1950) and was first axiomatized by Milnor (1954). The recent axiomatization of Stoye (2006) nicely illustrates how the foundations of maximin compare to those of subjective expected utility maximization and minimax regret. Formally, $\hat{f}$ attains *maximin* if

$$\hat{f} \in \arg\max_{f} \inf_{P} u\left(f, P\right)$$

where u is derived as under minimax regret.

It is straightforward to show in our setting that any strategy attains maximin. Minimal payoffs for any strategy f are generated by the distribution P that satisfies $u_j\left(P\right) = 0$ for all j. Consequently, all strategies are equally good in terms of their minimal payoff. In particular, the strategy that only tests action 1 in the test phase and then recommends action 1 regardless of the outcomes in the test phase attains maximin. A similar conclusion was derived by Manski (2005) for the case of a single unknown treatment.

Notice that the results using the maximin criterion are even more distinct than those derived for minimax regret when one assumes costly testing as in Section 5.4 with $c > 0$. In that setting, any rule that attains maximin does not run any tests.

7 Conclusion

In this paper we limited attention to exact results and demonstrated the power of game theory by deriving the first analytic results on sequential design of randomized experiments. Exposition and notation was detailed to bridge gaps between disciplines as the material is related to research in biostatistics, artificial intelligence and bounded rational learning. Proofs combine insights from the literature on selection procedures with the tools of game theory.

The perhaps astonishing result that only 11 tests are needed for 5% is due to the fact that estimates are measured in terms of *value* or performance and not in terms of *choice* itself. While in classical hypothesis testing the decision maker is worried about making the *wrong* choice, here he or she is worried about making a *bad* choice. Of course the choice itself can be important for inference and seems to be the only accepted practice when testing new medications. However there has been recently a debate on the ethics of running randomized experiments, driven by concern of "using subjects as guinea pigs for the good of man kind" (e.g. see Weijer et al., 2000). The fact that only 11 tests are needed to gather useful evidence about treatment response dramatically limits the possible damage as only few patients are treated with the worse medication. Thus, concern for value could not only be insightful per se but could also help dampen ethical concerns relating to randomized clinical trials.

A feature of the binomial average rule is that it is typically randomized. In fact, when the underlying distribution has a continuous density, due to the properties of the binomial transformation, the recommendation will never be deterministic. On the other hand, the recommendation of the rounded binomial average rule is always deterministic at the cost of not attaining minimax regret. Maximal regret is at most doubled which means that it requires around at most four times the observations. More research is needed to investigate how much maximal regret is really increased. It turns out that less randomness is needed to attain minimax regret. Initial insights can be found in Eozenou et al. (2006) who present an alternative rule with less variance and verify numerically for $N \leq 86$ that it also attains minimax regret.

We make two brief comments relating to the common misunderstanding of randomized recommendations, illustrating these by assuming that the recommendation is 96% on a new program and 4% on the old program. a) A randomized recommendation can be useful as it can be implemented as a heterogeneous recommendation, here the new program has to be implemented in 96% of all sites. b) If the program should only be implemented on a single site, the fact that the recommendation is randomized does not introduce time inconsistency. While the old program will only be implemented with a low probability, if this event should occur then the decision maker should find no difficulty in accepting this realization. The fact that the new program is recommended with very high probability does not mean that it is better. We do acknowledge the value of a deterministic recommendation for some applications. When faced with this additional constraint one has two possible options. i) Use the rounded binomial average rule where this paper gives insights on how much regret is possibly increased over the value of minimax regret. ii) In our example above, recommend the new program almost surely, this increases maximal regret by at most 4% (see Eozenou et al., 2006). Note that (ii) is preferred to (i) if and only if the value of minimax regret is at most 4% which means that $N \leq 18$.

The focus of the paper was not pure statistical decision theory as "learning" was also investigated, both in terms of design of testing and in terms of performance over time. A common objection to worst case analysis is that there is no concern for good performance in situations that are not worst case. It was therefore important that we spent some effort on describing properties of the binomial average rule when facing general distributions. We were able to verify that the binomial average rule causes ex-ante expected payoffs to increase with the number of tests, putting arbitrarily large weight on a best treatment when the sample is sufficiently large.

Last but not least, we hope that there will be more research on exact nonparametric hypothesis testing and decision making. The close connection between hypothesis testing and statistical decision making is manifested in Section 5.5 where we find that the binomial average rule both attains minimax regret and maximizes the probability of correct selection.

References

[1] Anscombe, F.J. and R.J. Aumann (1963), A Definition of Subjective Probability, *Ann. Math. Stat.* **34**, 199-205.

[2] Bechhofer, R.E. and R.V. Kulkarni (1982), "Closed Adaptive Sequential Procedures for Selecting the Best of $k \geq 2$ Bernoulli Populations," In *Proceeding of the Third Purdue Symposium on Statistical Decision Theory and Related Topics* (S.S. Gupta and G. Berder, eds.), 61-108, Academic Press, New York.

[3] Bergemann, D. and K.H. Schlag (2006), *Robust Monopoly Pricing - The Case of Regret*, Unpublished Manuscript, European University Institute.

[4] Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis* (2nd Ed.), Berlin, New York: Springer Verlag.

[5] Berry, D.A. and B. Fristedt (1985), *Bandit Problems: Sequential Allocation of Experiments*, Chapman-Hall, London.

[6] Börgers, T., Morales, A.J., and R. Sarin (2004), "Expedient and Monotone Learning Rules," *Econometrica* **72** (2), 383-405.

[7] Canner, P.L. (1970), "Selecting one of Two Treatments when the Responses are Dichotomous," *J. Amer. Stat. Asoc.* **65(329)**, 293-306.

[8] Cozzi. G. and P. Giordani (2006), "Do Sunspots Matter under Complete Ignorance", *Res. Econ.* **60**, 148-154.

[9] Droge, B. (1998), "Minimax Regret Analysis of Orthogonal Series Regression Estimation: Selection versus Shrinkage," *Biometrika* **85**, 631–643.

[10] Eldar, Y., Ben-Tal, A. and A. Nemirovski (2004), "Linear Minimax Regret Estimation of Deterministic Parameters with Bounded Data Uncertainties," IEEE Trans. Sign. Proc. 52 (8), 2177-2188.

[11] Eozenou, P., J. Rivas and K.H. Schlag (2006), *Minimax Regret in Practice - Four Examples on Treatment Choice*, Unpublished Manuscript, European University Institute.

[12] Freedman B. (1987), "Equipoise and the Ethics of Clinical Research," *N. Engl. J. Med.* **317**, 141-145.

[13] Gupta, S. S. and S. N. Hande (1992), "On Some Nonparametric Selection Procedures," *Nonparametric Statistics and Related Topics*, A.K.Md.E. Saleh (Editor), Amsterdam: Elsevier, 33-49.

[14] Hayashi, T. (2006), "Regret Aversion and Opportunity-Dependence," Unpublished Manuscript, University of Texas.

[15] Hodges, J.L.Jr. and E.L. Lehmann (1950), "Some Problems in Minimax Point Estimation," *Ann. Math. Stat.* 21(2), 182-197.

[16] Hoel, D.G. (1972), "An Inverse Stopping Rule for Play-the-Winner Sampling," *JASA* **67**(337), 148-151.

[17] Jennison, C. and B.W. Turnbull (2000), *Group Sequential Tests with Applications to Clinical Trials*, New York: Chapman and Hall.

[18] Lai, T.L. and H. Robbins (1985), "Asymptotically Efficient Adaptive Allocation Rules," *Adv. Appl. Math.* **6**, 4-22.

[19] Lakshimivarahan, S. and M.A.L. Thathachar (1973), "Absolutely Expedient Learning Algorithms for Stochastic Automata," *IEEE Trans. Syst., Man and Cybern.*, SMC-3, 281-286.

[20] Linhart, P.B. and R. Radner (1989), "Minimax-Regret Strategies for Bargaining over Several Variables," *J. Econ. Theory* **48**, 152-178.

[21] Manski, C. (2004). "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica* **72**(**4**), 1221-1246.

[22] Manski, C. (2005), *Social Choice with Partial Knowledge of Treatment Response.* Princeton, Oxford: Princeton University Press.

[23] Milnor, J. (1954), Games Against Nature. In Decision Processes, ed. R.M. Thrall, C.H. Coombs & R.L. Davis. New York: John Wiley & Sons.

[24] Puppe, C. and K.H. Schlag (2006), *Choice under Complete Uncertainty when Outcome Spaces are State-Dependent*, Unpublished Manuscript, European University Institute.

[25] Savage, L.J. (1951), "The Theory of Statistical Decision," *J. Amer. Stat. Assoc.* **46(253)**, 55-67.

[26] Savage, L.J. (1954), *The Foundations of Statistics*, New York: John Wiley & Sons..

[27] Schlag, K.H. (1998), "Why Imitate, and if so, How? A Boundedly Rational Approach to Multi-Armed Bandits," *J. Econ. Theory* **78(1)**, 130-156.

[28] Schlag, K.H. (2002), *How to Choose – A Boundedly Rational Approach to Repeated Decision Making*, Unpublished Manuscript, European University Institute.

[29] Schlag, K.H. (2003), *How to Minimize Maximum Regret under Repeated Decision-Making*, Unpublished Manuscript, European University Institute.

[30] Schlag, K. H. (2006a), *Eleven - Tests needed for a Recommendation*, European University Institute Working Paper ECO **2006-2**, January 17.

[31] Schlag, K. H. (2006b), *Distribution-Free Learning*, Working Paper ECO 2007/1**,** European University Institute, Economics Department.

[32] Schlag, K.H. (2007), *How to Attain Minimax Risk with Applications to Distribution-Free Nonparametric Estimation and Testing*, Mimeo, European University Institute.

[33] Sobel, M. and M.J. Huyett (1957), "Selecting the One Best of Several Binomial Populations," *Bell Sys. Tech. J.* **36**, 537–576.

[34] Stoye, J. (2005), *Minimax Regret Treatment Choice with Finite Samples*, Unpublished Manuscript, New York University, November 15.[17]

[35] Stoye, J. (2006), *Minimax Regret Treatment Choice with Finite Samples*, Unpublished Manuscript, New York University.

[36] von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton: Princeton Univ. Press.

[37] Wald, A. (1947), *Sequential Analysis*, New York: John Wiley & Sons.

[38] Wald, A. (1950), *Statistical decision functions*, New York: John Wiley & Sons.

[39] Weijer, C. Shapiro, S.H, Cranley Glass, K and M.W. Enkin (2000), "Clinical Equipoise and not the Uncertainty Principle is the Moral Underpinning of the Randomised Controlled Trial (For and Against)," *BMJ* **321**, 756-758.

[17] see http://www.iue.it/Personal/Schlag/papers/StoyeTreatmentchoice.pdf for this version, the last before (Schlag, 2006a) was put online.

A Minimax Regret when $I = N = 3$

In the following we investigate minimax regret behavior when there are three actions and $N = 3$. Let $\hat{f}$ be the binomial average rule. Then

$$\begin{aligned} p_1\left(\hat{f}, P\right) &= u_1\left(1-u_2\right)\left(1-u_3\right)+\frac{1}{2}u_1\left(u_2\left(1-u_3\right)+u_3\left(1-u_2\right)\right) \\ &\quad +\frac{1}{3}\left(u_1u_2u_3+\left(1-u_1\right)\left(1-u_2\right)\left(1-u_3\right)\right) \end{aligned}$$

and

$$r\left(\hat{f}, P\right) = u_1 - \left(p_1u_1 + p_2u_2 + p_3u_3\right) \text{ if } u_1 = \max\left\{u_1, u_2, u_3\right\}.$$

Given $i \in \{1,2,3\}$, let $P^i \in \Delta\{0,1\}^3$ be such that $P^i\left((1,1,1)\right) = 1 - P^i\left(e_i\right) = \frac{4}{3} - \frac{1}{3}\sqrt{7}$ (≈ 0.45), where e_i is the i-th unit vector, $i \in \{1,2,3\}$. So $\mu_i\left(P^i\right) = 1 > \mu_j\left(P^i\right) = \mu_k\left(P^i\right)$ when $\left|\{i,j,k\}\right| = 3$. It is easily verified that $\arg\max r\left(\hat{f}, P\right) \cap \Delta\{0,1\}^3 = \left\{P^1, P^2, P^3\right\}$, in particular P^i is a least favorable distribution of the binomial average rule for each $i = 1,2,3$.

Assume simultaneous testing and that $\hat{f}$ attains minimax regret. It is easily shown that there exists some prior $\hat{Q}$ such that $\left(\hat{f}, \hat{Q}\right)$ is a saddle point. This is because the fictitious zero sum game has an equilibrium if we restrict attention to distributions with binary payoffs which we may due to the properties of the binomial average rule. Given our calculations above on least favorable distributions, invoking standard arguments involving symmetry, it must be that $\left(\hat{f}, \hat{Q}\right)$ is a saddle point if $\bar{Q}$ is the prior where $\bar{Q}\left(P^i\right) = \frac{1}{3}$ for $i = 1,2,3$. In the following we show however that $\hat{f}$ is not a best response to $\bar{Q}$.

In our analysis we can focus on the probability q of guessing which action is best. Let $\alpha = P^i\left((1,1,1)\right)$. Note that if payoff 0 is observed after testing action i then it is known that i is not the best action. We easily verify for the binomial average rule that $q\left(\hat{f}\right) = \left(1-\alpha\right)^2 + 2\alpha\left(1-\alpha\right)\frac{1}{2} + \alpha^2\frac{1}{3} = 1 - \alpha + \frac{1}{3}\alpha^2$. Assume now that one action is tested twice. Let f^* be a best response recommendation where note that f^* recommends the action that yielded most success if only successes are observed. Then $q\left(f^*\right) = \frac{1}{3} + \frac{1}{3}\left(1-\alpha^2\right) + \frac{1}{3}\left(1-\alpha\right)\left(1-\alpha^2\right)$ where the first term reflects testing the best action twice, the second term represents the event of testing the best action only once and the third term is associated to not testing the best action at all. As $q\left(f^*\right) > q\left(\hat{f}\right)$ it follows that the binomial rule $\hat{f}$ is not a best response to $\bar{Q}$. We combine this finding with Proposition 6.

Remark 1 *Any rule that attains minimax regret under simultaneous testing when* $I = N = 3$ *does not often almost surely test each action once.*

B Tables

Table 1: Probability of recommending treatment 1 as function of number of tests.

N	0	1	3
p_1	$\frac{1}{2}$	$\frac{1}{2}+\frac{1}{2}\left(u_1-u_2\right)$	$\frac{1}{2}+\frac{1}{2}\left(2-u_1-u_2+2u_1u_2\right)\left(u_1-u_2\right)$

Table 2: Minimax regret and related values as function of number of tests.

N	0	1	3	5	7	9	11	13	15	17	19	25	29	33	39	47	73	99	199	289	∞
r^*_N	0.5	0.125	0.087	0.0706	0.0609	0.0543	0.0495	0.0458	0.0428	0.0403	0.0382	0.0335	0.0311	0.0292	0.0269	0.0246	0.0198	0.017	0.012	0.00998	0
$\frac{0.17}{\sqrt{N_{even}}-0.2}$	/	0.127	0.0872	0.0706	0.0609	0.0543	0.0495	0.0458	0.0428	0.0403	0.0382	0.0335	0.0311	0.0292	0.0269	0.0246	0.0198	0.017	0.012	0.00999	0

Table 3: Value of the the lower bound d of $|u_1(P) - u_2(P)|$ as a function of sample size N that ensures that the probability of recommending the better action is at least 0.05.

N	5	11	15	19	25	29	39	49	59	79	99	139	149	199	269
d	0.621	0.45	0.4	0.36	0.317	0.296	0.257	0.23	0.211	0.183	0.164	0.139	0.134	0.116	0.1