ELEVEN - Tests needed for a Recommendation[1] EUI Working Paper ECO 2006 - 2

Karl H. Schlag[2]

January 17, 2006

[1]The author would like to appologize. This paper was written in a rush to allow others to cite the results. Proofs yet have to be polished, examples recalculated, additional references added and existing references fine tuned. In particular, if any reader feels that his or her work was not appropiately credited then this was not the intention of the author who then wishes to be informed how to change the presentation or material herein. Any comments are gratefully received.

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

Abstract

A decision maker has to recommend a treatment, knows that any outcome will be in $[0, 1]$ but only has minimal information about the likelihood of outcomes (there is no prior). The decision maker can design a finite number of experiments in which treatments are tested.

For the case of two treatments we present a rule for designing experiments and making a recommendation that attains minimax regret and can thus ensure a given maximal error with the minimal number of tests. 11 tests are needed under the so-called *binomial average rule* to limit the error to 5%. We also consider the setting where there is covariate information to then identify minimax regret behavior and drastically reduce the number of tests needed to attain a given maximal error as compared to the literature (over 200 to 22 given two covariates). We extend the binomial average rule to more than two treatments and use it to derive a bound on minimax regret.

Keywords: statistical decision making, treatment response rule, binomial average rule.

1 Introduction

Consider the following simple statistical decision problem faced by a so-called decision maker. There are two actions each associated to a real random variable that yields payoffs in $[0, 1]$ with unknown mean. The decision maker wishes to choose the action that has the higher mean. Before making this choice the decision maker is allowed to gather information by running randomized experiments. In such a randomized experiment the decision maker selects one of the two actions and then observes a random outcome generated by the random variable associated to the action selected. These tests are run sequentially so that the design of future tests can be conditioned on previous outcomes. Of course this problem is simple but unrealistic if an unlimited number of tests can be made. So how many randomized experiments are needed to be able to make a choice that is guaranteed to be within 5% of the true maximum?[1] How should the tests then be designed? We solve this problem for any error. In the following we provide some intuition on how to achieve and how not to achieve this objective.

There are two underlying problems: (i) how should tests be conducted, in particular, is the decision of which action to test next dependent on payoffs observed earlier during experimentation and (ii) which action should be chosen based on the observations during the experiments.

If one decides to make an even number N of experiments and would not be allowed to observe realized payoffs until after all tests have been made then it is intuitive to select each variable equally often during experimentation. However, we assume that the payoff realized in each test is observed before the next test is implemented. So which action should one test next when half the tests have been conducted and one action always yielded payoff 1 while the other always payoff 0? We show that it is best to ignore the information about payoffs and to test each treatment equally often (conditional on N being even).

After running an equal number of experiments of each variable it is intuitive to use the *empirical success rule*, to choose the action that achieved the higher average payoffs during experimentation as an estimate. This turns out *not* to be the best thing to do when payoffs can take any value in $[0, 1]$. We do not know how many experiments the empirical success rule would need but we show that at least 14 are

[1] The decision maker has to ensure that the expected payoff generated by his choice given these unknown random variables is at least 95% of the highest mean of the two actions.

needed. The problem with the empirical success rule is that use of absolute differences in observed payoffs is not a good way to determine differences in means as the absolute outcomes only give limited information about the mean (a similar intuition drives the findings in Schlag (1998)). In fact, if one additionally restricts the problem to binomial random variables then we show that it is best to use the empirical success rule after all (conditional on N being even). Stoye (2005) has proven a partial result for this case of binomial random variables, namely conditional on running the same number of experiments on each action (so how to design testing is not an objective) and shows that only 12 rounds are needed.

The setting of binomial variables is very restrictive. We solve the case of general payoffs by finding a way to ignore variables that are not binary valued. This can be done with the help of *binomial averages*. Accordingly, first transform each payoff y obtained from an experiment into a binary payoff by assigning value 1 with probability y and value 0 otherwise. Then apply the empirical success rule to the transformed payoffs, selecting each variable equally likely as estimate if there is a tie. We call this rule the *binomial average rule* and extend it to odd samples. We then prove that this rule minimizes the maximal error, in particular 11 experiments are sufficient to guarantee the error to be at most 5%. In general, the error converges to zero at a rate equal to the square root of sample size. Half the error (so 2.5%) requires about four times as many experiments, the precise value equals 47. The fact that we find that there is no better way to conduct future tests by some complicated scheme based on past observations of payoffs stems from the fact that the binomial average rule acts in the worst case as if each treatment can be observed for each test. In particular we obtain that performance cannot be improved by allowing the decision maker to observe the outcomes of all treatments in each test.

As an aside, it follows directly from our analysis that these values of maximal error for given number of experiments are also tight upper bounds on the error that can be achieved by a rational (or Bayesian) decision maker who is endowed with a prior over the possible outcome distributions. So any rational decision maker can guarantee an error of at most 5% in 11 tests.

Assume now that the mean of one treatment is known so only the unknown (innovation) treatment is tested. Then our results above can be used to attain a bound on minimax regret. One can act as if one has $2N$ experiments of two unknown treatments, hence only 6 tests are needed for 5% error and 24 for 2.5%. Of course this bound is

too crude. The setting of only one unknown treatment has been solved by Manski (2004) with an explicit formula given by Stoye (2005) for the binomial setting. With the insights of this paper (using binomial averages) it follows that the value of the smallest maximal error under binomial payoffs is the same as under general payoffs. Combining this with values in Table 3.1 in Manski (2005) indicate that at most 3 tests are needed for 5%.

Consider next a more intricate problem that we present in a more specific context: each variable is associated to a treatment, an experiment is conducted by testing a treatment on a subject belonging to some large population. The new aspect is that information about *covariates* (or attributes) is available for each subject. The same treatment may yield different means for different covariates and treatments can be recommended based on the observed covariates. The decision maker is assumed to know the distribution of covariates in the population of subjects. During experimentation one may decide on which covariate to test a selected treatment on. New questions arise. Should all be treated the same or will different covariates be recommended different treatments when observed outcomes differ? Is it necessary to perform experiments on specific covariates (called *stratified random sampling*)? If so, then how does the frequency of a covariate influence the number of experiments? Here it becomes important whether future tests may depend on the outcomes of earlier tests (*sequential testing*) or not (*simultaneous testing*). Notice that while we show that sequential testing adds no value when there is no covariate information, it will when there are at least two covariates. Most of our findings concern *simultaneous testing*, the setting considered in Manski (2004).

Manski (2004) allows payoffs to be in $[0,1]$, derives upper and lower bounds and uses these to show that it is better to condition treatment choice on covariate information than to recommend the same treatment for all provided N is sufficiently large. We identify rules that minimize the maximal error which shows this to be true for all N. These rules first determine randomly how many tests to assign to each covariate and then apply the binomial average rule to each covariate separately. The table for two covariates in Manski (2004, Table 2) *seem to indicate* that 52 experiments are needed to obtain an error below 11% when one covariate is three times as frequent as the other.[2] This extrapolates via the upper bound on the convergence rate shown

[2] Of course the bounds in Manski (2004) are upper bounds and the tables were only presented to show how performance increases with sample size.

in Manski (2004) to requiring over 200 tests to get below 5%. However, when testing each covariate equally often and then applying the binomial average rule we limit the number of tests to at most 11 times the number of covariates regardless of how covariates are distributed. Of course, tests should be assigned differently depending on the distribution. For the case of large N and two covariates we use a general method of Stoye (2005) to demonstrate how to assign tests to each covariate as a function of its frequency. It turns out that the number of tests can be reduced by at most 10% when compared to testing proportionally to the frequencies.

An interesting side effect is that we find that the value of minimax regret is even lower under sequential testing provided N is not too small. This is shown without entering into more detail how minimax regret behavior would look under sequential testing.

We then consider the case of more than two treatments and return to simultaneous testing. We generalize the binomial average rule by making recommendations via pairwise comparisons. Whether or not this rule attains minimax regret is not known. Instead we derive an upper bound on the maximal regret of this rule and find the same rate of convergence $1/\sqrt{N}$ as in the case of two treatments. Of course, it is more difficult to learn with more treatments, choosing T instead of 2 treatments is like multiplying the number of experiments by $T^3/2$.

Formally speaking, the above problem is set in the framework of minimax regret going back to Savage's (1951) interpretation of Wald (1950) and first axiomatized by Milnor (1954). Minimax regret is derived by finding a saddle point of the zero sum game between the decision maker and nature. The trick to reduce general distributions to binomial distributions and thus complexity was first used in Schlag's (2003) analysis of repeated decision making. For a literature review on treatment response we refer the reader to Manski (2004, 2005) and Stoye (2005).

We would like to emphasize that while we use the term treatment it is only one of many examples. Basically the results in this paper apply whenever someone has the possibility to experiment and gain information via independent draws at no cost and then has to commit to some action. Examples are easily found in marketing, operations research, production planning, crime prevention and profiling, etc..

2 Setting

Following Manski (2004), consider a decision maker (or policy maker or planner) that has to recommend (or choose) some treatment (in decision theory also called arm or action) that belongs to a finite set of treatments $\{1, .., T\}$.

Choice (or implementation) of treatment $i \in T$ generates a random outcome, outcomes belong to a set of outcomes Y where Y contains at least two elements. Randomness is generated as follows. There is an unknown (joint) distribution (or *environment*) $P \in \Delta\left(Y^T\right)$ and the random outcome generated from choosing treatment i is drawn from the marginal of P with respect to the i-th component, denoted by P_i.[3] Depending on the specification of P, treatments can but need not generate independent outcomes. A treatment i is called *binary valued* if only one of two outcomes can occur under this treatment. The distribution P is called *binary valued* if all treatments are binary valued. For instance, all distributions are binary valued if $|Y| = 2$.

The analysis in this paper will also apply if attention is restricted to $|Y| = 2$. However, even if outcomes are only measured in terms of success or failure, the restriction to binomial P is only applicable if the value of a success and of a failure does not depend on which treatment was tested (e.g. treatment specific side effects would be ruled out).

Before making a recommendation the decision maker is allowed to run a given number of N independent tests (or randomized experiments or trials or samples). In each of N rounds the decision maker may choose a treatment and observe a random payoff generated by this treatment where payoffs are generated independently across rounds. This testing we also call the *test phase*, the choice thereafter also the *recommendation* or *final choice*. So a *strategy* (or *treatment rule*) of the decision maker consists of two parts: (i) which treatments to test in the test phase and (ii) which treatment to recommend based on the observations in the test phase. We consider two informational settings for the test phase. We say that N *sequential randomized experiments* are performed if earlier observations within the test phase are allowed to influence which treatments are tested in later rounds. If on the other hand the decision maker has to pre-commit to the number tests run with each treatment before starting the test phase we speak of N *simultaneous randomized experiments*.

[3] 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 $A \subset \Delta A$. Y^T denotes the set of all functions $\{1, .., T\} \to Y$.

We first formally describe strategies for running sequential randomized experiments.

After running $n \in \{1,..,N\}$ tests, the *history* h *of length* n is given by $h = ((t_1, y_1), (t_2, y_2), ..., (t_n, y_n))$ where t_k is the treatment chosen and y_k is the outcome generated in the k-th round of the test phase. So y_k is an outcome randomly drawn from the distribution P_{t_k}. The strategy σ of the decision maker for running sequential tests is to assign to each history of length n with $n < N$ the treatment to test next and after testing is over to decide based on the history of length N which treatment to recommend. We allow for the decision maker to randomize over treatments both during the test phase and when making the recommendation. Formally,[4]

$$\sigma : \cup_{n=0}^{N} (\{1,..,T\} \times Y)^n \to \Delta \{1,..,T\}$$

where σ is described as a behavioral strategy. If the recommendation is always deterministic, i.e. for any history h of length N we have $\sigma(h) \in \{1,..,T\}$, then the treatment rule is also referred to as a *singleton rule* (Manski, 2004). We say that σ is *deterministic* if $\sigma(h) \in \{1,..,T\}$ holds for any history of any length $n \in \{0,..,N\}$.

Now we analogously describe strategies for the more restricted setting of simultaneous randomized experiments. It is more restricted as the decision maker can ignore previous information during the test phase of a sequential randomized experiment and thus behave as if simultaneous randomized experiments are executed. If not mentioned otherwise we will be considering the setting of sequential randomized experiments.

An element $n \in \Delta_d N := \left\{ n \in \mathbb{N}_0^T \text{ s.t. } \sum_{i=1}^{T} n_i = N \right\}$ specifies that treatment i will be tested in n_i rounds, $\sigma(\emptyset)(n)$ denotes the probability of this assignment of treatments. The history h of observations generated by the set of observations during the test phase is as above except that it is now unordered, so $h = \{(t_1, y_1), (t_2, y_2), ..., (t_n, y_n)\}$. Together this means that

$$\begin{aligned} \sigma &: \emptyset \to \Delta(\Delta_d N) \\ \sigma &: \cup_{k=1}^{N} (\{1,..,T\} \times Y) \to \Delta \{1,..,T\}. \end{aligned}$$

Note that sequential sampling is often not implementable as for instance it might take time until outcomes are generated. However we will see that sometimes sequential sampling outperforms simultaneous sampling.

In a more detailed description of the problem that involves treatments per se one would also introduce an infinite population, members referred to as *subjects*, and assume that the decision maker has to recommend a treatment for each subject. In each

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

round of the testing phase, a subject would be drawn randomly and the treatment is applied in a more medical sense to this subject. Different subjects react perhaps differently to the same treatment which then naturally generates a random outcome of a test. As the population is infinite and the testing phase is finite, which treatments are run on which subjects during the testing phase does not influence the aggregate outcome for the population given the recommendation of the decision maker (unlike what would happen if the population were finite).

3 Risk and Uncertainty

We speak of a *risky* environment when P is known as compared to an *uncertain* environment when P is unknown. We first make some assumptions on preferences of the decision maker in risky environments and later extend these to uncertain environments.

Our decision maker has a complete strict preference ordering over the set of outcomes Y and has a most preferred outcome y_H and a least preferred outcome y_L. Preference satisfy the von Neumann Morgenstern (1945) axioms and hence there is a utility function $u : Y \to \mathbb{R}$ such that the decision maker would recommend the treatment that maximizes expected utility. Since u is uniquely determined up to an affine linear transformation let $u(y_L) = 0$ and $u(y_H) = 1$. Consequently, we can assume without loss of generality that elements of Y are payoffs, that $Y = [0, 1]$ and that the decision maker aims to maximize expected payoffs when P is known. Notice that this does not mean that the decision maker is risk neutral in risky environments as payoffs are measured in terms of utility and not in monetary value. If $Y = \{0, 1\}$ then we also refer to payoff 1 as a success and 0 as a failure, a binary valued P will then also be called *binomial*.

Let $\pi(i, P)$ denote the expected payoff generated by choosing treatment i so

$$\pi(i, P) = \int_{y \in [0,1]^T} y_i dP(y) = \int_0^1 y_i dP_i(y_i).$$

A decision maker who knows P will recommend a treatment belonging to $\arg\max_i \pi(i, P)$. Treatments in $\arg\max_i \pi(i, P)$ are called *best* (given P).

A decision maker who knows P does not have to test treatments as he or she already knows which treatment(s) are best. However, as mentioned above, in the main setting of the paper, P will not be known by the decision maker. Regardless of the

knowledge of the decision maker, P describes the true environment. Best treatments remain defined as above, the decision maker simply does not know which treatment is best when he or she does not know P. For later analysis we need to understand how each strategy performs as the decision maker might have some conjecture about which environments he or she may be facing (more on this later).

Let $p_i(\sigma, P)$ denote the probability of recommending treatment i given strategy σ when facing distribution P.[5] Let $\pi(\sigma, P)$ denote the *expected* payoff of the recommendation induced by using strategy σ when facing distribution P where expectations are calculated based on the distribution P ex-ante before entering the test phase, so

$$\pi(\sigma, P) = \sum_{i=1}^{T} p_i(\sigma, P)\, \pi(i, P).$$

Notice that payoffs realized during the testing phase do not directly influence the value of recommending treatment i that is given by $\pi(i, P)$. Tests only possibly influence the recommendation indirectly as which treatment is recommended can (and typically is expected to) influence the recommendation. Thus, in the end, the decision maker only cares about $\pi(\sigma, P)$.

4 On the Rationality of Taking Averages

Assume in the following that there are only two treatments, so $T = 2$.

Before we move to the analysis we add some informal discussion on a strategy that appears natural. Assume for this that N is even. As N is even, the obvious candidate for how to behave during the test phase is to test each treatment equally often (and hence $N/2$ times, no real reason to choose a specific order). The obvious candidate for the recommendation is to choose whichever treatment yielded the higher average payoffs during the testing phase and to choose each treatment with equal probability if both treatments yielded the same average payoff. Any rule that combines these two candidates will be called an *empirical success rule* (Manski, 2004), representatives of this class will be denoted by $\bar{\sigma}$.[6][7]

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

[6] There are many different empirical success rules according to the order of testing, e.g. first test treatment 1 $N/2$ times and then test treatment 2 $N/2$ times, e.g. alternate between treatments starting with a random treatment.

[7] The empirical success rule in Manski (2004) has a tie breaking rule when both treatments yield

In the following we briefly illustrate when this strategy may or may not be optimal from the perspective of a rational decision maker. According to Savage (1951), a rational (or Bayesian) decision maker is endowed with beliefs over the true P, beliefs are modelled by a prior Q over distributions P so $Q \in \Delta\left(\Delta\left(\left[0,1\right]^T\right)\right)$ and the decision maker chooses a *best response given* Q by selecting a strategy that solves $\max_\sigma \int \pi\left(\sigma, P\right) dQ\left(P\right)$.

Notice that it need not be rational to test each treatment equally often. Clearly, if beliefs are such that the decision maker is only unsure about payoffs generated by treatment 1 then it is optimal to test treatment 1 for N times. Even if each treatment is ex-ante believed to be equally likely to be the best treatment given the prior Q then it is simple to construct beliefs under which the treatment tested next non-trivially depends on the payoffs that have been realized in earlier rounds of the testing phase.

Notice furthermore that it need not be rational to recommend the treatment that yielded the higher average payoff even if each treatment was tested equally often. Consider for instance a decision maker who believes the following: one treatment yields payoff z for sure, the other is binomial with payoff 1 occurring with probability λ where λ and z are known and $1 > \lambda > z > 0$ and each treatment is equally likely the binomial one.[8] After a single test the decision maker knows which treatment is best. Assume never-the-less that the decision maker tests each treatment equally often as there is yet no harm to this testing. However to then use average payoffs to determine the recommendation specifies the worse treatment with probability at least $(1-\lambda)^{N/2} > 0$. This is clearly not rational as the decision maker knows which treatment is best after the testing phase. We will return to this particular example later.

Of course, the empirical success rule is the natural choice when N is large. Due to the law of large numbers it will select the best treatment with arbitrarily large probability provided N is sufficiently large; it yields a consistent estimator for the best treatment (a formal proof of this statement is given below). However the empirical success rule need not make "sense" when N is small where we of course first have to specify how to quantify "sensible" for a decision maker without a prior Q.

Of course the decision maker may have additional information about the treatments, he or she could be rational. In this case we show below that our paper provides on the side a tight upper bound on many tests a rational decision maker (endowed with a

same average payoff.

[8] Formally, we refer to a decision maker endowed with the prior Q such that $Q\left(P^1\right) = Q\left(P^2\right) = \frac{1}{2}$ where $P^1\left(z,1\right) = 1 - P^1\left(z,0\right) = \lambda$ and $P^2\left(1,z\right) = 1 - P^2\left(0,z\right) = \lambda$.

prior) needs for a given error.

5 The Binomial Average Rule for T=2

Next we introduce the strategy that we call the *binomial average rule* that will be selected later as the "best" strategy when $T = 2$ and hence we have to spend some time on explaining it in detail.

The binomial average rule will be described first for the case where N is an even number. Test as follows. In the first round, select equally likely one of the two treatments. Then continue to test the two treatments alternatingly until the test phase is over. Let t_k be the treatment tested in round k so $t_{k+2} = t_k$ for $1 \leq k \leq N-2$. Transform any payoff realized during the test phase that belongs to $(0,1)$ into a payoff in $\{0,1\}$ by the following randomization. When observing payoff $y_k \in (0,1)$ in the k-th round of the testing phase then with probability y_k act as if payoff $\tilde{y}_k = 1$ was observed and with probability $1-y_k$ act as if payoff $\tilde{y}_k = 0$ was observed in that round ($\tilde{y}_k \in \{0,1\}$ denotes the value of the transformed payoff where $\tilde{y}_k = y_k$ if $y_k \in \{0,1\}$). Given this transformation, it is as if only payoffs in $\{0,1\}$ were realized in each round of the test phase. After the test phase is over, recommend the treatment that yielded the higher average number of successes after payoffs were transformed, choose each treatment equally likely if both treatments generated the same average number of successes. Formally, treatment 1 is recommended with probability 1 if

$$\frac{2}{N}\sum_{k:t_k=1}\tilde{y}_k > \frac{2}{N}\sum_{k:t_k=2}\tilde{y}_k,$$

with probability $1/2$ if equality holds above and with probability 0 otherwise.

The binomial average rule specifically tests treatments in an alternating fashion starting with a random treatment. This is not important for the performance of the rule for given N. It only matters that each treatment is chosen equally often when N is even. However, the alternating character allows the binomial average rule also to have nice properties when N is unknown at the beginning of the test phase (see result below).

The *binomial average*

$$\frac{1}{\#\{k : t_k = i\}}\sum_{k:t_k=i}\tilde{y}_k$$

is an unbiased estimator of the mean or expected value $\pi(i,P)$ of treatment i. So just like the empirical success rule, the binomial average rule estimates the expected payoff of each treatment and then selects the treatment with the higher estimate. In the special case in which P is binomial, the binomial average rule and the empirical success rule coincide. This will no longer be true in the case where N is odd which we consider next.

The binomial average rule is defined for N odd as it is when N is even except for the following adjustment of the recommendation to correct for the unbalanced sample. Follow the recommendation based on the sample of the first $N-1$ rounds of the test phase if some treatment is recommended with probability 1 and hence there is no tie of the binomial averages. Otherwise, recommend the treatment tested in round N if and only if it yielded a transformed payoff $\tilde{y}_N$ equal to 1. Hence, if there is a tie up to round $N-1$ and the treatment tested in round N yields a transformed payoff equal to 0 then recommend the treatment not tested in the last round. Formally, treatment 1 is recommended with probability 1 if

$$\begin{aligned}
\frac{2}{N-1}\sum_{k<N:t_k=1}\tilde{y}_k &> \frac{2}{N-1}\sum_{k<N:t_k=2}\tilde{y}_k \\
\text{or } \frac{2}{N-1}\sum_{k<N:t_k=1}\tilde{y}_k &= \frac{2}{N-1}\sum_{k<N:t_k=2}\tilde{y}_k \text{ and } t_N=1 \text{ and } \tilde{y}_N=1 \\
\text{or } \frac{2}{N-1}\sum_{k<N:t_k=1}\tilde{y}_k &= \frac{2}{N-1}\sum_{k<N:t_k=2}\tilde{y}_k \text{ and } t_N=2 \text{ and } \tilde{y}_N=0,
\end{aligned}$$

and with probability 0 otherwise.

We would like to point out two differences to the behavior under an even sample due to this special treatment of the last test when N is odd. (i) The recommendation after an odd sample is always deterministic; one of the two treatments is recommended with probability 1. In other words, the binomial average rule is a singleton rule when N is odd but not when N is even. (ii) The binomial average rule and the empirical success rule no longer coincide when P is binomial and N is odd. E.g. assume that $N=3$ and that the test phase yielded history $((1,1),(2,1),(1,1))$. Then the empirical success rule recommends each treatment equally likely while the binomial average rule recommends treatment 1 with probability 1.

As in the case of N even, when N is odd then the order of tests does not influence the performance of the rule as long as (a) a flip of a fair coin determines which of the two treatments is tested once more than the other, and (b) a random observation of

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

N	0	1	2	3	4
p_1	$\frac{1}{2}$	$\frac{1}{2}+\frac{1}{2}(\pi_1-\pi_2)$	$p_1(1)$	$\frac{1}{2}+\frac{1}{2}(2-\pi_1-\pi_2+2\pi_1\pi_2)(\pi_1-\pi_2)$	$p_1(3)$

the treatment tested more often is used to break ties analogously to how this is done above.

Notice that the binomial average rule can also be described as a rule for simultaneous random sampling. In this alternative setting the only adjustment is that there is no longer an alteration between treatment testing.

There is an alternative method for intuitively deriving the recommendation under the binomial average rule that applies regardless of whether N is odd or even. The underlying interpretation will be important to understand later results. The idea is to assume that the decision maker believes for any test that yields payoff y that the treatment not tested would have yielded payoff $1-y$ and to recommend the treatment that yields the higher binomial average based on these beliefs, choosing each treatment equally likely when there is a tie. Formally, treatment 1 is recommended with probability 1 (with probability 1/2) if

$$\frac{1}{N}\left(\sum_{k:t_k=1}\tilde{y}_k+\sum_{k:t_k=2}(1-\tilde{y}_k)\right)>(=)\,\frac{1}{N}\left(\sum_{k:t_k=1}(1-\tilde{y}_k)+\sum_{k:t_k=2}\tilde{y}_k\right)$$

and hence if

$$\sum_{k:t_k=1}\tilde{y}_k+\sum_{k:t_k=2}(1-\tilde{y}_k)>(=)\,\frac{N}{2} \tag{1}$$

which is equivalent to the separate formulae given above for N even and N odd. Notice that (1) is also useful for saving on the number tests, namely if

$$\left(\sum_{k\leq M:t_k=1}\tilde{y}_k+\sum_{k\leq M:t_k=2}(1-\tilde{y}_k)\right)>N/2$$

holds for some $M<N$ then treatment 1 can be recommended after M tests without running the remaining $N-M$ tests.

We show the probability of choosing treatment 1 using the binomial average rule and facing a binomial distribution for small values of N in Table 1, setting $p_1(N)=p_1(\sigma^*,P,N)$ and $\pi_i=\pi(i,P)$ for $i\in\{1,2\}$.

The fact that performance in any even period in the above table is the same as in the preceding odd period is a more general phenomenon (see part (i) below). Of course, a tie between the binomial average of each treatment does not help in selecting a specific treatment and ties occur more likely when N is even. However, the fact that the last test in an even number of tests has no impact on the expected recommendation is nevertheless somewhat surprising. Notice that expectation refers here and throughout the paper to ex-ante calculations made before running the first test assuming a specific P.

Proposition 1 *Assume that σ^* is the binomial average rule and $N = 2n$ is even. Then*

(i) $p_i(\sigma^*, P, 2n-1) = p_i(\sigma^*, P, 2n)$ *and* $\pi(\sigma^*, P, 2n-1) = \pi(\sigma^*, P, 2n)$,

(ii) $\pi(j, P) > \pi(i, P)$ *implies* $p_j(\sigma^*, P, 2n+1) > p_j(\sigma^*, P, 2n)$ *and* $\pi(\sigma^*, P, 2n+1) > \pi(\sigma^*, P, 2n)$

Part (ii) shows that the binomial average rule increases "performance" as the sample size grows. In a later result (Proposition 4) this will be complemented with an efficiency type result:

$$\lim_{N\to\infty} \pi(\sigma^*, P, N) = \max\{\pi(i, P), \pi(j, P)\}.$$

Proof. Recall that $N = 2n$ is assumed to be even. Fix some distribution P. As we are interested in p_i, given the definition of the binomial average rule, we can assume without loss of generality that P is binomial. The proof is purely combinatorial.

It is simpler for the argument if we act as if any payoff y generated choosing treatment 2 during the test phase was really payoff $1-y$ generated by treatment 1. With this transformation, the binomial average rule recommends treatment 2 with certainty if strictly less than half the tests yielded payoff 1 and recommends both treatments equally likely if half the tests yielded payoff 1.

We now proceed to prove part (i) by showing $p_i(\sigma^*, P, N-1) = p_i(\sigma^*, P, N)$ which immediately implies $\pi(\sigma^*, P, N-1) = \pi(\sigma^*, P, N)$.

Consider the recommendation after $N-1$ tests as if generated by running N tests and ignoring the outcome of the test in round N. We focus on the situations in which the recommendations after $N-1$ tests and after N tests differ. It is easily verified that, given the recommendation after N tests, the recommendation after $N-1$ tests would have been different if and only if half of the N tests yielded payoff 1 (and hence each treatment is recommended equally likely after the N tests). If the last test yielded payoff 0 (payoff 1) then the recommendation after $N-1$ tests is treatment 1 (treatment 2) with certainty. Symmetry of the binomial average rule shows that among all

realizations in which half of the N tests yield payoff 1, the ex-ante probability that the last test yields payoff 1 equals $1/2$. Thus, the two possible changes in recommendation cancel each other and we obtain the same expected recommendation when performing N tests as when running $N-1$ tests which was the statement to be proven.

We now prove part (ii). Assume $\pi(i,P) < \pi(j,P)$. Consider the recommendation based on the even number of N samples. Then the recommendation can only change after an additional sample (drawn using the binomial rule) if there was a tie in the empirical success of transformed payoffs of each treatment after N samples. So we can limit our attention to an event of such a tie. Due to the symmetry of the binomial average rule, the expected payoff of the recommended treatment after N samples conditional on such a tie equals

$$\frac{1}{2}\pi(1,P) + \frac{1}{2}\pi(2,P). \tag{2}$$

Given the event of such a tie assume that treatment i is tested in round $N+1$. Then the probability that treatment i is recommended after $N+1$ samples equals $\pi(i,P)$, so the expected payoff of the recommended treatment equals

$$\pi(i,P)^2 + (1-\pi(i,P))\,\pi(j,P).$$

As it is equally likely that a tie occurs and treatment j is tested in round $N+1$, the overall effect conditional on a tie equals

$$\begin{aligned} & \frac{1}{2}\left(\pi(1,P)^2 + (1-\pi(1,P))\,\pi(2,P)\right) + \frac{1}{2}\left(\pi(2,P)^2 + (1-\pi(2,P))\,\pi(1,P)\right) \\ = \; & \frac{1}{2}\pi(1,P) + \frac{1}{2}\pi(2,P) + \frac{1}{2}\left(\pi(1,P) - \pi(2,P)\right)^2. \end{aligned}$$

Thus, if τ_N is the probability that a tie occurs after N samples, then

$$\pi(\sigma^*,P,N+1) - \pi(\sigma^*,P,N) = \frac{1}{2}\left(\pi(1,P) - \pi(2,P)\right)^2 \tau_N$$

which completes the proof of part (ii). ■

6 Minimax Regret with 2 Treatments

We now move to the main objective of the paper, to analyze which strategy the decision maker should select when P is unknown. First we have to postulate how to deal with the environment being uncertain.

How to select a strategy σ^* if the policy maker does not have a prior? We may like σ^* to generate a *consistent estimator* so

$$\lim_{N\to\infty} \Pr\left(t_N(\sigma^*) \in \arg\max_{i\in T} \pi(i,P)\right) = 1$$

holds for all P where $t_N(\sigma)$ denotes the random treatment recommended by strategy σ after a test phase of length N. We may like σ^* to be rational for some prior Q, a property that is called *admissible*. This puts some discipline on σ^* and ensures that selection is as close as possible to the rationality setting. Finally, we may like to have some axiomatic foundations of the procedure for selecting a strategy. Selection according to "minimax regret" will yield a strategy that satisfies these three properties. The alternative of selecting according to maximin will be discussed later.

Regret $r(\sigma,P)$ is defined as the difference between the expected payoff of the best treatment and the expected payoff realized by using strategy σ when the environment is given by P. This is what we refer to in the introduction as *error*. Formally

$$r(\sigma,P) = \max_{i\in T} \pi(i,P) - \pi(\sigma,P).$$

In this paper we search for a strategy σ^* that attains *minimax regret*, formally

$$\sigma^* \in \arg\min_{\sigma} \sup_{P\in\Delta(Y^T)} r(\sigma,P)$$

and set $r_N^* = \inf_\sigma \sup_{P\in\Delta(Y^T)} r(\sigma,P)$. The underlying idea is that each strategy is evaluated according to the maximum regret it creates among all possible distributions P. According to our assumption, the only information the decision maker has about the environment P is that $P \in \Delta\left(Y^T\right)$. Strategies with lower maximal regret are preferred and hence a strategy σ^* (if it exists) that yields the lowest maximum regret is most preferred, in which case $r_N^* = \sup_{P\in\Delta(Y^T)} r(\sigma^*,P)$. Minimax regret was introduced by Savage (1951) based on an interpretation of Wald (1950) for making decisions in uncertain environments, an axiomatic framework underlying minimax regret is due to Milnor (1954).[9][10] Two papers (Manski, 2004, Stoye, 2005) specifically investigated minimax regret in this setting, their results will be compared to ours later.

[9] The connection to the formal settings of Savage (1951) and Milnor (1954) are established by identifying each $P \in \Delta\left(Y^T\right)$ with a state of the world.

[10] Two axioms underlying the axiomatization of Milnor (1954) are central: (i) independence of irrelevant alternatives is relaxed; preferences are not allowed to change when adding an action that does not change the best payoff in any of the states (ii) the independence axiom is strengthened by replacing the constant lottery by one where payoffs are only required to be constant across actions in any given state.

The nice thing about minimax regret is that it also yields information about how large the error of a rational (or Bayesian) decision maker endowed with a prior $\bar{Q}$ and who hence chooses a best response $\bar{\sigma}\left(\bar{Q}\right)$. Note first that a rational decision maker will minimize expected regret as $\sigma \in \arg\max_{\sigma} \pi\left(\sigma, \bar{Q}\right)$ if and only if $\sigma \in \arg\min_{\sigma} r\left(\sigma, \bar{Q}\right)$. Given the saddle point characterization of Savage (1954) that will be used in the proof below, provided there is a minimax regret strategy, then

$$r_N^* = \inf_{\sigma} \sup_{P \in \Delta(Y^T)} r\left(\sigma, P\right) = \sup_{Q \in \Delta(\Delta(Y^T))} \inf_{\sigma} r\left(\sigma, Q\right).$$

So $r\left(\bar{\sigma}\left(\bar{Q}\right), \bar{Q}\right) \leq r_N^*$ holds for all $\bar{Q}$ which means that the expected error of a rational decision maker is bounded above by r_N^*. Later we will see via the saddle point characterization that this bound is tight as it is attained for some prior $\bar{Q}$.

A key result of this paper is part (i) and (ii) of the following. Let $B\left(j, m, z\right) = \binom{m}{j} z^j \left(1-z\right)^{m-j}$ be the probability of drawing j successes among m independent samples of a Binomial distribution with success probability z where $j, m \in \mathbb{N}_0$ with $0 \leq j \leq m$ and $z \in [0,1]$.

Proposition 2 *(i) The binomial average rule attains minimax regret. The empirical success rule does not. The value r_N^* of minimax regret is given by*[11]

$$r_N^* = \max_{u \in \left(\frac{1}{2}, 1\right)} \left(\left(2u-1\right) \sum_{n < N_{odd}/2} B\left(n, N_{odd}, u\right) \right). \tag{3}$$

where $N_{odd} = \max\left\{n \in \mathbb{N} \text{ s.t. } n \leq N \text{ and } \left(n+1\right)/2 \in \mathbb{N}\right\}$. r_N^* is the maximal regret that can be generated if the decision maker instead has a prior.

(ii) If N is odd then any rule that attains minimax regret makes the same (deterministic) recommendation as the binomial average rule.

(iii) Assume that P is restricted to be binomial. Then the binomial average rule attains minimax regret. The empirical success rule attains minimax regret if and only if N is even. The value of minimax regret equals r_N^* specified in (3).

Notice that the binomial average rule has been constructed in way that the testing procedure during the first n rounds of the testing phase does not depend on the size of N provided $N \geq n$. Consequently, we can apply the binomial average rule even when N is not known.

[11] The first order conditions are $2\sum_{n=0}^{(N-1)/2} \left(\binom{N}{n} u^n \left(1-u\right)^{N-n}\right) - \left(2u-1\right) N \binom{N-1}{\frac{N-1}{2}} \left(u\left(1-u\right)\right)^{\frac{N-1}{2}} = 0$.

Corollary 1 *The binomial average rule attains minimax regret if the sample size N is randomly drawn from an unknown distribution.*

On the other hand, if N is known then we can use Proposition 1 to save tests which immediately implies that all subjects are be treated equally.

Corollary 2 *If N is known then minimax regret can be attained with an odd number of tests followed by some treatment recommended with probability one.*

Proof. (of Proposition 2) Ever since Savage (1954, Theorem 1, p. 186), minimax regret is best calculated by finding an equilibrium (or saddle point) of the following simultaneous move zero sum game between the decision maker and nature. Pure strategies of the decision maker are given by deterministic strategies σ_d defined above. The set of environments $\Delta\left(Y^T\right)$ is the set of pure strategies of nature. Both parties may choose mixed strategies so the decision maker chooses strategy σ and nature chooses a prior Q. The payoff of the decision maker equals $r(\sigma, Q)$ while that of nature is defined as $-r(\sigma, Q)$. Here

$$r(\sigma, Q) := \int r(\sigma, P)\, dQ(P) = \int \max_i \pi(i, P)\, dQ(P) - \pi(\sigma, Q).$$

Given this equation, the decision maker minimizes regret given Q if and only if she chooses a best response to Q, i.e. maximizes expected payoffs $\pi(\sigma, Q)$ over all σ. We will find a prior Q^* such that (σ^*, Q^*) is an equilibrium of this game. Thus, σ^* will be admissible. Moreover, given that this is a saddle point, $Q^* \in \arg\max_Q \inf_\sigma r(\sigma, P)$. So Q^* is also called a *worst case prior* as it is the prior under which the best response is furthest away from the benchmark of an omniscient decision maker who chooses the best treatment in each environment.

We will show that (σ^*, Q^*) is an equilibrium of this game where Q^* is defined as follows. Let

$$u^* \in \arg\max_{u \in \left(\frac{1}{2}, 1\right)} \left[(2u-1) \left(\sum_{n < N/2} B(n, N, u) + \frac{1}{2} B(N/2, N, u)\, 1_{\{N \text{ even}\}} \right) \right] \tag{4}$$

where some $u^* \in \left[\frac{1}{2}, 1\right]$ clearly exists as the expression in the bracket is continuous and bounded in u. The fact that $u^* \in \left(\frac{1}{2}, 1\right)$ exists follows as the value of the bracket is generally non negative while for $u^* \in \left\{\frac{1}{2}, 1\right\}$ it equals 0. Let P^1 be the

binomial distribution such that $P^1((1,0)) = 1 - P^1((0,1)) = u^*$. Let P^2 be the binomial distribution such that emerges when swapping labels of treatments in P^1, so $P^2((0,1)) = 1 - P^2((1,0)) = u^*$. Let Q^* be such that $Q^*(P^1) = Q^*(P^2) = \frac{1}{2}$.

First we show that σ^* is a best response to Q^*. Neither P^1 nor P^2 puts weight on events in $\{(0,0),(1,1)\}$ so each treatment yields the same information when facing Q^*. So we can act as if only treatment 1 is tested. Given the symmetry of Q^* it is clear that the recommendation by σ^* given the information from the test phase is a best response. In fact it is the unique best response unless treatment 1 yielded the same number of successes as failures.

Now we show that Q^* maximizes regret given σ^*. Given the definition of σ^* we can restrict attention to binomial P. In the following we show for P such that $\pi(1,P) - \pi(2,P)$ is constant that $p_2(\sigma^*,P)$ is maximized if and only if $\pi(1,P) + \pi(2,P) = 1$.

Given Proposition 1 we can restrict attention to the case where N is even. To simplify notation, let $x = \pi(1,P)$, $y = \pi(2,P)$, $n = N/2$ and $p_2 = p_2(\sigma^*,P)$. We will show that $\frac{d}{dx}p_2 + \frac{d}{dy}p_2 \leq 0$ if and only if $x + y > 1$.

Concerning the definition of $B(j,m,x)$, if $j < 0$ or $j > m$ then set $B(j,m,x) = 0$. Treatment 2 is recommended is recommended with probability $\frac{1}{2}$ if both treatments yielded the same number of successes in the test phase and is recommended with certainty if it yields strictly more successes than treatment 1 in the test phase. So

$$p_2 = \frac{1}{2}\sum_{k=0}^{n} B(k,n,y)\,B(k,n,x) + \sum_{k=1}^{n} B(k,n,y)\sum_{j=0}^{k-1} B(j,n,x)$$

Using the fact that

$$\frac{d}{dz}B(j,m,z) = m\left(B(j-1,m-1,z) - B(j,m-1,z)\right)$$

we obtain after some intermediary steps involving rearranging terms that

$$\frac{1}{n}\left(\frac{d}{dx}p_2 + \frac{d}{dy}p_2\right) = \frac{1}{2}(x-y)\sum_{k=0}^{n-1} n\binom{n-1}{k}^2 x^k(1-x)^{n-k-1}\,y^k(1-y)^{n-k-1}\left(\frac{1}{k+1} - \frac{1}{n-k}\right).$$

Collecting terms k and $n-k-1$ for $k < (n-1)/2$ then yields

$$\frac{1}{n}\left(\frac{d}{dx}p_2 + \frac{d}{dy}p_2\right) = \sum_{k<(n-1)/2} n\binom{n-1}{k}^2 \left(\begin{array}{c}(xy)^k\left[(1-x)(1-y)\right]^{n-k-1} \\ -(xy)^{n-k-1}\left[(1-x)(1-y)\right]^k\end{array}\right)\left(\frac{1}{k+1} - \frac{1}{n-k}\right)$$

where the statement to be proven then follows from the fact that

$$\frac{(xy)^k\left[(1-x)(1-y)\right]^{n-k-1}}{(xy)^{n-k-1}\left[(1-x)(1-y)\right]^k} = \left(\frac{(1-x)(1-y)}{xy}\right)^{n-2k-1}$$

is strictly decreasing in x, taking the value 1 if and only if $x = 1 - y$ which is what we wanted to show.

Since the binomial average rule is symmetric, maximal regret is attained for some binomial P^* such that $\pi(1, P^*) > \pi(2, P^*)$ where

$$r(\sigma^*, P^*) = (\pi(1, P^*) - \pi(2, P^*)) p_2(\sigma^*, P^*).$$

Our result above shows that $\pi(1, P^*) + \pi(2, P^*) = 1$ and hence

$$r(\sigma^*, P^*) = (2\pi(1, P^*) - 1) p_2(\sigma^*, P^*)$$

which equals the expression in the brackets in (4). Given the definition u^* and Q^* it follows that Q^* maximizes regret given σ^*. This proves part (i).

Part (ii) follows from the fact that the best response is unique when N is odd.

Concerning part (iii) assume that P is restricted to be binomial. The "if" statement follows from the above together with the fact that the binomial average rule and the empirical success rule coincide. Assume N is odd. Then the empirical success rule $\bar{\sigma}$ does not achieve minimax regret as it is not a best response against Q^*. To see this, consider the event in which both treatments yield payoff 1 up to round $N - 1$ and the treatment chosen in round N yields payoff 0. Then the empirical success rule recommends each treatment equally likely while the best response is to recommend the treatment not chosen in the final round. ■

In the following we illustrate why the binomial average rule does not yield the only possible recommendation when N is even. Assume that N is even and that outcomes are binary valued. Only an odd number of tests is needed. So the decision maker can also attain minimax regret when N is even by applying the binomial average rule but simply ignoring the last test. So this is a rule based on an equal number of tests of each treatment that attains minimax regret but that makes a different recommendation than the binomial average rule.

We present some properties of r_N^*. Using the central limit theorem to approximate r_N^* given in (3) one easily derives its rate of convergence when N is large. Combining this with Proposition 1 shows the following.

Proposition 3 *$r_{N-1}^* = r_N^* > r_{N+1}^*$ when N is even ($N \geq 2$) and $r_N^* \approx \frac{0.17}{\sqrt{N}}$ when N is large.*[12]

[12] 0.17 is the solution z to $2\int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = z\sqrt{\frac{2}{\pi}} e^{-z^2/2}$ rounded to four digits after the comma.

6.1 Foregone Payoffs

In the proof of Proposition 2 we found that there is a worst case prior Q^* that puts weight only on binomial distributions P in which $\pi(1,P) + \pi(2,P) = 1$ so $P(\{(0,0),(1,1)\}) = 0$. When faced with a worst case prior, it is as if the decision maker observes in each round of the test phase the payoff that the treatment not tested would have achieved. Payoffs realized by treatments not chosen are called *foregone* payoffs. So it is as if the decision maker can observe foregone payoffs when facing a worst case prior. Consider now a decision maker who always observes foregone payoffs during the test phase, hence tries both treatments on the same subject in each round. Thus, there is nothing to decide on during the test phase (as both treatments are tested simultaneously), the only question is what to recommend given the N random observations of both outcomes.

Since there is no harm to also observing foregone payoffs, minimax regret is weakly smaller when a decision maker can observe foregone payoffs than when he or she cannot (as in the main model of this paper). On the other hand, when faced with the worst case prior Q^* defined in the proof of Proposition 2, there is no advantage in observing the payoff of each treatment in each test. Thus, the maximal regret among the recommendations based on foregone payoffs is at least as large as the maximal regret under the binomial average rule. Combining these two observations we obtain in terms of minimax regret that there is no need to (or "sense" in) observing the payoff of each treatment in each test.

Corollary 3 *Assume that the decision maker observes the outcome of each treatment in each round of test phase. Then the recommendation of the binomial average rule (evaluated by disregarding this additional information) attains minimax regret.*

6.2 Literature

We briefly comment on the literature. Stoye (2005) is interested in recommendations when each treatment is tested equally often. For $T = 2$, P binomial, N even and a given random sample of $N/2$ observations of each treatment, Stoye (2005, Proposition 1) shows that the recommendation of the empirical success rule attains minimax regret and claims that the value of minimax regret is given by (3). Given our results above, this statement is also true if P is not restricted to the binomial case and can more generally realize payoffs in $[0,1]$. Stoye (2005, Proposition 4 (i)) also derives the

recommendation under minimax regret based on a single test of each treatment (so $N = 2$) for the more general case of payoffs in $[0, 1]$. Given our results we provide the minimax regret recommendation whenever it is exogenously given that each treatment has to be tested equally often. Moreover, since sampling is part of the decision we find that sampling each treatment equally often when N is even is sufficient to minimize maximum regret.

6.3 The Value of Minimax Regret for Small Samples

We illustrate the value of minimax regret for small values of N by a cross in Figure 1 below. Figure 2 shows $\sqrt{N} r_N^*$ for some odd N.

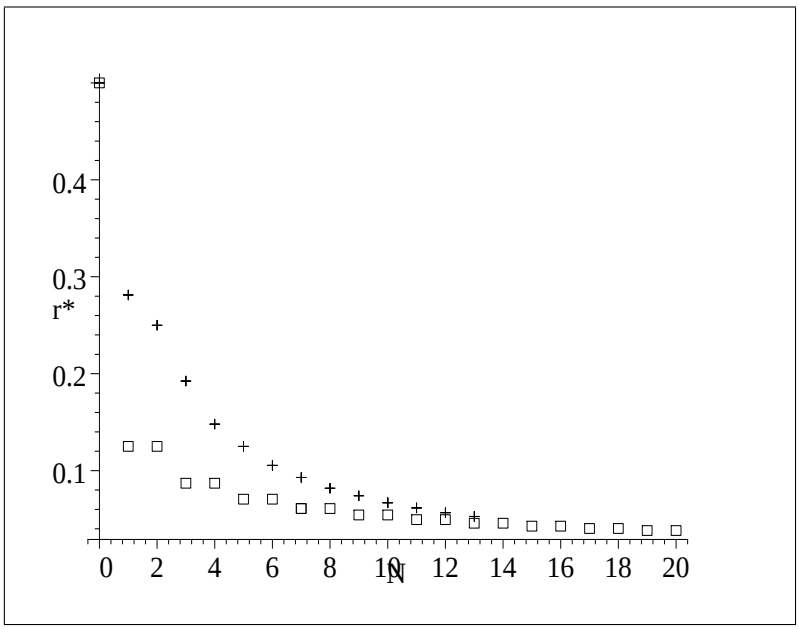

Figure 1: Value of minimax regret r_N^* (box) and lower bound on maximal regret under emprical success rule (cross) as function of sample size N.

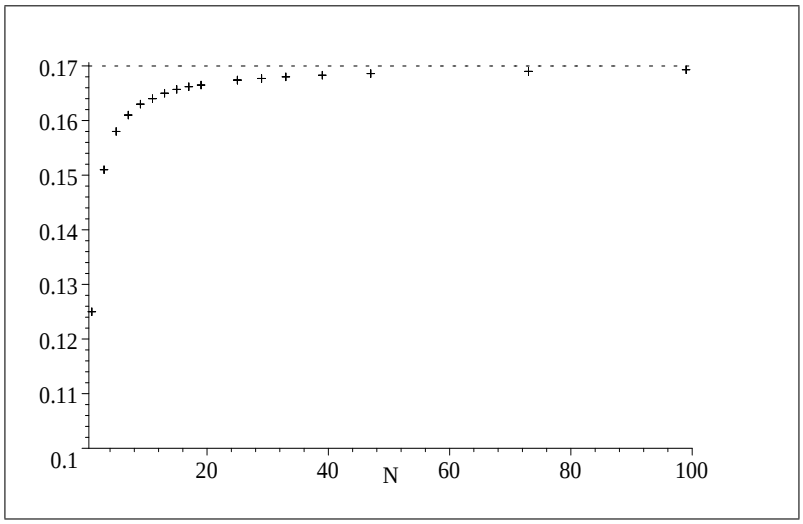

Figure 2: Plot of $\sqrt{N}r_N^*$ with limit value 0.17 as $N \to \infty$ added as dotted line.

The numerical values are given in the following table, we include the first value of N for which minimax regret is below 5%, 4%, 3%, 2.5%, 2% and 1%. In particular, observe the following.

Remark 1 11 *tests yield a maximal regret (or error) of* 0.0495.

It turns out that $r_N^* \approx \frac{0.17}{\sqrt{N+0.8}}$ is a very good approximation when N is a small odd number, we provide the values in the table below.

N	0	1	3	5	7	9	11	13
r_N^*	0.5	0.125	0.087	0.0706	0.0609	0.0543	0.0495	0.0458
$\sqrt{N}r_N^*$	0	0.125	0.151	0.158	0.161	0.163	0.164	0.165
$\frac{0.17}{\sqrt{N+0.8}}$	/	0.127	0.0872	0.0706	0.0609	0.0543	0.0495	0.0458

N	15	17	19	25	29	33	39
r_N^*	0.0428	0.0403	0.0382	0.0335	0.0311	0.0292	0.0269
$\sqrt{N}r_N^*$	0.1657	0.1662	0.1665	0.1674	0.1677	0.168	0.1683
$\frac{0.17}{\sqrt{N+0.8}}$	0.0428	0.0403	0.0382	0.0335	0.0311	0.0292	0.0269

N	47	73	99	199	289	∞
r_N^*	0.0246	0.0198	0.017	0.012	0.00998	0
$\sqrt{N}r_N^*$	0.1686	0.169	0.1693	0.1695	0.1697	0.17
$\frac{0.17}{\sqrt{N+0.8}}$	0.0246	0.0198	0.017	0.012	0.00999	0

6.4 Empirical Success

We would like to show how much worse the empirical success (or empirical success) rule $\bar{\sigma}$ performs in terms of maximal regret when N is small and payoffs belong to $[0,1]$. We do not attempt the demanding task of calculating the precise value of maximal regret under the empirical success rule. Instead we provide a lower bound for $\bar{r}_N = \sup_{P\in\Delta(Y^T)} r(\bar{\sigma}, P)$.

To obtain a lower bound, let $P = P_{x,z} \in \Delta\left([0,1]^2\right)$ be such that treatment 1 yields the value z for sure while treatment 2 is binomial with success probability given by x, $x, z \in [0,1]$. If N is even, $n = N/2$ and $\frac{1}{n}x > z > 0$ then it is easily verified that

$$\begin{aligned} r(\bar{\sigma}, P_{x,z}) &= (x-z)(1-x)^n \\ \bar{r}_N &\geq \sup_{x,z:x>nz>0} r(\bar{\sigma}, P_{x,z}) = \frac{n^n}{(1+n)^{1+n}}. \end{aligned} \tag{5}$$

If instead $N = 2n+1$ for some $n \in \mathbb{N}$ and sampling alternates as under the binomial average rule then

$$\begin{aligned} r(\bar{\sigma}, P_{x,z}) &= \frac{1}{2}(x-z)\left((1-x)^{n+1} + (1-x)^n\right) \\ \bar{r}_N &\geq \sup_{x,z:2x>Nz>0} r(\bar{\sigma}, P_{x,z}) = \frac{1}{2+n}\left(\frac{n}{2+n}\right)^{\frac{n}{2}}. \end{aligned} \tag{6}$$

We add the values given in (5) and (6) for $N \leq 10$ to Figure 1, these values remain above r_N^* for $N < 20$. However, this is no longer true for $N = 20$ when the lower bound

equals 0.035 while $r^*_{20} = 0.0382$. This clearly shows that our lower bound is not tight for all N.

6.5 Large Sample Behavior

In the following we show formally what happens when samples are large. Both the binomial average rule σ^* and the empirical success rule $\bar{\sigma}$ recommend the best treatment with arbitrary high probability provided that the sample is sufficiently large. In other words, both rules generate consistent estimators. The bounds necessary for this statement to be true can be chosen uniform provided that the absolute difference in performance of the two treatments is bounded below. For the formal statement, let $t^N(\sigma, P)$ be the random variable that specifies the treatment recommended by strategy σ given N and P.

Proposition 4 *For any $\varepsilon > 0$ there exists N_0 such that for any $N > N_0$ and any P with $|\pi(1,P) - \pi(2,P)| \geq \varepsilon$ we have*

$$\Pr\left(t^N(\sigma^*, P) = t^N(\bar{\sigma}, P) = \max\{\pi(1,P), \pi(2,P)\} | P\right) > 1 - \varepsilon \text{ and } r_N(\sigma^*, P) \leq \varepsilon.$$

The additional statement that regret is uniformly bounded above is easily verified.

Corollary 4 *The recommended treatment under the binomial average rule is a uniformly consistent estimator.*

We do not provide a formal proof as it is an immediate consequence of the law of large numbers and the fact that the possible variance of P is bounded above.

7 Maximin and T=2

Depending on the discipline there is an alternative popular method 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 procedure was introduced by Wald (1950) and was first axiomatized by Milnor (1954). Formally, $\hat{\sigma}$ attains *maximin* if

$$\hat{\sigma} \in \arg\max_{\sigma} \inf_{P} \pi(\sigma, P).$$

It is straightforward to show in our setting that any strategy attains maximin. Minimal payoffs for any strategy σ are generated by the trivial binomial distribution P that satisfies $\pi(i, P) = 0$ for all i. Consequently, all strategies are equally good in terms of their minimal payoff. In particular, the strategy that only tests treatment 1 in the test phase and then recommends treatment 1 regardless of the outcomes in the test phase attains maximin.

Proposition 5 *Any strategy attains maximin. In particular, a strategy that attains maximin need not generate a recommendation that is a consistent estimator of the best treatment.*

This result is analogous to the finding of Manski (2005) for the case of a single unknown treatment (see also below) that the unknown treatment is never recommended under maximin regardless of how large the sample is.

8 Covariates and T=2

In the following we enrich the setting with two treatments and include covariates (or attributes) as in Manski (2004, Section 3). A *covariate* is an observable characteristic of a subject and treatments can be recommended depending on the value of this covariate. In addition, the decision maker may sample among subjects with some specified covariate, in the formal model below the decision maker has to specify for each test which specific covariate it should be tested on.

We adapt our previous notation and formalism to this setting. Let X be a non-empty finite set of covariates (or attributes) with typical element ξ. Our previous setting will be embedded as the special case where $|X| = 1$. The distribution of covariates is known to the decision maker where p_ξ denotes the probability that a random subject has covariate ξ where each covariate is assumed by a strictly positive fraction so $p_\xi > 0$ and $\sum_{\xi \in X} p_\xi = 1$. To keep notation simple, we assume that the set of possible outcomes Y is the same for each covariate and later we discuss how matters change if different covariates are known to yield different outcomes. The outcome realized by a treatment now also depends on the covariate so $P \in \Delta\left(\left(Y^2\right)^X\right)$ with $P(\xi)(y)$ being the probability that outcome y_i results assigning treatment i to the class of subjects

with covariate ξ for each $i \in \{1,2\}$.[13] Let

$$\pi(\xi, i, P) = \int y_i dP(\xi)(y)$$

be the expected payoff of treatment i for covariate ξ. Consequently, there is a best treatment for each covariate where a best treatment for one covariate may not be best for a different covariate.

Again we have to differentiate between sequential and simultaneous experimentation.

8.1 Simultaneous Experimentation

In the following we consider the setting in which the covariate and treatment tested next may not depend on previous outcomes. So the decision maker predetermines the number $n_{\xi,t}$ of tests of treatment t to be run on covariate ξ. The complete description will be denoted by n so $n = (n_{\xi,t})_{\xi \in X, t \in \{1,..,2\}}$. Let $\Delta_d N$ denote the set of such assignments so $\Delta_d N$ is a subset of $\Delta(\mathbb{N}_0 \times \{1,2\})^X$ such that $n \in \Delta_d N$ implies $\sum_{\xi \in X} \sum_{i=1}^{2} n_{\xi,i} = N$. Of course assignment can be random so $\nu \in \Delta(\Delta_d N)$ will denote the distribution. In addition the decision maker has to make a recommendation $\sigma_\xi(n)$ for each covariate ξ separately given the number of tests $n_{\xi 1}$ and $n_{\xi 2}$ run with each treatment. σ_ξ is formally defined in Section 2, taking $n_{\xi 1} + n_{\xi 2}$ as sample size. We denote the strategy by (ν, σ) where $\sigma = (\sigma_\xi(n))_{\xi,n}$. If ν puts all weight on a single element n then we speak of *stratified random sampling* (Manski, 2004).

We will apply a variant of the binomial rule. Accordingly, the decision maker randomly chooses how many tests to run on each covariate and then follows the binomial rule to execute how many treatments and how to recommend. Let M be the set of all $m \in \mathbb{N}_0^{|X|}$ such that $\sum m_\xi = N$. Let $\bar{\sigma}(m)$ be the binomial average rule applied independently to each covariate separately where covariate ξ is tested m_ξ times. So $\bar{\sigma}(m)$ is the description of the binomial average rule applied to a sample of size m_ξ. Random assignment of tests means that the decision maker may choose $\mu \in \Delta M$. So this yields a rule we denote by $\bar{\sigma}(\mu)$ that of course can be formally embedded in the above notation (ν, σ).

Proposition 6 *There exists $\mu^* \in \Delta M$ such that $\bar{\sigma}(\mu^*)$ attains minimax regret.*

[13] We use $\left(Y^T\right)^X$ instead of $Y^{X \times T}$ as the former is more useful for proofs.

Proof. We will first construct a specific zero sum game between the decision maker and nature, show than an equilibrium exists and later show that its equilibrium is a saddle point of the zero game related to minimax regret.

M will be the set of pure strategies of the decision maker.

Let P_u^1 and P_u^2 be such that

$$u = P_u^1((1,0)) = 1 - P_u^1((0,1)) = P_u^2((0,1)) = 1 - P_u^1((1,0)).$$

Let $Q_u \in \Delta\left(\{0,1\}^2\right)$ be such that $Q_u\left(P_u^1\right) = Q_u\left(P_u^2\right) = \frac{1}{2}$. Given $v \in \left[\frac{1}{2},1\right]^X$ let $\hat{Q}_v$ be the prior that is independent across variables where the marginal $\left(\hat{Q}_v\right)_\xi$ on the covariate ξ is set equal to Q_{v_ξ}. The set of pure strategies of nature will be $\left[\frac{1}{2},1\right]^X$.

The game is set up as a zero sum game where the payoff of nature is given by $r\left(\left(m,\bar{\sigma}(m)\right),\hat{Q}_v\right)$.

As argued using Glicksburg (1952) in similar games above, this game has a saddle point (μ^*,ν^*) where $\mu^* \in \Delta M$ and $\nu^* \in \Delta\left(\left[\frac{1}{2},1\right]^X\right)$. As behavior of the decision maker when recommending for ξ only depends on outcomes in ξ we can assume that $\nu^* \in \times_{\xi\in X}\Delta\left[\frac{1}{2},1\right]$.

Now consider the game associated to minimax regret where the decision maker chooses σ and nature chooses Q. In the following we will show that $\left(\bar{\sigma}(\mu^*),\hat{Q}_{v^*}\right)$ is a saddle point and hence that $\bar{\sigma}(\mu^*)$ attains minimax regret. We build on the properties of the binomial rule established in the proof of Proposition 2.

First consider the decision maker. Notice that for any given m_ξ, $\bar{\sigma}(m_\xi)$ is a best response to $Q_{v_\xi^*}$ as $Q_{v_\xi^*}$ is symmetric. Hence $\bar{\sigma}(\mu^*)$ minimizes regret given $\hat{Q}_{v^*}$.

Now consider nature. Due to independent behavior under $\bar{\sigma}(\mu^*)$ across covariates we can restrict ourselves to covariate ξ and to only binary outcomes. In the following we will show that $\left\{\arg\max_P r\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,P\right)\right\} \cap \{P \text{ s.t. } \pi(1,P)+\pi(2,P)=1\} \neq \emptyset$ where $r\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,P\right) = \max_{i\in\{1,2\}}\pi(\xi,i,P) - \pi\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,P\right)$. Note that $r\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,\bar{P}\right) = \sum_m \mu^*(m)\, r\left(\xi,\bar{\sigma}(m_\xi),\bar{P}\right)$. In our proof of Proposition 2 we showed that $r\left(\xi,\bar{\sigma}(m),P\right)$ can be increased for each m by replacing $\bar{P}$ with $\tilde{P}$ such that $\pi\left(1,\tilde{P}\right) - \pi\left(2,\tilde{P}\right) = \pi\left(1,\bar{P}\right) - \pi\left(2,\bar{P}\right)$ and $\pi\left(1,\bar{P}\right) + \pi\left(2,\bar{P}\right) = 1$. Thus $r\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,P\right)$ can be maximized by some P with $\pi(1,P)+\pi(2,P)=1$. Given our the description of the game presented at the beginning of this proof, we obtain that $r\left(\xi,\bar{\sigma}\left(\mu^*\right)_\xi,P\right)$ is maximized by $Q_{v_\xi^*}$ which completes the proof. ■

We immediately extend a result of Manski (2004) shown for sufficiently large N to all $N \geq 1$, namely that treatment choice can be conditioned on each covariate.

Corollary 5 *For all N, minimax regret can be attained by ignoring outcomes realized during testing on covariates $\xi' \neq \xi$ when recommending a treatment for covariate ξ.*

Given our proof of Proposition 6, it follows immediately that the value of minimax regret is given by

$$\min_{\mu \in \Delta M} \sup \left\{ r\left(\bar{\sigma}\left(\mu\right), P\right) \text{ s.t. } \pi\left(1, P_\xi\right) + \pi\left(2, P_\xi\right) = 1 \text{ for all } \xi \right\} \tag{7}$$

which is bounded above by $\sum p_\xi r^*_{m_\xi}$ as

$$\sup \left\{ r\left(\bar{\sigma}\left(m\right), P\right) \text{ s.t. } \pi\left(1, P_\xi\right) + \pi\left(2, P_\xi\right) = 1 \text{ for all } \xi \right\} = \sum p_\xi r^*_{m_\xi}.$$

Thus $\min_{m \in M} \sum p_\xi r^*_{m_\xi}$ can be used as an upper bound on the value of minimax regret, a very useful result that has already been shown by Stoye (2005) for a more general setting and then used for binomial distributions.

Notice that however that Stoye (2006, Proposition 6) also claims that the decision maker should completely separate the decision problem across covariates. This would mean that Corollary 5 would be due to him and would imply that the bound $\min_{m \in M} \sum p_\xi r^*_{m_\xi}$ is tight. However this is not true, it is (7) that is tight. We give some intuition using the interpretation of the solution to minimax regret via a game between the decision maker and nature. If nature would know (by receiving additional information) how many tests are run on each covariate then $\min_{m \in M} \sum p_\xi r^*_{m_\xi}$ would be a tight bound. For instance, it is as if nature knows how many tests are run if the decision maker chooses a deterministic allocation of treatments to covariates during testing as under stratified random sampling. In other words, this is the correct bound if the decision maker is restricted to stratified random sampling. So Proposition 6 in Stoye (2005) is correct if one adds this restriction. We provide a counter example to show that the minimax regret rule actually sometimes mixes and chooses a non deterministic μ. In this example we show how to apply the tight bound given in 7.

Assume $N = 1$, $X = \{a, b\}$ and $p_a \leq \frac{1}{2}$. Then using the formula in Table 1 we obtain

$$\begin{aligned} r\left(\bar{\sigma}\left(\mu\right), Q_v\right) &= p_a\left(2v_a - 1\right)\left[\mu_{(1,0)}\left(\frac{1}{2} + \frac{1}{2}\left(1 - 2v_a\right)\right) + \mu_{(0,1)}\frac{1}{2}\right] \\ &\quad + \left(1 - p_a\right)\left(2v_b - 1\right)\left[\mu_{(1,0)}\frac{1}{2} + \mu_{(0,1)}\left(\frac{1}{2} + \frac{1}{2}\left(1 - 2v_b\right)\right)\right]. \end{aligned}$$

Finding a saddle point we formally obtain the following. If $p_a \leq \frac{1}{5}$ then $\mu_b = 1$ and consequently $v_a = 1$ and $v_b = \frac{3}{4}$ so $r = p_a r^*_0 + p_b r^*_1 = p_a \frac{1}{2} + p_b \frac{1}{8}$. If $\frac{1}{5} < p_a \leq \frac{1}{2}$ then

$\mu_b = \frac{1}{2}\sqrt{\frac{p_b}{p_a}} < 1$, $v_a = 1$ and $v_b = \frac{1}{2}\left(1 + \sqrt{\frac{p_a}{p_b}}\right)$ so that $r = \frac{1}{2}\sqrt{p_a p_b} < p_a r_0^* + p_b r_1^*$. This is also very intuitive. When one covariate is too unlikely then it is never tested. Above the threshold (here $p_a \geq \frac{1}{5}$) the decision maker tests both covariates with positive probability, increasing the probability of testing covariate a as p_a increases until each covariate is tested equally likely when $p_a = \frac{1}{2}$. We illustrate minimax regret and the bound given by the corollary in Figure 3.

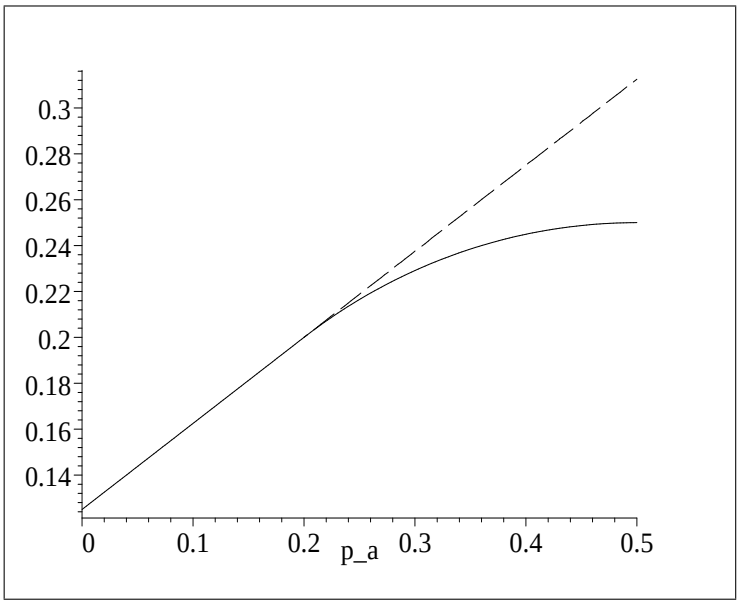

Figure 3: Minimax regret as a function of p_a with the upper bound from the Corollary given as dotted line.

For larger sample sizes we expect the advantage of mixing over sample sizes to be less dramatic as the above result seems to rest on the strong curvature of r_N^* for small N. However it will generally not be a good idea to never test one covariate while sampling some other one that is equally likely. We expand briefly on this and investigate when it is best to only sample covariate b and hence invoke stratified sampling when $X = \{a, b\}$. First note that following Proposition 1, if N is even then both covariates need to be tested in order to attain minimax regret regardless of how small or how large p_a is (provided $p_a \in (0, 1)$). Assume N is odd. If a is not tested under minimax regret rule then $r^* = p_a \frac{1}{2} + p_b r_N^*$. On the other hand, testing covariate a in a single test with probability ε will yield at most regret $p_a\left((1-\varepsilon)\frac{1}{2} + \varepsilon\frac{1}{8}\right) + p_b\left((1-\varepsilon)r_N^* + \varepsilon r_{N-1}^*\right)$. This is because we act as if nature observes the testing of the decision maker which

then of course increases regret. So if a is not tested then by taking derivatives with respect to ε we obtain $p_b\left(r^*_{N-1} - r^*_N\right) \geq p_a\left(\frac{1}{2} - \frac{1}{8}\right)$ or $r^*_{N-2} - r^*_N \geq \frac{3}{8}\frac{p_a}{1-p_a}$. So for instance if a is not tested and $N = 11$ then $p_a \leq 0.0127$.

In the following we briefly investigate how one should assign tests to covariates in order to minimize the bound under the stratified random sampling provided above (see Stoye, 2005). Consider the case of $X = \{a, b\}$. First we assume that N is sufficiently large so that we can replace r^*_N be $\frac{0.17}{\sqrt{N}}$. Then $\lambda_i = n_i/N$ be the proportion of tests assigned to covariate i (so $\lambda_a + \lambda_b = 1$) where we ignore integer constraints. Let r^*_p be the value of minimax regret under covariate distribution p. Then

$$r^*_p \leq p_a \frac{0.17}{\sqrt{\lambda_a N}} + p_b \frac{0.17}{\sqrt{\lambda_b N}} = \frac{0.17}{\sqrt{N}}\left(p_a \frac{1}{\sqrt{\lambda_a}} + p_b \frac{1}{\sqrt{\lambda_b}}\right).$$

The expression on the right hand side is minimized when

$$p_a = \frac{(\lambda_a)^{\frac{3}{2}}}{(\lambda_a)^{\frac{3}{2}} + (\lambda_b)^{\frac{3}{2}}} \tag{8}$$

and hence

$$r^*_p \leq \frac{0.17}{\sqrt{N}\left((\lambda_a)^{\frac{3}{2}} + (\lambda_b)^{\frac{3}{2}}\right)} \tag{9}$$

where $\lambda_a = \lambda_a(p_a)$ is the solution to (8). We plot (8) in Figure 4.

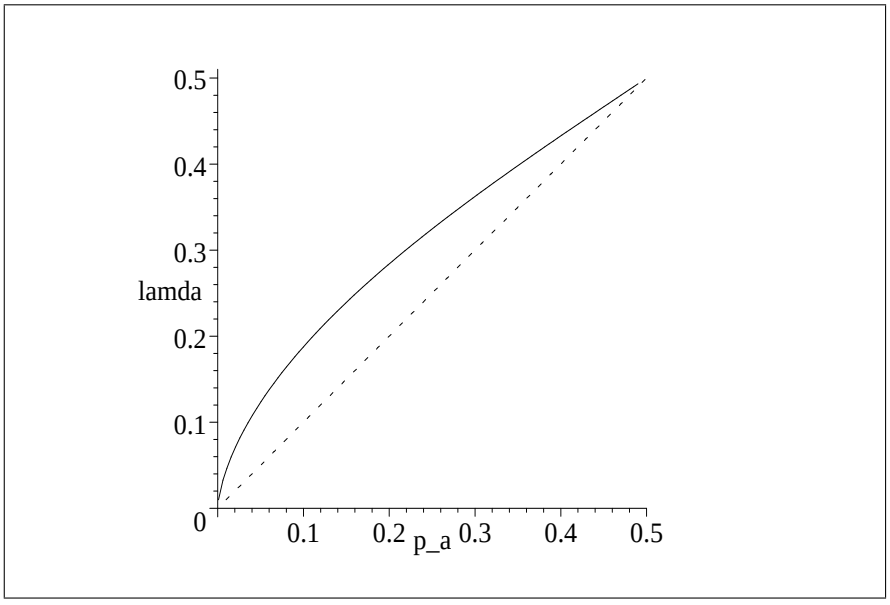

Figure 4: Asymptotically optimal frequency of testing covariate a as a function of the frequency of covariate a.

How many tests can be saved asymptotically by running them optimally as compared to sampling randomly? In order to guarantee regret to be below the value r_N^* achieved without covariates, our results above say that you have to multiply the number of tests by $1/\left(\left(\lambda_a\right)^{\frac{3}{2}}+\left(\lambda_b\right)^{\frac{3}{2}}\right)^2$ where $\lambda_a=\lambda_a\left(p_a\right)$ is the solution to (9). We compare this to the factor $\left(\sqrt{p_a}+\sqrt{p_b}\right)^2$ that emerges asymptotically from sampling covariates at random in Figure 5. The difference is maximally 0.097 when $p_a\approx 0.05$. So up to 10% tests can be saved asymptotically when going from random sampling to optimal sampling.

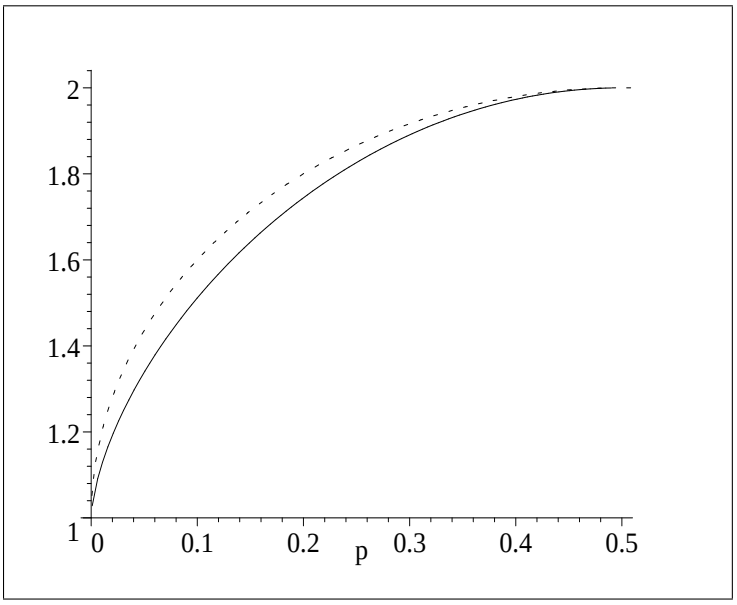

Figure 5: Impact factor due to assignment of asymptotically best covariates as function of frequency of covariate a (dotted shows the analogous expression when covariates are tested at random).

We briefly mention an important extension. What should the decision maker do if he or she does not know the distribution of covariates? Let us assume that $\left(p_\xi\right)_\xi$ is added to the quantifier maximizing regret, so $\left(p_\xi\right)_\xi$ is chosen by nature. Disregarding integer constraints, if the decision maker tests each covariate equally often then minimax regret is given by $r_{N/|X|}^*$. If the decision maker decides to deterministically allocate tests to covariates (so $\mu^*\left(m\right)=1$ for some m) then this bound cannot be improved on.

Corollary 6 *If p_ξ is not known by the decision maker then minimax regret is bounded above by $r^*_{\lfloor N/|X| \rfloor}$.*

In particular, we can ensure an error of 5% with $11 * |X|$ samples.

The above results indicate that performance is worse the more covariates there are. However this need not be true if different covariates yield different outcome ranges. We can only briefly comment on the possible results one can obtain when outcome ranges are covariate specific.

Let Y_ξ be the set of outcomes that can be realized by some treatment on covariate ξ. Assume that the decision maker has a single preference order over $\cup_{\xi \in X} Y_\xi$. Continue as in the setting above, in particular assuming that Y_ξ has a most preferred outcome y^ξ_H and a least preferred outcome y^ξ_L. Normalize utility u in the same way as above, the least preferred outcome in $\cup_{\xi \in X} Y_\xi$ is assigned value 0 and the most preferred is assigned value 1. Now binomial distributions on covariates assign only probabilities to the best outcome y^ξ_H and worst outcome y^ξ_L in Y_ξ. Consequently, to apply the binomial average rule, payoffs realized are transformed relative to these covariate specific best and worst outcomes. However, in order to maintain comparability of outcomes across covariates, the payoffs achieved on each covariate have to be scaled down using the range of utilities achievable by the covariate specific outcomes. It is sufficient to make the following adjustment. In all formulae above, replace $\pi(\xi, i, P)$ by $\left(u\left(y^\xi_H\right) - u\left(y^\xi_L\right)\right) * \pi(\xi, i, P)$. This works because regret is based on differences only so we need not control for the differences in $u\left(y^\xi_L\right)$ across ξ. This covariate specific transformation of payoffs can have a substantial impact. For instance, regret with four covariates that each can yield the same outcomes is bounded in Corollary 6 above by $r^*_{N/4}$. Now assume for illustration that it is known that the range of each covariate in terms of utility is half of the total range, so $u\left(y^\xi_H\right) - u\left(y^\xi_L\right) = \frac{1}{2}$ for all $\xi \in X$. Then invoking Corollary 6 but adjusting for the smaller outcome range of each covariate we obtain that minimax regret is bounded above by $\frac{1}{2} r^*_{N/4}$ which is approximately equal to r^*_N when N is large.

8.2 Sequential Sampling

Consider now briefly the more general setting of sequential sampling. We would like to point out that simultaneous sampling is not executed when sequential sampling is available and $N \geq 5 * |X|$. We do not need to add formal arguments as intuition is simple.

Consider a decision maker using a minimax regret strategy that we can assume can be represented by $\bar{\sigma}(\mu^*)$ given Proposition 6. Draw $\bar{m}$ using μ^* and start testing the covariate ξ' that is tested most often under $\bar{m}$. This covariate will be tested at least 5 times. As mentioned in Section 5, it is not necessary to run all tests to obtain the recommendation. Particularly, if in the first $\lfloor m_{\xi'}/2 \rfloor + 1$ tests treatment 1 yielded only successes and treatment 2 only failures then treatment 1 will be recommended to covariate ξ' regardless of future outcomes. Thus one can use the remaining $m_{\xi'} - \left(\lfloor m_{\xi'}/2 \rfloor + 1\right)$ tests to gather more information about some other covariate. Since at least 2 tests are reallocated to a different covariate, maximal regret decreases strictly. Hence the minimax regret rule will never be based on simultaneous sampling when $N \geq 5 * |X|$.

9 Finitely Many Unknown Treatments

Consider now the general setting with T treatments with uncertain outcomes. We generalize the binomial average rule in the obvious way by invoking pairwise comparisons. While we conjecture that it attains minimax regret under simultaneous experiments, due to the increased complexity we are unable to prove this conjecture. Thus, for completeness we first ensure existence and later use the binomial average rule to provide an upper bound on minimax regret.

Proposition 7 *Under either simultaneous or sequential testing there is a strategy that attains minimax regret and that first randomly transforms payoffs obtained during the testing phase into binary values as under the binomial average rule before it then conditions the recommendation on these binary values.*

Proof. We only sketch the proof as it is very simple. As in the proof of Proposition 2, all we have to do is to establish existence of a saddle point in the fictitious zero sum game between the decision maker and nature. We first consider only binomial environments. So the pure strategy of nature is to choose $P \in \Delta\left(\{0,1\}^T\right)$. The decision maker chooses a deterministic strategy that only evaluates histories in which all outcomes are binary. Nature aims to maximize expected regret while the decision maker aims to minimize expected regret.

We establish existence of an equilibrium. The set of pure strategies of the decision maker is finite. While the set of pure strategies for nature is infinite, it is convex

and compact if we consider the topology induced by considering the weights on the corners $\{0,1\}^T$. Furthermore, regret is continuous in P. Following Glicksburg (1952) a saddle point exists. Precisely, first associate each rule σ with a pure strategy, observe that sets of pure strategies are compact and convex and that payoffs are continuous to obtain existence. Then note for the decision maker that the equilibrium mixed strategy which is a mixture over rules can actually be identified with a rule itself given the representation of rules as behavior strategies.

Now consider such a strategy that attains minimax regret when facing only binomial P. Extend it to a general strategy by first transforming payoffs randomly into binary ones as done under the binomial average rule. Given the linearity of this transformation, the expected payoff attained by this strategy when facing any general P' is the same as the expected payoff it achieves when facing the binomial P that satisfies $\pi(i,P) = \pi(i,P')$ for all i. Thus, maximal regret under this strategy is attained under some binomial P. Consequently, this strategy also attains minimax regret when facing any P'. ■

9.1 Binomial Average Rule for Many Treatments

In the following we generalize the binomial average rule to more than two treatments.

One obvious way of extending the binomial rule to more than two treatments is to compare success pairwise. This easily generates a well defined recommendation when N is a multiple of T as in this case each treatment can be tested N/T times. However we want to have a rule for general sample sizes which is constructed as follows.

The binomial average rule defined for an arbitrary number of treatments is defined as follows for sequential randomized experiments. Randomly select an ordering or permutation of the treatments where each permutation is selected equally likely. Let $\rho \in \{1,..,T\}^{\{1,..,T\}}$ be such that ρ_o is the treatment assigned to the o-th element of the order. Test treatments in this order until the test phase is over and N tests have been run. So test treatment $\rho_{1+((k-1) \bmod T)}$ in round k. Then transform payoffs achieved in each round of the test phase randomly into a binary payoff as done for the case of $T = 2$. Let n_i be the number of tests run on treatment i. Search for a treatment i such that

$$\sum_{k:t_k=i} \tilde{y}_k + \sum_{k:t_k=j} (1-\tilde{y}_k) \geq \frac{1}{2}(n_i + n_j) \text{ holds for all } j \neq i. \tag{10}$$

Below we show that such an i exists. If there are several treatments that have this

property then randomize among them with equal probability. More formally, let $A \subseteq \{1,..,T\}$ be the set of treatments for which (10). Then recommend $i \in A$ with probability $1/|A|$.

Notice that the above description is also easily adapted to the setting of simultaneous experiments. Notice also that the above algorithm reduces to comparing the binomial average realized by each treatment when N is a multiple of T.

We show that the algorithm described above yields a well defined recommendation in $\Delta\{1,..,T\}$. This follows once we show that the following preference ordering $\succsim$ on $\{1,..,T\}$ is complete and transitive. Accordingly,

$$i \succsim j \text{ if either } j = i \text{ or if } \sum_{k:t_k=i} \tilde{y}_k + \sum_{k:t_k=j} (1-\tilde{y}_k) \geq \frac{1}{2}(n_i + n_j).$$

Clearly this preference ordering is complete. We easily verify that it is also transitive. Consider $i,j,k \in \{1,..,T\}$ such that $|\{i,j,l\}| = 3$. Assume that $i \succsim j$ and $j \succsim l$. Then

$$\left(\sum_{k:t_k=i} \tilde{y}_k + \sum_{k:t_k=j} (1-\tilde{y}_k)\right) + \left(\sum_{k:t_k=j} \tilde{y}_k + \sum_{k:t_k=l} (1-\tilde{y}_k)\right) \geq \frac{1}{2}(n_i + n_j) + \frac{1}{2}(n_j + n_l)$$

which implies

$$\sum_{k:t_k=i} \tilde{y}_k + \sum_{k:t_k=l} (1-\tilde{y}_k) \geq \frac{1}{2}(n_i + n_l)$$

and hence $i \succsim l$.

Notice that the above rule sometimes recommends a random treatment. In the following we show that the binomial average rule can be slightly adjusted to yield a singleton rule at not cost to its behavior in terms of regret. Notice that if $i,j \in A$ then $n_i = n_j$. Now we learned for the setting of $T = 2$ that we can ensure a deterministic recommendation by avoiding sampling the same number of times. Similarly we can do here by comparing two treatments going back to the last round in which there was a different number of observations. More formally, let $n_i(m)$ be the number of times that treatment i is sampled up to and including the m-th round of the test phase. Let $m^*_{ij} = \max\{m \leq N : n_i(m) \neq n_j(m)\}$. Then define

$$i \succ_s j \text{ if } \sum_{k \leq m^*_{ij}:t_k=i} \tilde{y}_k + \sum_{k \leq m^*_{ij}:t_k=j} (1-\tilde{y}_k) > \frac{1}{2}\left(n_i\left(m^*_{ij}\right) + n_j\left(m^*_{ij}\right)\right).$$

By construction we obtain that $\succ_s$ induces a complete strict preference ordering. We verify transitivity. Given the transitivity of $\succsim$ all we have to do is to consider the case

where $n_i = n_j = n_k$. We show that there is a unique selected treatment among i, j, and k. Assume that treatment k was tested last. If $\tilde{y} = 1$ then $k \succ_s i$ and $k \succ_s j$ which means that k is selected. If instead $\tilde{y} = 0$ then $i \succ_s k$ and $j \succ_s k$ and either i or j is selected. We summarize.

Remark 2 *The binomial average rule can be adjusted to yield a singleton rule that yields the same expected payoff.*

9.1.1 Simultaneous Randomized Experiments

Consider the setting in which tests are predetermined. We derive the following upper bound on regret under the binomial average rule which is hence an upper bound on the value of minimax regret.

Proposition 8

$$\sup_P r(\bar{\sigma}, P) \leq (T-1)\left(\frac{\binom{T-N \bmod T}{2}}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor - 1} + \left(1 - \frac{\binom{T-N \bmod T}{2}}{\binom{T}{2}}\right) r^*_{2\lfloor N/T \rfloor + 1}\right).$$

Given that the value of minimax regret is generally non increasing, the bound can also be replaced by the simpler expression $(T-1) r^*_{2\lfloor N/T \rfloor - 1}$. When N is large, using our result on the convergence rate of r^*_N (see Proposition 3), either $r^*_{\frac{2N}{T(T-1)^2}}$ or $(T-1)\sqrt{\frac{T}{2}} r^*_N$ can be used as an approximate bound. In particular this means that the rate of convergence does not depend on the number of treatments but that going from $T = 2$ to $T = 3$ treatments, the number of tests has to be multiplied by $\frac{T(T-1)^2}{2}$ to guarantee the same error.

Proof. We derive an upper bound on $r(\bar{\sigma}, P^*)$ where $P^* \in \arg\max_P r(\bar{\sigma}, P)$. Since the binomial average rule is symmetric in the sense that it does not depend on labelling of treatments, we can assume that $\pi(1, P^*) = \max_i \{\pi(i, P^*)\}$. So

$$r(\bar{\sigma}, P^*) = \pi(1, P^*) - \sum_{i=1}^{T} p_i(\bar{\sigma}, P^*, N)\, \pi(i, P^*) = \sum_{i \neq 1} p_i(\bar{\sigma}, P^*, N)\left(\pi(1, P^*) - \pi(i, P^*)\right).$$

Let $\bar{\sigma}_{ij}$ be the recommendation made under the same testing procedure but where instead the decision maker chooses only between treatment i and j using the recommendation of the original binomial average rule for $T = 2$ applied to the observations of tests on treatments i and j. Notice that under the testing procedure for $T > 2$ two events may only occur. Either treatment i and treatment j were tested equally often

or one of the two was tested once more often than the other where each treatment is equally likely to be the one tested more often.

If N is multiple of T so $N \bmod T = 0$ then each is tested N/T times. If instead $N \bmod T > 0$ then treatment i is either tested $\lfloor N/T \rfloor$ or $\lfloor N/T \rfloor + 1$ times where $\lfloor x \rfloor = \max \{x' \in \mathbb{N}_0 \text{ such that } x' \leq x\}$. Let $\tilde{n}_i$ and $\tilde{n}_j$ be the random variables describing how often treatments i and j were tested. Then

$$\begin{aligned} \Pr\left(\tilde{n}_i = \tilde{n}_j = \lfloor N/T \rfloor\right) &= \frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}} \\ \Pr\left(\tilde{n}_i = \tilde{n}_j = \lfloor N/T \rfloor + 1\right) &= \frac{\binom{N \bmod T}{2}}{\binom{T}{2}} \\ \Pr\left(\tilde{n}_i \neq \tilde{n}_j\right) &= \frac{(T - N \bmod T)(N \bmod T)}{\binom{T}{2}} \end{aligned}$$

where $\binom{m_2}{m_1} := 0$ if $m_2 < m_1$.

If treatment i is recommended by $\bar{\sigma}$ then it is also recommended by $\bar{\sigma}_{1i}$, hence

$$p_i(\bar{\sigma}, P^*, N) \leq p_i(\bar{\sigma}_{1i}, P^*, N).$$

In the following we will put a bound on

$$p_i(\bar{\sigma}_{1i}, P^*, N)(\pi(1, P^*) - \pi(i, P^*)).$$

Assume that both treatments were tested $\lfloor N/T \rfloor$ times. Then under $\bar{\sigma}_{1i}$ it is as if the original binomial rule for two treatments was applied to a sample of size $2\lfloor N/T \rfloor$. Thus,

$$p_i(\bar{\sigma}_{1i}, P^*, N)(\pi(1, P^*) - \pi(i, P^*)) = p_i(\bar{\sigma}, P^*, 2\lfloor N/T \rfloor)(\pi(1, P^*) - \pi(i, P^*)) \leq r^*_{2\lfloor N/T \rfloor}.$$

Using such arguments it follows that

$$\begin{aligned} & p_i(\bar{\sigma}_{1i}, P^*)(\pi(1, P^*) - \pi(i, P^*)) \\ \leq\ & \frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor} + \frac{(T - N \bmod T)(N \bmod T)}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor + 1} + \frac{\binom{N \bmod T}{2}}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor + 2} \\ =\ & \frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor - 1} + \left(1 - \frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}}\right) r^*_{2\lfloor N/T \rfloor + 1} \end{aligned}$$

and hence

$$r(\bar{\sigma}, P^*) \leq (T - 1)\left(\frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}} r^*_{2\lfloor N/T \rfloor - 1} + \left(1 - \frac{\binom{T - N \bmod T}{2}}{\binom{T}{2}}\right) r^*_{2\lfloor N/T \rfloor + 1}\right).$$

■

Consider $T = 3$. For $N = 3$ we find $\sup_P r(\bar{\sigma}, P) = -\frac{20}{81} + \frac{14}{81}\sqrt{7} \approx 0.21$ and the upper bound equal to $2r_1^* = \frac{1}{4}$. For $N = 6$ we find the corresponding values 0.134 and 0.174, for $N = 9$ we find 0.109 and 0.141. Our results on the convergence rate of r_N^* would indicate a drop in regret of $\frac{1}{\sqrt{2}} \approx 0.707$ between $N = 3n$ and $N = 3(n+1)$ for n large which should be compared to $\frac{0.134}{0.21} \approx 0.638$ for $N = 3$ to $N = 6$ and to $\frac{0.109}{0.134} \approx 0.813$ for $N = 6$ to $N = 9$. The comparison between $T = 2$ and $T = 3$ above hints that the factor $\frac{2}{T(T-1)^2}$ can be multiplied to the number of tests to translate regret for $T = 2$ to $T = 3$. Note that $\frac{2}{3*2^2} * 9 = 1.5$ where $r_1 = 0.125$.

9.1.2 Sequential Randomized Experiments

Unlike the case of two treatments it seems that sequential testing can outperform simultaneous testing when $T = 3$. In the following we illustrate for the case of $T = N = 3$ that the binomial average (or empirical success) rule can be outperformed by appropriate sequential sampling when P is binomial.

Assume $T = N = 3$ and consider only binomial P. Consider the empirical success rule $\bar{\sigma}$ defined analogously as under the setting of two treatments.[14] In the appendix we verify the following. The recommendation of the empirical success rule achieves minimax regret conditional on testing each treatment. However it does not achieve minimax regret when the sample is endogenous as we find an alternative strategy that yields lower maximal regret. While the empirical success rule $\bar{\sigma}$ yields $\sup_P r(\bar{\sigma}, P) = -\frac{20}{81} + \frac{14}{81}\sqrt{7} \approx 0.21$, the alternative strategy $\hat{\sigma}$ reduces maximal regret to $\sup_P r(\hat{\sigma}, P) = \frac{11}{64} \approx 0.172$. The strategy $\hat{\sigma}$ starts by testing a random treatment but then continues by testing any treatment that yielded a success again. Treatments not tested are assigned payoff 0. Recommendation is like under the empirical success rule.

10 One Known Treatment

Consider the situation in which the mean of one of the treatments is known. As this is an important case in particular for applications we briefly comment on what insights our analysis provides. When $T = 2$ then one might imagine the unknown treatment to

[14] Test each treatment once, then recommend the treatment with the most successes and randomize equally when there are ties.

be an innovation. Generally the case of a known treatment can also be interpreted as an outside option or default that one can follow should the unknown treatments not be sufficiently successful. Our analysis is easily extended to this setting and we present two results: (i) existence of minimax regret strategy and (ii) a uniform upper bound on the value of minimax regret.

Consider the same basic setting with $T \geq 2$ as above except that we now assume that $\pi(1, P)$ is known. Formally the strategies do not change. Of course, any strategy that attains minimax regret will not test the known treatment.

We immediately obtain existence. The statement of Proposition 7 holds for this setting too, the proof is analogous.

Consider the special case of $T = 2$. Then we can in fact construct a strategy that attains minimax regret. Of course only treatment 2 will be tested. So the objective is to determine which treatment should be recommended based on N independent tests of treatment 2. However, no new rule or analysis is necessary as we can build on existing results obtained for binomial P. For the case of P binomial, Manski (2005) derived the formula to determine a minimax regret strategy numerically, Stoye (2005) specifies an equation that implicitly defines the minimax regret strategy. With our trick of transforming general payoffs from $[0, 1]$ into binary outcomes in $\{0, 1\}$ it follows immediately that the results of Manski (2005) and Stoye (2005) can be applied to the setting of general payoffs. Simply apply their cutoff rules to the binomial average of the tests of the unknown treatment 2 to obtain a rule that attains minimax regret.

When $T > 2$ then we do not know of a rule that attains minimax regret. One might of course conjecture that binomial averages are used to compare the two unknown treatments.

Of course one can always apply the binomial rule when one treatment is known by acting as if there are only unknown treatments but instead of testing the unknown treatment, acting as if $\pi(1, P)$ was observed whenever it was tested. Without integer constraints this means that we fictitiously add $\frac{N}{T-1}$ tests to obtain a total sample of $\frac{T}{T-1}N$ observations. Respecting integer constraints we can use a sample $N' = N + \lfloor N/(T-1) \rfloor$ in order to be sure that the random allocation of which treatments should be tested more often is compatible with N. Combining this with 8 we obtain an upper bound on minimax regret for the case where the mean of one treatment is known. Of course the real value of minimax regret under one known treatment will depend on

$\pi(1, P)$ and can substantially differ from this bound. For instance in the trivial cases $\pi(1, P) \in \{0, 1\}$ the value of minimax regret is 0 as a best treatment is known.

Proposition 9 *The value of minimax regret under one known treatment and N samples is bounded above by the value of maximal regret achieved by the binomial average rule under T unknown treatments and sample size $N + \lfloor N/(T-1) \rfloor$. For large N this means that the value of minimax regret under one known treatment is approximately bounded above by $r^*_{\frac{2N}{(T-1)^3}}$.*

References

[1] Glicksburg, I.L. (1952) "A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points," *Proc. Nat. Acad. Sci.* **38**, 170-174.

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

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

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

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

[6] Savage, L.J. (1954), *The Foundations of Statistics*, John Wiley & Sons..

[7] 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.

[8] Schlag, K. H. (2003), *How to Minimize Maximum Regret under Repeated Decision-Making*, Mimeo, European University Institute, http://www.iue.it/Personal/Schlag/papers/regret7.pdf.

[9] Stoye, J. (2005), Minimax Regret Treatment Choice with Finite Samples, Mimeo.

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

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

A Three Treatments

The aim of this section is to show that the binomial average rule does not attain minimax regret under sequential sampling when $T = N = 3$.

Consider P binomial and let $x = \pi(1, P)$, $y = \pi(2, P)$ and $z = \pi(3, P)$.

Let $\bar{\sigma}$ be the empirical success rule generalized from the setting of $T = 2$.[15] Then

$$\begin{aligned} p_1(\bar{\sigma}) &= x(1-y)(1-z) + \frac{1}{2}x(y(1-z) + z(1-y)) \\ &\quad + \frac{1}{3}(xyz + (1-x)(1-y)(1-z)) \\ p_2(\bar{\sigma}) &= y(1-x)(1-z) + \frac{1}{2}y(x(1-z) + z(1-x)) \\ &\quad + \frac{1}{3}(xyz + (1-x)(1-y)(1-z)) \\ p_3(\bar{\sigma}) &= z(1-x)(1-y) + \frac{1}{2}z(x(1-y) + y(1-x)) \\ &\quad + \frac{1}{3}(xyz + (1-x)(1-y)(1-z)) \end{aligned}$$

and

$$r(\bar{\sigma}) = x - (p_1 x + p_2 y + p_3 z) \text{ if } x = \max\{x, y, z\}.$$

We find $r(\bar{\sigma}, P)$ is maximized for $x = 1$ and $y = z = \frac{4}{3} - \frac{1}{3}\sqrt{7}$ and hence $\sup_P r(\bar{\sigma}, P) = -\frac{20}{81} + \frac{14}{81}\sqrt{7} \approx 0.21$. Let P^i be binomial such that $P^i((1,1,1)) = 1 - P^i(e_i) = \frac{4}{3} - \frac{1}{3}\sqrt{7}$ (≈ 0.45) and let $\bar{Q}$ be such that $\bar{Q}(P^i) = \frac{1}{3}$ for $i = 1, 2, 3$ where e_i is the unit vector on treatment i. Then $\sup_P r(\bar{\sigma}, P) = r\left(\bar{\sigma}, \bar{Q}\right)$. While the recommendation of the binomial average rule is a best response to this prior given the specification of the testing we demonstrate below that alternative testing improves performance.

Consider the following alternative strategy $\hat{\sigma}$. Select each treatment with equal probability to be tested in the first round of the test phase and proceed as follows until 3 tests have been made. If a treatment yields a success then test it again, if it yields a failure then test the treatment with the next higher index modulo 3. Treatments not tested are associated with payoff 0. Recommend the treatment that yielded the highest

[15] Results also apply to general $P \in \Delta[0,1]^T$ by first transforming payoff $y_k \in [0,1]$ into payoff $\tilde{y}_k \in \{0,1\}$ as in the $T = 2$ setting and then continuing as if P were binomial.

average payoff, randomizing equally likely if there are ties. Then

$$\begin{aligned} p_1(\hat{\sigma}) &= \frac{1}{3}x\left(x+(1-x)(1-y)\right)+\frac{1}{3}(1-x)(1-y)(1-z) \\ &\quad +\frac{1}{3}z(1-z)x+\frac{1}{3}(1-z)x+\frac{1}{3}(1-y)(1-z)x \\ p_2(\hat{\sigma}) &= \frac{1}{3}y\left(y+(1-y)(1-z)\right)+\frac{1}{3}(1-x)(1-y)(1-z) \\ &\quad +\frac{1}{3}x(1-x)y+\frac{1}{3}(1-x)y+\frac{1}{3}(1-z)(1-x)y \\ p_3(\hat{\sigma}) &= \frac{1}{3}z\left(z+(1-z)(1-x)\right)+\frac{1}{3}(1-x)(1-y)(1-z) \\ &\quad +\frac{1}{3}y(1-y)z+\frac{1}{3}(1-y)z+\frac{1}{3}(1-x)(1-y)z \end{aligned}$$

and $r(\hat{\sigma},P)$ is maximized for $x=\frac{3}{4}$ and $y=z=\frac{1}{4}$ and hence $\sup_P r(\hat{\sigma},P)=\frac{11}{64}\approx 0.172$.

We do not claim that $\hat{\sigma}$ attains minimax regret as it would be too tedious to verify such a conjecture.