Learning From The Data: A Theory Without Guessing*

Karl H. Schlag

February 5, 2025

Abstract

A decision maker who needs to choose between two actions has an exogenous "pseudo prior" that determines for each action the likelihood that they would choose this action. They also have access to a data set (small or large) on how these actions performed in the past. The decision maker seeks to use this data set in a way that makes them better off with the data than without it in the environment they face. Bayesians generically will not satisfy this property as they "guess" which environment they face and hence might be wrong. A best choice is identified.
Keywords: distribution-free, learning, treatment choice, A/B testing

1 Introduction

Data is constantly used to aid decision-making. There are several common ways of how to use data. Bayesians maximize subjective expected utility based on some prior and use the data to update this prior. Frequentists often use the data to estimate which action is best and then choose this action. A/B testing typically suggests to make choices based on the outcomes of hypothesis tests using the data.

We point out a common shortcoming of all of these approaches. Any of these decision makers can be worse off in the true environment if they sometimes alter behavior after observing the data. We provide a solution that does not have this shortcoming. We show how the decision-maker can use the data to make them weakly better off in the true

*We would like to thank Bernhard Kasberger, Anton Kolotilin, Macro Mariotti, Philipp Peitler, Larry Samuelson and Joergen Weibull for useful comments.

environment without knowing this environment, typically they will even be strictly better off. Not knowing which action is best the decision-maker mixes between the different actions and updates this mixture (or randomized action) whenever they receive data. The updating of this mixture has many features that are in common with the process of updating a prior within the Bayesian paradigm.

More specifically, we consider a decision-maker who has to choose from a finite set of actions. We limit attention to only two actions in the introduction as well as in the later part of the paper. Actions generate payoffs that belong to a given interval and that are drawn from some unknown distribution. Any distribution (also referred to as environment) generating payoffs in this interval might be the true one. The action that realizes the highest mean payoff under the true distribution is most preferred. This alone does not help much for making decisions as the true distribution is not known and there is no prior over the possible distributions.

Our decision maker has access to a data set. This data set contains some payoffs that these actions have realized. The sequence of actions in this data set is given and the payoff of each action in the data set is independently drawn from the true distribution.

The decision maker does not know the true environment. Their indecisiveness about which action is best is reflected in their choosing a mixed action. This mixed action is updated whenever they receive new information. The mixed action they would choose if they did not have this data set or if they had to make a choice prior to receiving data set is called their pseudo prior.

The pseudo prior is denoted by q. It can be the mixed action that resulted from updating a previous mixed action using previous data. When facing a novel decision the likelihood q_a can reflect the initial understanding of how likely action a will be the best one. It can represent the fraction of experts who recommend choosing action a. The uninformative pseudo prior that puts equal weight on each action captures a decision-maker who approaches the decision without any bias.

The decision-maker formally has a rule that specifies how they update the pseudo prior after observing the data set. We introduce a way to evaluate such a rule. We say that a decision-making rule can be caught guessing under q if there is some candidate true distribution under which the decision-maker is worse off with the data set than without it. This evaluation takes place prior to observing the payoffs in the data set and is computed using expectations under the true distribution. We hasten to point out the true distribution is simply there and not a random draw from some prior as implicitly assumed in the Bayesian paradigm. The term guessing comes from the observation that a Bayesian assesses (or guesses) by means of their prior the likelihood of each environment. They will not be guessing in the true environment if they happen to assign a sufficiently high likelihood to this environment. We will be interested in rules that cannot be caught guessing under q. These rules will be called non-guessing rules under q. They make the decision maker on average (under the true distribution) better off after observing the data set. These rules are unambiguously better than following the recommendation of a random expert when q_a is the frequency of experts recommending action a prior to having the data set. Notice that the rule that specifies to choose q regardless of which payoffs are observed in the data set is a non-guessing rule. Including this rule in our set of desirable rules ensures existence.

Observe that a decision-making rule cannot be caught guessing under q if and only if every Bayesian weakly prefers following this rule to choosing q. Hence, such a rule can be recommended to others (in an organization) as it can be accepted by all. With only two actions a rule that cannot be caught guessing under q chooses on average (as weighted by the true environment) the best action more likely than under q. So a rule that cannot be caught guessing under q can weakly improve the understanding of which action is best in the true environment even if this environment is not known.

As a first insight we obtain that it is not possible to improve the understanding of which action is truly best (so best in the true environment) if at the outset the decision-maker is already convinced about which action is truly best. More specifically, assume that the pseudo prior q puts all probability mass on action a. Then some mass can only be put on action a' after observing the data set if the decision-maker is convinced that action a' is better than a. However, given the richness of possible distributions, it is not possible in any finite data set to reach this conclusion. Thus, the trivial rule that chooses a regardless of which payoffs are observed in the data set is the only rule that cannot be caught guessing when $q_a = 1$. In the following we consider pseudo priors that put probability mass on each action.

We find that a rule that cannot be caught guessing is expected to choose actions as under q if both actions are equally good in the true environment. It is as if the decision-maker ignores the data set when both actions are equally good. In particular this means that any non-guessing rule randomizes whenever all payoffs in the data set are equal. We use this to show that any generic Bayesian can be caught guessing as Bayesians generically do not use mixed actions. Similarly, hypothesis tests are traditionally non random and hence can be caught guessing. We also show that rules that cannot be caught guessing are continuous in each of the payoffs in the data set. This insight is used to show that any (frequentist) rule that chooses the action that is estimated (using the data set) as being the best can be caught guessing.

We then proceed to investigate rules that are linear in each of the payoffs in the data set. Linear rules are arguably the simplest continuous rules which makes them natural candidates for non-guessing rules. Moreover, linear rules are simple to evaluate as their performance only depends on the mean payoffs in the true environment. As our central result we identify a unique best rule among the linear rules. In any environment its performance is superior to that of any other rule that is both linear and non-guessing under q. So one can avoid being caught guessing if one uses the data appropriately.

We illustrate with a simple numerical example. Consider a data set in which action a yielded payoffs 1, 3 and 5 while action a' yielded payoffs 4 and 10 and where any payoff necessarily belongs to $[0, 10]$. This seems to be a difficult problem. Action a' looks better but also was observed less often than a and the sample is extremely small. Under the uninformative pseudo prior the best linear rule specifies to choose action a' with probability 0.804.

This is how choices are determined under the best linear rule. In a first step one has to apply the randomization trick (Schlag, 2003). Each payoff x in the data set is independently randomly transformed into one of the two extreme payoffs 0 or 10 by replacing payoff x by payoff 10 with probability $x/10$ and by payoff 0 with probability $1 - x/10$. Note that this random transformation does not change the mean payoff of that observation. Thereafter the data set only contains extreme payoffs. If the sample is balanced, so each action is observed equally often, and if the pseudo prior is uninformative then the next step is very simple. Choose whichever action yielded 10 more often, randomizing equally likely if there is a tie. If either the sample is unbalanced or the pseudo prior is not uninformative then the rule needs to be slightly more sophisticated in order to deal with the underlying asymmetries. Accordingly, there is a cutoff such that action a (action a') is chosen if the number of times action a yielded 10 is more (is less) than the cutoff, randomizing at the cutoff appropriately. This cutoff depends on the total number of times the payoff 10 is observed in the transformed sample. We provide a simple formula for the value of the cutoff as well as for the mixed action chosen at the cutoff.

Next we investigate how likely the best linear rule will be choosing the truly best action in different data sets and under different pseudo priors. Up to now we only know that the best linear rule is able to increase this likelihood in each environment when using the data set. As it is not possible to learn which action is better when the two actions are arbitrarily similar we assume that there is a minimal difference between their respective means. We find that the minimal probability of choosing the truly best action converges to one as the number of samples of each action tends to infinity. If the sample is balanced then under the uninformative prior the minimal probability of choosing the truly best action is weakly larger than any other rule, even including rules that can be caught guessing.

We then investigate how learning under the best linear rule depends on the sample size, in particular how it improves as samples get larger. In all of simulations we uncover an approximate universal constant. When doubling the number of times each action appears in the data set the minimal difference between the two means needed to guarantee a given minimal probability of choosing the truly best action decreases approximately by 29%. This simple relationship allows us to nicely quantify inefficiencies introduced by an unbalanced data seta or by a pseudo prior that is not uninformative. For example, consider starting with a balanced sample and doubling the number of times one action is observed. Then the numerical calculations show that one third of the additional observations are wasted when comparing to the more efficient approach to allocate the additional observations evenly to the two actions. One third seems small given the extreme unbalancedness of the data set and the apparent difficulty of comparing actions that are observed differently often.

We then present rules that are non-randomized and that are non-guessing if the means are not too similar. We also present a weaker concept of almost no guessing to allow for new actions to be introduced when there is sufficient evidence. We also show how to include covariates, uncover connections to hypothesis testing and to social learning and show how the same methodology applies to making statements about which action is better.

1.1 Related Literature

In terms of the application, our paper is most related to the treatment choice literature, notably Manski (2004). Therein the objective is as in this paper, to make a choice after observing outcomes of the different actions. The information provided within the data is the payoff yielded by an action (bandit setting), there is no additional information about payoffs that other actions would have achieved (as in the foregone payoff setting). We fix the number of times each action is observed as in stratified random sampling (Manski, 2004). Important is that the payoffs are generated independently in each of the observations in the data set. This is very different from the alternative setting in which decision-maker has to learn from own previous choices as in the machine learning literature (e.g. Cesa-Bianchi & Lugosi, 2006) and under reinforcement (e.g. Börgers etal., 2004). Covariates can be included as explained in Section 6, similar to Manski (2004) and Stoye (2009), by making decisions separately for each vector of covariates. It is also related to A/B and A/B/C testing (Fabijan etal., 2018) which basically uses the same framework as treatment choice.

In terms of theory, to avoid being caught guessing is a general and novel suggestion for how to process information. To our knowledge, the term "guessing" has not appeared yet when formalizing concepts in decision-making. In the theory proposed in this paper, there is a choice to be made, and information has been gathered. The objective is to attain a payoff that is better when using the information than when not, regardless of the underlying truth. This concept can be applied to any decision-making or strategic setting where information has to be processed. We have established this criterion to aid our understanding of how to make choices based on data. Minimax regret is an alternative distribution-free method that has also been extensively used in the context of decision-making based on data (Manski, 2004, Schlag, 2006b, Stoye, 2009 among many others). One major advantage of our criterion is that the constraints imposed by the definition allow to construct a rule. In contrast, under minimax regret the rules presented in the literature have been found by guessing and verifying.

The criterion that the decision-maker must be better off than some benchmark for any underlying distribution is not new. It can be found in Börgers etal. (2004), which is based on the concept of absolute expediency (Lakshmivarahan & Thathachar, 1973). Therein, the decision-maker has to be better off in the next round when conditioning on their own previous choice. The criterion to always be better off can also be found in the work on social learning by Schlag (1998) (see also Schlag, 1999 and Hofbauer & Schlag, 2000). Therein, individuals are learning from others in the population. It is as if each individual has data about what others have experienced. In these papers on social learning, the benchmark is endogenous and given by the current frequency of play in the population. In the present paper the benchmark q is an exogenous input to the model. The present paper reveals optimality properties of the learning rule used in Hofbauer

& Schlag (2000), see Section 7. There is also a close connection to hypothesis testing. It turns out that our best linear non-guessing rule is equivalent to a uniformly most powerful unbiased linear test, as explained in Section 8. More generally, this paper falls within the literature on robust decision-making, which can be found in many disciplines, from statistics (Huber, 1972) to engineering (Taguchi & Phadke, 1989) to economics and mechanism design (Bergemann & Morris, 2005).

The key to finding non-guessing rules is the randomization trick. In all its generality it can be found in Schlag (2003), independently it is used for the probability ratio test by Cucconi (1968) and for decision making by Gupta & Hande (1992). It has been effectively used to compute exact solutions for decision-making (Schlag, 2006b, Stoye, 2009, Tetenov, 2012, Chen & Guggenberger, 2025), statistical hypothesis testing (Schlag, 2006a) and econometric modeling (Gossner & Schlag, 2013).

The proof behind the characterization of the best linear non-guessing rule uses standard optimization techniques. After inserting the no guessing constraint into the payoff objective, one solves constrained optimization for a given payoff distribution and then shows that the solution does not depend on the underlying distribution. The proofs related to the ability to learn the best action follow the tradition in the related literature on minimax regret. Therein, a zero sum game against nature is formulated and equilibrium strategies are guessed (Schlag, 2006b, Stoye, 2009, Chen & Guggenberger, 2025 among others).

We proceed as follows. Section 2 contains the model. In Section 3 we present some first insights and in Section 4 we discuss linear rules. Section 5 contains the results for choosing between two actions. Therein, in Subsections $5.1 - 5.6$ we present a "best" rule, some examples, large sample performance, quantification of performance, non-random non-guessing rules an an extension on almost no guessing. In Section 6 we show how covariates can be included, in Sections 7 and 8 we connect to social learning and to hypothesis testing. In Section 10 compare SEU maximizing to non-guessing. In Sectino 11 we conclude. In the appendix we present the proof of the main theorem from Section 5.1 and parameters of the "best" rule for two actions in small samples.

2 The Model

A decision-maker has some data and wants to use this to make a choice. The underlying decision problem is given as follows. The decision-maker has to choose an action a from a finite set A consisting of n actions. Let ΔA be the set of mixed actions. When deriving general properties we allow for $n \geq 3$ while in our characterization we assume $n = 2$. Each action $a \in A$ yields a random payoff Z_a, the (joint) distribution of $\{Z_a\}_{a \in A}$ is denoted by G. Let G_a be the marginal distribution of Z_a for $a \in A$. For simplicity we assume that G has finite support. Allowing also for distributions that have non finite support only complicates the proofs and exposition. We assume that the payoff of any action belongs to a common given bounded interval. Without loss of generality (given the preferences stated below), by affinely transforming payoffs, we may assume that the payoff realized by any action is contained in $[0, 1]$. So it is assumed that $Z_a \in [0, 1]$ for all $a \in A$. We will say that G is binary valued if $G_a(\{0, 1\}) = 1$ for all $a \in A$. Let μ_a^G be the expected payoff of action a, so $\mu_a^G = \int_x x dG_a(x)$ for $a \in A$.

The decision maker does not know the distribution G they face. When we wish to give particular emphasis to the distribution they face we denote it by $\bar{G}$ and refer to it as the true distribution. All the decision maker knows is that $\bar{G}$ satisfies the properties above. We refer to distributions that the decision maker thinks they could be facing as conceivable. Let $\mathcal{G}$ denote the set of conceivable distributions, so $\bar{G} \in \mathcal{G}$ and $\mathcal{G} = \{G \text{ with finite support} : G_a([0, 1]) = 1 \ \forall a \in A\}$. Later we will also consider the case where $\mathcal{G}$ is a subset of this set. All the decision-maker knows about $\bar{G}$ is that it belongs to $\mathcal{G}$. Without seeing the data (as explained below) they cannot distinguish the elements in $\mathcal{G}$. In particular, the decision maker does not have a prior over the elements of $\mathcal{G}$.

The decision-maker prefers higher expected payoffs to lower expected payoffs under the true distribution, so they weakly prefer a to a' if $\mu_a^{\bar{G}} \geq \mu_{a'}^{\bar{G}}$.[1] So a is the best action if $\mu_a^{\bar{G}} > \max_{a' \in A \setminus \{a\}} \mu_{a'}^{\bar{G}}$. This preference alone does not help much for making a choice as the decision-maker does not know $\bar{G}$ and hence does not know $\mu_a^{\bar{G}}$ for any $a \in A$.

The data or data set is given by a set $X = \{(a_i, x_i)\}_{i=1}^m$ that consists of m pairs of observations where x_i is a payoff that is has been independently realized by the action $a_i \in A$, so x_i is independently drawn from $\bar{G}_{a_i}$, $i = 1, ..., m$. In particular, the same action can appear multiple times in the data set. Moreover, not all actions from A need to be in the data set. Let $m_{\bar{a}} = |\{j : a_j = \bar{a}\}|$ be the number of times that action a appears

[1] In a slightly more general model, actions would generate outcomes, outcomes would be evaluated by von Neumann Morgenstern utilities and these utilities would be the payoffs mentioned above.

in the data set. So $m = \sum_{a\in A} m_a$. We refer to m_a as the sample size of a in the data set. In the following we will hold the sample size of each action fixed and known to the decision-maker, so only the payoffs generated are random. We call the data set *balanced* if $m_a = m_{\bar{a}}$ for $a \neq \bar{a}$. For binary valued data sets where $x_i \in \{0,1\}$ for all $i = 1,..,m$ let $r_{\bar{a}} = |\{j : a_j = \bar{a}, x_j = 1\}|$ be the number of times a success (payoff 1) was observed for action $a \in A$ and let $r = \sum_{a\in A} r_a$ be the total number of successes.

The decision-maker observes the data set X and then makes a choice. The way in which the decision-maker uses the data set to make their choice can be described by a mapping (or rule) f from $\cup_{m=1}^{\infty} (A \times [0,1])^m$ to ΔA where $f(X)_a$ is the probability of choosing action a after observing data set X. The expected payoff of their choice after observing the data set X under the true distribution is then equal to $f(X) \cdot \mu^{\bar{G}}$. Note here that it plays no role how payoffs are correlated between actions. The only relevant property of $\bar{G}$ is its marginals.

Note the randomness implicit in $f(X) \cdot \mu^{\bar{G}}$ that is due to the randomness in the realizations of the payoffs in X. To eliminate this randomness we evaluate a rule under a conceivable distribution G by fixing the actions observed in the data set as defined by $\{m_a\}_{a\in A}$ and calculate expected payoffs prior to observing the payoffs in the data set. This expectation is denoted by $E_G\left(f\left(\bar{X}\right)\right)$ and formally given by

$$E_G\left(f\left(\bar{X}\right)\right) = \int_x f\left(\{(a_j, x_j)\}_{j=1}^m\right) dG_{a_1}(x_1) ... dG_{a_m}(x_m).$$

Following our assumptions above, rule f is preferred to rule g under the true distribution $\bar{G}$ if $E_{\bar{G}}\left(f\left(\bar{X}\right)\right) > E_{\bar{G}}\left(g\left(\bar{X}\right)\right)$.

The choice remains difficult as the decision-maker does not know the true distribution $\bar{G}$. To aid their choice we introduce a mixed action $q \in \Delta A$ and use $q \cdot \mu^G$ as a benchmark for evaluating the success of the decision-maker after observing the data set. The following interpretations can be given for q. It can be the choice resulting from learning from previous data sets. The value q_a might be the proportion of times the decision-maker chose action a in similar situations. The mixed action q might be what the decision-maker would choose if they did not have data at their disposal. The value q_a might be the likelihood that the decision-maker thinks that action a is best or their degree of openness to choosing action a. The value of q_a can be seen as a model of the decision-maker's initial knowledge and understanding of how good action a is before observing the data set. Alternatively, q_a might be the proportion of experts who recommend to choose action a, capturing the wisdom of the crowd. In any of these interpretations, $q_a = 1$ would mean that under the benchmark the decision-maker would have no slightest doubt that action a should be chosen.

We refer to q as the pseudo prior. Two candidates for pseudo priors stand out. One is the pseudo prior that puts equal weight on each action, so $q_a = \frac{1}{n}$ for all $a \in A$. We call it the uninformative pseudo prior as it does not give any action an advantage or special attention, reflecting a decision-maker who has no prior information on this decision problem. Another candidate is the pseudo prior q under which an action is chosen according to the frequency with which it occurs in the data set. This pseudo prior might make sense when the actions in the data set have been sampled from some population in which case this choice of q reflects (on average) the frequencies with which the actions occur in this population. Under this pseudo prior a close connection to social learning will be revealed in Section 7.

If the decision-maker would be a Bayesian (more precisely, a subjective expected utility maximizer) then they would have a prior over the conceivable distributions in $\mathcal{G}$ and would choose an expected payoff maximizing action. Let Q be this prior. Then the Bayesian would choose a rule $\hat{f}$ that maximizes $\int_{G\in\mathcal{G}} E_G\left(f\left(\bar{X}\right)\right)\cdot\mu^G dQ\left(G\right)$. Note that the Bayesian expects that they will do better when conditioning on the data than when choosing q for any q. This is because the rule $\hat{f}$ they would choose satisfies

$$\int_{G\in\mathcal{G}} E_G\left(\hat{f}\left(\bar{X}\right)\right)\cdot\mu^G dQ\left(G\right) = \max_f \int_{G\in\mathcal{G}} E_G\left(f\left(\bar{X}\right)\right)\cdot\mu^G dQ\left(G\right) \geq \int_{G\in\mathcal{G}} q\cdot\mu^G dQ\left(G\right).$$

However, the prior Q is hypothetical. The distribution $\bar{G}$ they face is given and not drawn from any distribution. In the end the decision-maker only cares about the expected payoffs they obtain under the distribution $\bar{G}$. As the above inequality need not hold for all distributions G, they might strictly prefer q to their rule under the true distribution $\bar{G}$. Namely, it might be that

$$E_{\bar{G}}\left(\hat{f}\left(\bar{X}\right)\right)\cdot\mu^{\bar{G}} < q\cdot\mu^{\bar{G}}. \tag{1}$$

Note that this can only happen if the weight put by the decision-maker's prior Q on $\bar{G}$ is sufficiently small. In this sense they would be caught guessing that they are not facing the distribution they are actually facing.[2] Note that the property of being caught guessing does not depend on a particular realization of the payoffs, it is an observation about average performance when weighing realizations according to the underlying payoff

[2]To guess is to form an opinion from little or no evidence (Meriam-Webster Dictionary, 2024). Guessing implicitly means to take wrong decisions in to account.

distribution. This discussion raises two questions. Will Bayesians potentially be caught guessing? What properties do rules have that cannot be caught guessing? Both of these questions will be formalized and answered.

In this paper we look for rules that cannot be caught guessing. Specifically we search for rules where the decision-maker is always better off with the data than when choosing the mixed action q. We call such rules non-guessing.

Definition 1 *A rule f* can be caught guessing *under the pseudo prior q if there exists a distribution $G \in \mathcal{G}$ such that $E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G < q \cdot \mu^G$. A rule f is called a* non-guessing rule *under the pseudo prior q if it cannot be caught guessing under q, so if $E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G \geq q \cdot \mu^G$ holds for $G \in \mathcal{G}$.*

If q_a is the frequency of experts recommending action a for each $a \in A$ then with a non-guessing rule the decision maker will on average beat the wisdom of the crowd. If q is the choice made after observing the previous data set then a non-guessing rule on average makes the decision maker weakly better off after observing each additional data set. If there are only two actions (so $n = 2$) then a non-guessing will learn on average which action is better as it will choose the best action weakly more likely than the pseudo prior whenever the two means are not equal. The term average refers above to expectations taken based on the true distribution.

Note the difference between non-guessing and SEU maximization. A Bayesian stores their understanding of which environment they might be facing in a prior and updates this when observing data. A decision-maker who uses a non-guessing rule does not directly try to infer which joint distribution they are facing. Instead they store some understanding of how good each action is in their pseudo prior and update this understanding when observing data. In Section 10 we provide a more extensive comparison between SEU maximization and non-guessing.

Let $\mathcal{F}^*$ be the set of all non-guessing rules. Note that the rule to choose q regardless of the information provided by the data set is a non-guessing rule. Criteria for selecting among non-guessing rules are straightforward to add.

We obtain the following trivial characterization.

Remark 2 *Fix $q \in \Delta A$. A rule f is non-guessing under q if and only if exante before observing the data set any Bayesian weakly prefers following this rule to choosing q.*

In this spirit non-guessing rules can be used to reach consensus among a committee consisting only of Bayesians. Assume that there is a given default that is chosen if there is no agreement. Then each member of the committee will agree to a given non-guessing rule under q if q is set equal to this default. Assume instead that disagreement leads to random dictatorship. Then only give the committee access to the data set if they agree in which case each member of the committee will agree on a given non-guessing rule under q if q_a is set equal to the fraction of them who without the data set have action a as their most preferred action.[3]

In the next section we will find out whether Bayesians can be caught guessing. As our first illustration of how to evaluate the non-guessing property we choose a simple frequentist rule.

Example 3 (The Random Empirical Success Rule) *A natural way to make decisions when there is no information is to choose the action that achieved the higher average payoffs in the data set, randomizing with equal probability whenever there are ties for the highest. Formally, this is the rule* f *where* $f\left((a_j, x_j)_{j=1}^m\right)_a = \frac{1}{L}$ *if* $\frac{1}{m_a}\sum_{j:a_j=a} x_j \geq \frac{1}{m_{a'}}\sum_{j:a_j=a'} x_j$ *for all* $a' \neq a$ *where* $L = \left|\arg\max_{a''}\left\{\frac{1}{m_d}\sum_{j:a_j=a''} x_j\right\}\right|$. *We call this rule the* random empirical success rule *(short, rES).*[4] *It is easy to verify that, for any given distribution, rES learns with high probability which action is best if the data set contains sufficiently many observations of each action. However, rES can be caught guessing under any pseudo prior and any data set as we explain after Proposition 8 below. To gain some intuition let rES be denoted by* f^r, *assume* $A = \{a, a'\}$ *and consider the simplest data set in which* $m_a = m_{a'} = 1$. *Assume that the decision-maker is unknowingly facing a decision problem in which the payoff to action* a *is Bernoulli distributed while action* a' *is deterministic. Specifically, assume that* $\bar{G}$ *is such that* $\bar{G}_a(\{0,1\}) = \bar{G}_{a'}(\{z\}) = 1$ *for some* $z \in (0,1)$. *Then* $E_{\bar{G}}(f^r)_a = \mu_a^{\bar{G}}$ *so* $E_{\bar{G}}(f^r) \cdot \mu^{\bar{G}} = \left(\mu_a^{\bar{G}}\right)^2 + \left(1 - \mu_a^{\bar{G}}\right) z$. *Note that* $E_{\bar{G}}(f^r) \cdot \mu^{\bar{G}} \geq q \cdot \mu^{\bar{G}}$ *holds if and only if* $\left(E_{\bar{G}}(f^r)_a - q_a\right) \cdot \left(\mu_a^{\bar{G}} - z\right) = \left(\mu_a^{\bar{G}} - q_a\right) \cdot \left(\mu_a^{\bar{G}} - z\right) \geq 0$. *Consequently,* $E_{\bar{G}}(f^r) \cdot \mu^{\bar{G}} < q \cdot \mu^{\bar{G}}$ *when* $q_a < \mu_a^{\bar{G}} < z$. *This reveals that rES can be caught guessing for any choice of* q *when* $m_a = m_{a'} = 1$. *Treating the data as represen-*

[3]Clearly, no rule can be acceptable by any set of Bayesians if each of them is free to choose whatever they wish if they do not have access to the data set. This is because some might have a degenerate prior that puts all weight on an environment that has a unique best action and where all data sets are possible. They would choose the best action in that environment strictly not prefer any rule that does not stubbornly recommend that action.

[4]The only difference to the empirical success rule of Manski (2004) appears when several actions achieve the same maximal empirical success $\frac{1}{m_b}\sum_{j:a_j=b} x_j$. While BAR randomizes equally likely among these, the empirical success rule is non randomized. All actions are indexed and the rule selects the action with the lowest index among those with the highest empirical success.

tative ignores the underlying random nature of the decision problem and leads to some guessing.

3 First Insights

We now present some properties of non-guessing rules. These properties give initial guidance for how to construct rules and also help in the comparison to SEU maximization.

First we identify data sets in which the understanding of the decision-maker about how good an action is, as captured by the pseudo prior, may not change. It may not change if this action is not contained in the data set. Similarly it may not change if this action is ruled out a priori before observing the data. In particular, if the pseudo prior puts all mass on a single action then the decision-maker who uses a non-guessing rule must ignore the data set and choose this action regardless of the observations in the data set. These findings are closely connected to the insight that any improving social learning rule has to be imitating, as further explained in Section 7. Formally:

Proposition 4 *Let f be a non-guessing rule under q. If $m_a^X = 0$ or $q_a = 0$ then $f(X)_a = q_a$.*

We provide some intuition behind this finding. Proofs are omitted as they are straightforward. To simplify the arguments assume that there are only two actions a and a'. Consider first the case where action a does not occur in the data set, so $m_a^X = 0$. Without any observations of action a it is hard to learn which of the two actions is better. Only understanding how good or how bad action a' is does not help in understanding whether action a' is better or worse than action a. Hence, a non-guessing rule will not change the weight put on action a, so $f(X)_a = q_a$ if $m_a^X = 0$. Now consider the case where action a would never be chosen without the data, so $q_a = 0$. To change the weight on action a after observing the data means to increase the probability of choosing action a at the expense of decreasing the probability of choosing some other action. This may only be done if one is sure that action a is better than some other action. However, no data set can reveal with certainty that this is true. Hence, the understanding about action a will not change if $q_a = 0$.

Given these insights it seems hard to learn from data without guessing. However, as we see below, this is possible when $0 < q_a < 1$. One can increase the weight on action a

when action a looks better than action a' and decrease it when action a looks worse than action a'. When done appropriately, one can move on average (as weighted by the true underlying distribution) in the right direction.

Next we identify distributions under which the decision-maker's understanding may not change in expectation. This holds true for distributions in which all actions are equally good. Note that non-guessing definition has no bite when all actions are equally good. The proof of this finding utilizes continuity arguments.

Proposition 5 *Assume that f is a non-guessing rule under q. If G is such that $\mu_a^G = \mu_{a'}^G$ for all $a, a' \in A$ then $E_G\left(f\left(\bar{X}\right)\right) = q$.*

Proof (of Proposition 5). Assume that f is a non-guessing rule. Consider some G such that $\mu_{a'}^G = \mu_{a''}^G$ for all $a', a'' \in A$ and $E_G\left(f\left(\bar{X}\right)\right)_a > q_a$. Then we can change G slightly to $\hat{G}$ such that $\mu_a^{\hat{G}} < \mu_{a'}^{\hat{G}} = \mu_{a''}^{\hat{G}}$ for all $a, a'' \in A \backslash \{a\}$. Yet, as $E_G\left(f\left(\bar{X}\right)\right)_a$ is continuous in G, we still have $E_{\hat{G}}\left(f\left(\bar{X}\right)\right)_a > q_a$ (provided $\hat{G}$ is sufficiently close to G) which contracts the fact that f is a non-guessing rule. ■

An immediate implication is that a decision-maker using a non-guessing rule chooses the pseudo prior whenever all payoffs in the data set are equal.

Corollary 6 *If f is a non-guessing rule under q then $f\left((a_i, y)_{i=1}^m\right) = q$ for all $y \in [0, 1]$.*

In particular, this shows that non-guessing rules sometimes randomize whenever the pseudo prior does not put all mass on a single action.

Given these insights we can now answer whether Bayesians can be caught guessing. To qualify the behavior of a Bayesian it is natural to let the pseudo prior q be their choice under their prior when there is no data set. In the following arguments we also allow for more general pseudo priors as one might wonder what happens if, when starting from some default choice q, the decision is delegated to a Bayesian. We say that a decision maker is responsive to the data set if there is some data set for which they choose something different from the pseudo prior. It turns out that any Bayesian who is response to the data set can typically be caught guessing. We first explain intuitively and then formally. This statement follows immediately from Proposition 4 if the pseudo prior is deterministic. If the pseudo prior is not deterministic then Corollary 6 implies that the Bayesian has to randomize after some data sets. However, as we show below, Bayesians typically do not randomize where typically is given in a well defined sense.

More formally, we say that Bayesians generically can be caught guessing under q if the set of all priors under which any SEU maximizing rule can be caught guessing is dense in $\mathcal{G}$ while its complement is nowhere dense in $\mathcal{G}$. So to find a SEU maximizer who uses a non-guessing rule is an exception. There will be arbitrarily close priors where all SEU maximizers can be caught guessing. On the other hand, whenever there is a prior under which it is not possible to be both SEU maximizing and non-guessing then this property will also hold for any nearby prior.

Corollary 7 *Any responsive Bayesian generically can be caught guessing under q for any $q \in \Delta A$.*

Proof. Given the arguments in the text above we only need to consider a non deterministic pseudo prior q. Let $a', a'' \in A$ be such $q_{a'}, q_{a''} > 0$. Following Proposition 5 a non-guessing rule randomizes between a' and a'' when all payoffs in the data set are equal. In the following we show that Bayesians generically will not randomize between a' and a'' for any data set.

Let $Q'(X|Q)$ be the posterior given prior Q and data set X. Let $P(X|G)$ be the probability of observing data set X under joint distribution $G \in \mathcal{G}$. Then

$$\frac{Q'(X|Q)(\bar{G})}{Q'(X|Q)(G)} = \frac{P(X|\bar{G})\, Q(\bar{G})}{P(X|G)\, Q(G)}.$$

Let $\bar{Q}$ be the prior of the Bayesian decision-maker and let $\hat{Q} = (1-\varepsilon)\bar{Q} + \varepsilon\left[\hat{G}\right]$ for some $\hat{G} \in \mathcal{G}$ where $\left[\hat{G}\right] \in \mathcal{G}$ is defined by $\left[\hat{G}\right]\left(\hat{G}\right) = 1$. Then for all $G, \bar{G} \in \mathcal{G} \backslash \left\{\hat{G}\right\}$ we have

$$\frac{Q'\left(X|\hat{Q}\right)(G)}{Q'\left(X|\hat{Q}\right)(\bar{G})} = \frac{Q'(X|\bar{Q})(G)}{Q'(X|\bar{Q})(\bar{G})}$$

and for $G \neq \hat{G}$ we have

$$\begin{aligned}\frac{Q'\left(X|\hat{Q}\right)\left(\hat{G}\right)}{Q'\left(X|\hat{Q}\right)(G)} &= \frac{P\left(X|\hat{G}\right)\left((1-\varepsilon)\bar{Q}\left(\hat{G}\right)+\varepsilon\right)}{P(X|G)(1-\varepsilon)\bar{Q}(G)} = \frac{P\left(X|\hat{G}\right)\bar{Q}\left(\hat{G}\right)}{P(X|G)\bar{Q}(G)}\frac{(1-\varepsilon)\bar{Q}\left(\hat{G}\right)+\varepsilon}{(1-\varepsilon)\bar{Q}\left(\hat{G}\right)} \\ &> \frac{Q'(X|\bar{Q})\left(\hat{G}\right)}{Q'(X|\bar{Q})(G)}.\end{aligned}$$

So $Q'\left(X|\hat{Q}\right) = (1-\lambda)\, Q'(X|\bar{Q}) + \lambda\left[\hat{G}\right]$ for some $\lambda > 0$.

Let $\mu_a^Q = \int \mu_a^G dQ(G)$ be the expected payoff of action $a \in A$ under prior Q. Assume that $\bar{Q}$ is such that $\mu_{a'}^{Q'(X|\bar{Q})} = \mu_{a''}^{Q'(X|\bar{Q})}$ for some $a', a'' \in A$ with $a' \neq a''$. Consider any $\hat{G}$ such that $\mu_{a'}^{\hat{G}} > \mu_{a''}^{\hat{G}}$. Then $\mu_a^{Q'(X|\hat{Q})} = (1-\lambda)\mu_a^{Q'(X|\bar{Q})} + \lambda\mu_a^{\hat{G}}$ for $a \in A$ and consequently $\mu_{a'}^{Q'(X|\hat{Q})} > \mu_{a''}^{Q'(X|\hat{Q})}$. Thus the Bayesian with prior $\hat{Q}$ which is arbitrarily close to $\bar{Q}$ will not be indifferent between actions a' and a''.

The above shows that the set of priors under which any SEU maximizing rule under this prior have some guessing is dense. We now show that its compliment is nowhere dense. Consider some open neighborhood $\mathcal{N}$ in $\mathcal{G}$. Let $\bar{Q} \in \mathcal{N}$. Then $\hat{Q} \in \mathcal{N}$ if ε is sufficiently small. As $\mu_{a'}^{Q'(X|\hat{Q})} > \mu_{a''}^{Q'(X|\hat{Q})}$, by continuity of Bayesian updating, there will be a neighborhood $\mathcal{N}_0$ of $\hat{Q}$ such that $\mu_{a'}^{Q'(X|Q)} > \mu_{a''}^{Q'(X|Q)}$ holds for all $Q \in \mathcal{N}_0$. Consequently, the non-guessing rules are not dense in $\mathcal{N}_0$ and hence not dense in $\mathcal{N}$. ■

Our third insight is that any non-guessing rule has to be continuous and monotone in the payoffs. In particular, the decision-maker should not "jump to conclusions" whenever some payoff in the data set changes slightly.

Proposition 8 *Assume that f is a non-guessing rule. Then $f\left(\{(a_j, x_j)\}_{j=1}^m\right)$ is continuous in x_i and $f\left(\{(a_j, x_j)\}_{j=1}^m\right)_{a_i}$ is weakly monotone increasing in x_i for all $i = 1, .., m$.*

Proof. Assume that f is a non-guessing rule. We wish to show that f has to be continuous in x_i by contradiction. Consider data set X in which $x_i < 1$ where $i \in \{1, .., m\}$. Assume that $f(X)_{a_i}$ makes a discontinuous jump when x_i slightly increases in the data set X. Take a distribution G such that X occurs with positive probability, no action yields a deterministic payoff and all actions are equally good. Such distributions exist. By replacing x_i by $x_i + \varepsilon$ and slightly changing the weights on some other payoffs in the support of G we can make it such that all μ_a remain all equal. As f makes a discontinuous jump when x_i slightly increases, so does $E_G\left(f\left(\bar{X}\right)\right)_{a_i}$ in the same direction. On the other hand, as all means stay the same, $E_G\left(f\left(\bar{X}\right)\right)_{a_i} = q_{a_i}$ has to remain unchanged by Proposition 5 which is a contradiction to f has to be continuous.

We now prove weak monotonicity. Consider a data set X and let (a, x_1) be the first data point in this set. Assume that $x_1 < 1$. We will show that f_a has to be weakly monotone increasing in x_1. Let $G \in \mathcal{G}$ be such that $\mu_{a'}^G = \frac{1}{2}$ for all $a' \in A$ and X occurs with positive probability. So $G_a(\{x_1\}) > 0$. As G has finite support, we can find $\varepsilon > 0$ sufficiently small so that $G_a(\{x_1 + \varepsilon\}) = 0$. Let $\hat{X}$ be defined by replacing (a, x_1) by $(a, x_1 + \varepsilon)$.

We now move some mass that G_a puts on x_1 to $x_1 + \varepsilon$ to obtain distribution $\hat{G}$. Let γ be the mass moved. So γ is the probability that $\hat{G}_a$ puts on $x_1 + \varepsilon$. Choose γ sufficiently small. Then the distributions of data sets under G and $\hat{G}$ are essentially identical apart from the probability γ under which data set X occurs under G while data set $\hat{X}$ occurs under $\hat{G}$. As γ is sufficiently small, all other data sets that cannot occur under both G and $\hat{G}$ can be ignored.

Assume $f(X)_a > f\left(\hat{X}\right)_a$ which violates weak monotonicity of f_a. Then $E_G\left(f\left(\bar{X}\right)\right)_a > E_{\hat{G}}\left(f\left(\bar{X}\right)\right)_a$ holds when γ is sufficiently small. As $\mu_{a'}^G = \frac{1}{2}$ for all $a' \in A$, by Lemma 5 this means that $E_G\left(f\left(\bar{X}\right)\right)_a = q_a > E_{\hat{G}}\left(f\left(\bar{X}\right)\right)_a$. On the other hand, as f is a non-guessing rule, and $\mu_{a'}^{\hat{G}} = \frac{1}{2} < \mu_a$ for $a' \neq a$ we have $E_{\hat{G}}\left(f\left(\bar{X}\right)\right)_a \geq q_a$, which is a contradiction. ■

A direct consequence of Proposition 8 is that the ability randomize with any intensity is an integral part of any responsive non-guessing rule. More specifically, consider a non-guessing rule that chooses action a with probability λ_0 for some data set and with probability λ_1 for some other data set with $\lambda_0 < \lambda_1$. Then for any $\lambda \in (\lambda_0, \lambda_1)$ there exists a data set for which the rule chooses action a with probability λ. This follows with continuity from the intermediate value theorem. In particular this shows why rES can be caught guessing under any q. Similarly, any other frequentist rule that estimates which action is best and only randomizes when there are ties can be caught guessing under any q.

Finally we show how one can sequentially combine two non-guessing rules that each have access to their own data set into a new non-guessing rule for the combined data set. The idea is to sequentially apply one rule after the other, using the output of the first rule as pseudo prior for the second. This is reminiscent of SEU maximization where the posterior can be used as a prior when facing the next data set.

Proposition 9 *Let $t \in \{2, ..., m-2\}$ and let $f, \bar{f}, \hat{f}$ be rules such that*

$$\hat{f}\left(\{(a_i, x_i)\}_{i=1}^m\right) = \bar{f}\left(\{(a_i, x_i)\}_{i=t+1}^m\right) \text{ for all } \{x_i\}_{i=1}^m \in [0,1]^m .$$

If f is a non-guessing rule under pseudo prior q and $\bar{f}$ is a non guessing rule under pseudo prior $f\left(\{(a_i, x_i)\}_{i=1}^t\right)$ then $\hat{f}$ is a non-guessing rule under pseudo prior q.

The proof follows from the definitions. Note that it is important that the intermediate round t does not depend on payoffs in the data set. Later we will provide some examples to evaluate the effectiveness of this sequential combination.

4 Linear Rules

Given the insights from the previous section it is not clear where to start looking for non trivial non-guessing rules as both SEU maximization and frequentist rules have been ruled out. Moreover the rule has to be very sensitive to changes in payoffs. The set of possible rules is infinitely dimensional so there are many ways of introducing this sensitivity. On the other hand we have a powerful identity that has to hold whenever all actions are equally good. At this point the search would be a lost cause were it not for the randomization trick. This is a method that allows to limit attention to binary valued data sets and linear rules, reducing the complexity of the problem enormously. Together with the mentioned identity we will be able to identify linear non-guessing rules. We also comment on the further value of linear rules.

Definition 10 *The rule f is* linear *if $f\left(\left(a_i,x_i\right)_{i=1}^m\right)$ is linear (or, more precisely, affine) in x_j for each $j\in\{1,..,m\}$, so for each $a\in A$ and $j\in\{1,..,m\}$ there exist functions $\gamma_j\left(x_{-j}\right)$ and $\delta_j\left(x_{-j}\right)$ such that $f\left(\left(a_i,x_i\right)_{i=1}^m\right)_a=\gamma_j\left(x_{-j}\right)x_j+\delta_j\left(x_{-j}\right)$.*

Linear functions are arguably the simplest continuous functions. Following Proposition 8, non-guessing rules have to be continuous. It is easy to see that any linear rule f satisfies

$$f\left(\left(a_i,x_i\right)_{i=1}^m\right)=\sum_{j_1=0}^{1}\cdots\sum_{j_m=0}^{1}\left(\prod_{i=1}^{m}x_i^{j_i}\left(1-x_i\right)^{1-j_i}\right)f\left(\left(a_i,j_i\right)_{i=1}^m\right).$$

In particular, linear rules are uniquely defined by their values for binary valued data sets. In fact, this equality shows how one can construct a linear rule f for data sets with payoffs in $[0,1]$ when starting with a rule f^0 that is defined on binary valued data sets. For $i=1,..,m$, independently randomly transform the ith payoff x_i into 1 with probability x_i and into 0 with probability $1-x_i$. After transforming all payoffs one obtains a binary valued data set to which then we evaluate the rule f^0. The rule f is then set equal to the expected choices under this construction, the formula is given by the equation above when replacing f on the right hand side by f^0.The construction of a linear rule from a rule for binary valued data is called the randomization trick and goes back to (Schlag, 2003). The value of this trick is that this construction nicely uncovers how properties of the rule f^0 for binary valued data automatically carry over to data where payoffs are in $[0,1]$.

The unique linear rule f for payoffs belonging to $[0,1]$ that emerges from this construction is called the linear extension of f^0. Note that one can also evaluate the linear extension of a rule that is defined for payoffs in $[0,1]$ by limiting this rule to binary valued payoffs and then creating the linear rule. This mapping is of course the identity mapping when the original rule was already linear.

Proposition 11 *If f^0 is a non-guessing rule under q for distributions with payoffs in $\{0,1\}$ then the linear extension of f^0 is a non-guessing rule under q when facing distributions with payoffs in* $[0,1]$.

This result will be used in the next section to find non-guessing rules by looking at the much simpler setting where payoffs are binary valued and thereby identifying linear non-guessing rules when payoffs are in $[0,1]$.

We point out three nice properties of linear rules. First of all, linear functions are easy to evaluate as due to linearity,

$$E_G f\left(\{(a_i,x_i)\}_{i=1}^m\right) = f\left(\left\{\left(a_i,\mu_{a_i}^G\right)\right\}_{i=1}^m\right) = \sum_{j_1=0}^{1} \cdots \sum_{j_m=0}^{1} \left(\prod_{i=1}^{m} \left(\mu_{a_i}^G\right)^{j_i} \left(1-\mu_{a_i}^G\right)^{1-j_i}\right) f\left((a_i,j_i)_{i=1}^m\right).$$

Second of all, linear non-guessing rules can be most preferred among the set of all non-guessing rules. The advantage of being linear emerges when selecting among non-guessing rules using a worst case approach. One might wish to minimax regret among the non-guessing rules, where regret is given by loss of not having chosen the best action. Alternatively, one might wish to maximize the minimal probability of choosing the best action when the means of the actions are not too similar (as in Section 5.4). In the most generality consider selecting among the non-guessing rule one that minimizes a maximum loss. Specifically, let $l(G,p) \in \mathbb{R}$ be the loss of choosing $p \in \Delta A$ given distribution G and assume that f is weakly preferred to $\bar{f}$ if

$$\sup_{G \in \mathcal{G}} l\left(G, E_G\left(f\left(\bar{X}\right)\right)\right) \leq \sup_{G \in \mathcal{G}} l\left(G, E_G\left(\bar{f}\left(\bar{X}\right)\right)\right).$$

It then follows with Proposition 11 that the linear extension of a non-guessing rule is a non-guessing rule that is weakly preferred to the original rule. When computing the maximum loss using the linear extension it is like maximizing the original expression only over the binary valued distributions. This shows that linear rules exist among the most preferred non-guessing rules under such worst case scenario preferences.

The third property refers to averaging. The availability of some data only in terms of averages often emerges when using data from sources of others. Averaging is also a way to deal with data sets that consist of independent groups of dependent observations by taking averages among those that are dependent. Linear non-guessing rules remain non-guessing when some of the payoffs in the data set are themselves averages.

More formally, let $G_a^{(t)}$ be the distribution of the average of t identically distributed random variables with distribution G_a, so $G_a^{(t)}$ is the distribution of $\frac{1}{t}\left(Z_a^{(1)}+...+Z_a^{(t)}\right)$ where $Z_a^{(j)}, j=1,..,t$ are (possibly non independent) copies of Z_a. Given $k \in N^m$ let X^k be a data set in which payoff z_j is independently drawn from $G_a^{(k_j)}$. So $X^k = X$ when $k_j = 1$ for all $j = 1,..,m$. We say that X^k is a data set that allows for averages.

Proposition 12 *If f is a linear non-guessing rule then f is also a non-guessing rule when the data set allows for averages.*

The statement follows immediately by observing that $\frac{1}{t}\left(Z_a^{(1)}+...+Z_a^{(t)}\right)$ has mean μ_a^G. As the performance of a linear rule only depends on the underlying mean of each observed payoff it does not matter how the payoffs are otherwise distributed.

5 Data Sets with Only Two Different Actions

Given the last two sections it now seems feasible to be able to identify non-guessing rules. Yet as we wish to consider any data set and select among the non-guessing rules tractability is still a concern. So we limit attention to data sets that contain only two actions. Given Proposition 4 we can assume without loss of generality that $n = 2$ and $m_a, m_{a'} \geq 1$ (so $m \geq 2$).

5.1 A "Best" Rule

Given the findings in Section 4 we search for non-guessing rules among the linear rules. It turns that there is a dominant linear non-guessing rule. Being dominant, for any given vector of the mean payoffs, it maximizes the minimal expected payoff, minimizes the maximal regret and maximizes the minimal probability placed on the best actions.

We call $\bar{f}$ a *dominant rule among the rules in the set* $\bar{\mathcal{F}}$ if $\bar{f} \in \bar{\mathcal{F}}$ and $\bar{f}$ achieves higher expected payoffs than any other rule f in $\bar{\mathcal{F}}$ under any distribution G, so if $E_G\left(\bar{f}\left(\bar{X}\right)\right) \cdot \mu^G \geq E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G$ for $f \in \bar{\mathcal{F}}$ and $G \in \mathcal{G}$.

Linear rules can be characterized by their behavior when data is binary valued. Cutoff rules for binary valued data play a central role in our main characterization. We say that a rule f^0 for binary valued data is a *cutoff rule* if for each $r \in \{0, .., m\}$ there exists $a \in A$, $s_r \in \{0, .., m_a\}$ and $\theta_r \in (0, 1]$ such that $f_a^0 = 1$ if $r_a > s_r$, $f_a^0 = \theta_r$ if $r_a = s_r$ and $f_a^0 = 0$ if $r_a < s_r$. Loosely speaking, choose action a if it yielded strictly more successes that a given threshold, choose the other action if it yielded strictly less and mix appropriately at the threshold.

Theorem 13 *Given $A = \{a, a'\}$ and q such that $q_a, q_{a'} > 0$ consider the cutoff rule f^{0*} for binary valued data defined as follows. Let $s_0 = 0$, $s_m = m_a$ and $\theta_0 = \theta_m = q_a$. For $r \in \{1, .., m-1\}$ let s_r with* $\max\{0, r - m_{a'}\} \leq s_r \leq \min\{r, m_a\}$ *and $\theta_r \in (0, 1]$ be such that*

$$\frac{1}{\binom{m}{r}} \left(\binom{m_a}{s} \binom{m_{a'}}{r-s} \theta_r + \sum_{i=s+1}^{\min\{r,m_a\}} \binom{m_a}{i} \binom{m_{a'}}{r-i} \right) = q_a, \tag{2}$$

so

$$\frac{1}{\binom{m}{r}} \sum_{i=s+1}^{\min\{r,m_a\}} \binom{m_a}{i} \binom{m_{a'}}{r-i} < q_a \leq \frac{1}{\binom{m}{r}} \sum_{i=s}^{\min\{r,m_a\}} \binom{m_a}{i} \binom{m_{a'}}{r-i}$$

and

$$\theta_r = \frac{\binom{m}{r}}{\binom{m_a}{s}\binom{m_{a'}}{r-s}} \left(q_a - \frac{1}{\binom{m}{r}} \sum_{i=s+1}^{\min\{r,m_a\}} \binom{m_a}{i} \binom{m_{a'}}{r-i} \right).$$

Then the randomized extension f^* of the rule f^{0*} is a dominant rule among the linear non-guessing rules and it is the unique rule with this property.

The proof of this result is in the appendix. At the end of this section we discuss the limitations introduced by limiting attention to linear rules. We proceed now by providing some intuition behind the construction of the dominant linear non-guessing rule. Following Proposition 11, all we need to do is to find a non-guessing rule for Bernoulli distributions. A rule for general distributions that inherits its properties is then created by using the randomization trick. So let us look at the rule f^{0*} for binary valued data sets. As we wish to choose action a when action a is better, it is natural to choose a when there are sufficiently many more successes of a than of a' in the data. The threshold is chosen so that, following Proposition 5, the non-guessing rule selects action a with probability q_a whenever both actions are equally good. This condition has to hold when $\mu_a = \mu_{a'}$ is small and there are most likely to be few successes in the data set as well as when $\mu_a = \mu_{a'}$ is large and there are many successes in the data set. To be able to condition on the total number of successes makes it possible to guarantee this condition across all values of $\mu_a = \mu_{a'}$. The equality (2) emerges directly from this condition. In Section 8 we give more insights by establishing a connection to unbiased hypothesis tests.

We further the understanding of the dominant linear non-guessing rule by presenting some comparative statics. We first consider properties defined by f^{0*} that govern how the dominant linear rule deals with the transformed data set. Equivalently, these are properties of the dominant linear rule in binary data sets. Specifically we present comparative statics on the cutoff s_r. We do this for large samples where it is easy to verify using the central limit theorem that the cutoff s_r approximately satisfies

$$q_a \approx 1 - \Phi\left(\frac{s_r}{r}; p, \sqrt{\frac{p(1-p)}{r}\left(1-\frac{r}{m}\right)}\right) \text{ for } 1 \leq r \leq m-1,$$

where $p = \frac{m_a}{m}$ and $\Phi(\cdot; \mu, \sigma)$ is the cdf of the Normal distribution with mean μ and variance σ^2.[5] Given this approximation we make the following three observations. (i) The cutoff is decreasing in the weight put on action a. The more convinced the decision maker is that a is the better action the more likely they will also choose that action. (ii) Note that $\frac{r_a}{r} \geq p$ if and only if $\frac{r_a}{m_a} \geq \frac{r_{a'}}{m_{a'}}$. Moreover, $\left|\frac{s_r}{r} - p\right|$ decreases when $\frac{p(1-p)}{r}\left(1 - \frac{r}{m}\right)$ decreases. Hence, if there are sufficiently many successes in the transformed data set or the sample is sufficiently unbalanced then action a is chosen if it yielded disproportionally more often a success than action a'. (iii) The more unbalanced the data set or the more successes there are in the transformed data set, the closer the threshold proportion of successes $\frac{s_r}{r}$ is to the frequency $\frac{m_a}{m}$ with which action a is observed in the data set.

Next we provide some comparative statics for how the dominant linear rule deals with the original data set. Choice probabilities are continuous in the payoffs of the original data set. We know this from Proposition 8 as the rule is non-guessing, here we point out that it emerges when applying the randomization trick. In fact, non random choices

[5]Let $X_i \in \{0,1\}$ such that $X_i = 1$ if the ith draw without replacement from an urn with r values 1 and $m-r$ values 0. Then for $j \neq i$ and $p = \frac{m_a}{m}$,

$$\begin{aligned} Cov(X_i, X_j) &= \frac{m_a}{m}\frac{m_a-1}{m-1}\left(1-\frac{m_a}{m}\right)^2 + \left(\frac{m_a}{m}\frac{m_{a'}}{m-1} + \frac{m_{a'}}{m}\frac{m_a}{m-1}\right)\left(1-\frac{m_a}{m}\right)\left(-\frac{m_a}{m}\right) \\ &\quad + \frac{m_{a'}}{m}\frac{m_{a'}-1}{m-1}\left(\frac{m_a}{m}\right)^2 \\ &= -(1-p)\frac{p}{m-1} \end{aligned}$$

and

$$Var\sum_{i=1}^{r} X_i = r * p(1-p) + r(r-1)\left(-(1-p)\frac{p}{m-1}\right) = rp(1-p)\left(1-\frac{r}{m}\right).$$

only emerge whenever payoffs 0 and 1 do not both occur in the data set. For example, assume that payoff 0 does not occur. Then the randomization trick transforms the data set with strictly positive probability into a one in which all payoffs are equal to 1, upon which the rule chooses q. So whenever the decision-maker is initially open to choosing either action (as $q_a \in (0,1)$), after observing a data set in which both extreme payoffs do not occur they remain open. Note also, following Proposition 8, that the weight placed by the dominant linear rule on any action is weakly monotone increasing in each of the payoffs in the data set that are associated to this action. This is however difficult to see from Theorem 13 as it would require a more detailed understanding of how the cutoffs $\{s_r\}_{r=0}^{m}$ compare to each other.

We proceed by presenting two corollaries of Theorem 13. The first relates to symmetry in labels. As the dominant linear non-guessing rule is unique the rule must be symmetric when the setting is completely symmetric, so when $q_a = \frac{1}{2}$ and $m_a = m_{a'}$. In that case the cutoff rule f^{0*} is defined by $s_r = \frac{r}{2}$ and $\theta_r = \frac{1}{2}$ if r is even and $s_r = \frac{r+1}{2}$ and $\theta_r = 1$ if r is odd. We see that this is rES. In fact, we can verify easily (from the equations in Theorem 13) that the cutoff rule f^{0*} is not equal to rES otherwise. We summarize these insights.

Corollary 14 *The randomized extension of rES is the dominant linear non-guessing rule if and only* $q_a = \frac{1}{2}$ *and* $m_a = m_{a'}$*.*

Following Schlag (2006b) we know that rES attains minimax regret in balanced samples. Thus, the result above shows that the dominant linear non-guessing rule attains minimax regret in the completely symmetric setting where the pseudo prior is uninformative and the sample is balanced. Note that there are no results in the literature on minimax regret for unbalanced samples.[6]

We use the insight that the randomized extension of rES is a non-guessing rule to show how the sequential combination of rules (see Proposition 9) can lead to inefficiencies. To see this, consider $q_a = \frac{1}{2}$ and the data set $((a,1),(a',0))$. Then the dominant rule selects action a and hence puts no weight on action a'. Consequently, when combining this rule with the dominant rule for the data set $((a,0),(a',1))$, action a' still receives no weight by Proposition 4. However, the dominant rule for the combined data set puts equal weight on

[6]It is straightforward to extend the findings of Manski (2004) to obtain bounds on maximal regret of BAR for unbalanced samples. Exact results as in Schlag (2006b) for unbalanced samples are hard to come by as the proofs require guesssing the Nash equilbrium of a zero sum game.

each action. As the dominant rule is unique it follows that it is not possible to efficiently evaluate these two data sets separately.

The second corollary contains a simple insight on the value of having more data.

Corollary 15 *The expected payoffs achieved by the dominant linear non-guessing rule weakly increase as more observations are added to the data set.*

The corollary follows from Theorem 13 as the dominant linear non-guessing rule for the smaller data set is always feasible in the larger data set. So the property of being dominant implies the statement. We however hasten to point out two things. First of all, more data need not lead to a strict increase. For example, below we show that $E_G\left(f^*\left(\bar{X}\right)\right)\cdot\mu^G$ does not change when adding one more observation given $m_a = m_{a'} = 1$ when $q_a = \frac{1}{2}$. Second of all, this corollary does not reveal any information on whether the dominant linear non-guessing rule will learn what is best in the long run. We will deal with this topic separately in Section 5.3 below.

Finally, note that Theorem 13 is far from being a characterization of all non-guessing rules. In particular, it is an open question whether the dominant linear non-guessing rules are admissible among the set of non-guessing rules. Clearly their property of being dominant among the linear rules means they are undominated when the underlying distributions are binary valued. It can be shown that they dominate all other non-guessing rules when $m_a = m_{a'} = 1$ (see more in Section 5.2). However, this is no longer necessarily true when $m_a, m_{a'} \geq 1$ and $m \geq 3$ (see Section 7).The value of the above characterization lies in the ability to identify rules for any m and, due to linearity, to be able to quantify their performance as done later in Section 5.4. Moreover, as shown in Section 5.4, linearity is not an obstacle when searching for rules that are good at learning which action is best.

5.2 Some Examples

The dominant linear non-guessing rule for $m_a = m_{a'} = 1$ is simple. We have $s_1 = 1$ and $\theta_1 = 2q_a$ if $q_a \leq \frac{1}{2}$ and $s_1 = 0$ and $\theta_1 = 2q_a - 1$ if $q_a \geq \frac{1}{2}$. Consider how this rule behaves after the randomization trick has been applied when $q_a \leq \frac{1}{2}$. If action a' yields a success while action a a failure then action a' is chosen. However, vice versa, if action a gave a success and action a' gave a failure then action a is only chosen with probability

$2q_a$ and action a' is chosen with probability $1-2q_a$. So the fact that $q_a \leq \frac{1}{2}$ leads to trusting evidence that action a' looks better while being more suspicious when action a looks better. It follows that

$$f^*\left(\left\{(a, x_a), (a', x_{a'})\right\}\right)_a = q_a + \min\left\{q_a, q_{a'}\right\}(x_a - x_{a'})$$

and hence $E_G f^*\left(\left\{(a, Z_a), (a', Z_{a'})\right\}\right)_a = q_a + \min\left\{q_a, q_{a'}\right\}\left(\mu_a^G - \mu_{a'}^G\right)$. In particular, dominant linear non-guessing rule for $m_a = m_{a'} = 1$ performs strictly better than q whenever both actions are not equally good. It strictly improves the decision-maker's understanding of which action is better as compared to q. As mentioned above, this rule dominates all other non-guessing rules, not only the linear ones. The proof is beyond the scope of this paper but it is essentially identical to the one provided by Schlag (1998) on dominant improving rules for imitation in a population.

With these insights we now uncover a value of non-guessing rules. Consider any data set in which each action appears at least once. To apply the dominant rule for $m_a = m_{a'} = 1$ to the first time each action appears in this data set constitutes a non-guessing rule for this data set. As the dominant linear non-guessing rules dominates all other linear non-guessing rules, it also dominates (performs better than) this rule. Combining this insight with the learning abilities of the dominant rule for $m_a = m_{a'} = 1$ shown above we obtain the following. The dominant linear rule under q performs strictly better than when choosing q whenever both means are not equal and each action appears at least once in the data set. In some sense, the decision-maker is expected to have a better understanding of the decision problem whenever the two actions are not equally good. We summarize this insight formally.

Remark 16 *$E_G f^*\left(\bar{X}\right) \cdot \mu^G > q \cdot \mu^G$ hold for G such that $\mu_a^G \neq \mu_{a'}^G$ whenever $m_a, m_{a'} \geq 1$.*

Consider now the case where $m = 3$ and $m_a = 1$. If $q_a \leq \frac{1}{3}$ then $s_1 = 1$ and $\theta_1 = 3q_a$, if $q_a > \frac{1}{3}$ then $s_1 = 0$ and $\theta_1 = \frac{3}{2}q_a - \frac{1}{2}$. If $q_a \leq \frac{2}{3}$ then $s_2 = 1$ and $\theta_2 = \frac{3}{2}q_a$, if $q_a \geq \frac{2}{3}$ then $s_2 = 0$ and $\theta_2 = 3q_a - 2$. Together this yields

$$f\left(\left\{(a, x_1), (a', x_2), (a', x_3)\right\}\right)_a = \begin{cases} q_a + \frac{1}{2}q_a\left(4x_1 - (2 + x_1)(x_2 + x_3) + 2x_2x_3\right) & \text{if } q_a \leq \frac{1}{3} \\ q_a - (1 - q_a + x_1)\frac{x_2 + x_3}{2} + x_2x_3(1 - 2q_a) & \\ \quad + \left(1 - (1 - (x_2 + x_3))q_a\right)x_1 & \text{if } \frac{1}{3} < q_a < \frac{2}{3} \\ q_a + \frac{1}{2}(1 - q_a)\left(2x_1 - (1 - x_1)(x_2 + x_3) - 2x_2x_3\right) & \text{if } q_a \geq \frac{2}{3}. \end{cases}$$

It follows that

$$E_G\left(f\left(\left\{\left(a,Z_a\right),\left(a',Z_{a'}\right),\left(a',Z_{a'}\right)\right\}\right)\right)_a=\begin{cases} q_a+q_a\left(2-\mu_{a'}\right)\left(\mu_a-\mu_{a'}\right) & \text{if } q_a\leq\frac{1}{3} \\ q_a+\left(1-\mu_{a'}-q_a+2q_a\mu_{a'}\right)\left(\mu_a-\mu_{a'}\right) & \text{if } \frac{1}{3}<q_a<\frac{2}{3} \\ q_a+\left(1-q_a\right)\left(1+\mu_{a'}\right)\left(\mu_a-\mu_{a'}\right) & \text{if } q_a\geq\frac{2}{3}. \end{cases}$$

The insights of Schlag (1999) can be used to show that there is no dominant rule among all rules when $m=3$ and $q_{a''}=\frac{m_{a''}}{m}$ for $a''\in A$ (as explained in Section 7).

We compare $m=3$ to $m=2$. We observe that the performance strictly improves when $\mu_a\neq\mu_{a'}$ when $q_a\notin\left\{0,\frac{1}{2},1\right\}$. However, when $q_a=\frac{1}{2}$ then $f_a=\frac{1}{2}+\frac{1}{2}\left(x_1-\frac{1}{2}x_2-\frac{1}{2}x_3\right)$ and $E\left(f\left(\bar{X}\right)\right)_a=\frac{1}{2}+\frac{1}{2}\left(\mu_a-\mu_{a'}\right)$. In the latter case the performance remains unchanged when adding one more data point to a balanced data set with one observation for each action. When we add yet another observation of action a', so $m=4$ and $m_a=1$, then we obtain $E\left(f\left(\bar{X}\right)\right)_a=\frac{1}{2}+\frac{1}{2}\left(1+\mu_{a'}-\mu_{a'}^2\right)\left(\mu_a-\mu_{a'}\right)$ if $q_a=\frac{1}{2}$. Now we find superior performance as compared to $m_a=m_{a'}=1$ when $\mu_a\neq\mu_{a'}$ and $\mu_{a'}\neq 0$.

In the appendix we present some values of s and θ in small data sets.

5.3 Performance in Large Samples

The performance of the dominant linear non-guessing rule increases as the data set gets larger. We now show that this will lead the decision-maker to choose the best action in the long run when the data set contains sufficiently many observations of each action. In fact, the minimal probability of choosing the best action will be arbitrary large for any given minimal distance between the two means.

For G with $\mu_a^G\neq\mu_{a'}^G$ let $a^{G*}\in A$ be such that $\mu_{a^{G*}}^G=\max\left\{\mu_a^G,\mu_{a'}^G\right\}$.

Proposition 17 *Let f^* be the dominant linear non-guessing rule.*

(i) For every $d>0$ and $\psi<1$ there exists $\bar{m}\in\mathbb{N}$ such that if $m_a,m_{a'}\geq\bar{m}$ and $\left|\mu_a^G-\mu_{a'}^G\right|\geq d$ then $E_G\left(f^*\left(\bar{X}\right)\right)_{a^{G*}}\geq\psi$.

(ii) For ever $\varepsilon>0$ there exists $\bar{m}\in\mathbb{N}$ such that if $m_a,m_{a'}\geq\bar{m}$ then $E_G\left(f^*\left(\bar{X}\right)\right)\cdot\mu^G\geq\max\left\{\mu_a^G,\mu_{a'}^G\right\}-\varepsilon$ for all $G\in\mathcal{G}$.

Note that part (ii) follows immediately from part (i). Note also that the value of $\bar{m}$ will depend on q. Any rule that is non-guessing when $q_a=1$ chooses action a regardless of the data set. Continuity arguments imply that behavior will be very similar when q_a is very close to 1. As a consequence, learning can be arbitrarily slow when q is close to the boundaries.

The proof of Proposition 17 is intuitive albeit a bit awkward. Although we know the exact formula for the dominant linear non-guessing rule we are not able to derive properties of its expected value unless the data set is balanced and the pseudo prior is symmetric.

Proof. Assume that $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$. We will first prove the claim when $m_a = m_{a'}$ and $q_a = \frac{1}{2}$. Given the randomization trick we only have to consider binary valued distributions. The claim follows easily. If m_a and $m_{a'}$ are sufficiently large then with arbitrarily high probability the better action will yield strictly more successes and will be chosen by the best-average rule. Here the property that $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$ ensures that convergence is uniform in the sense that how large m_a and $m_{a'}$ have to be only depends on d and not on the specific G. The minimal number of times each action should be contained in the data set can be set without knowing the specific values of μ_a^G and $\mu_{a'}^G$. Note that this argument extends immediately when $m_a \neq m_{a'}$ as the dominant linear non-guessing rule will perform better than the rule that first drops observations from the action sampled more often to make the sample balanced and then applying the dominant linear non-guessing rule for balanced sample.

Next we consider $q_a \neq \frac{1}{2}$, without loss of generality we can assume that $q_a > \frac{1}{2}$. We will iteratively apply Proposition 9. Let f^* be the dominant linear non-guessing rule when $q_a = q_{a'} = \frac{1}{2}$. Let $\bar{f}\left(\cdot|q\right)$ be the rule such that $\bar{f}_a\left(X|q\right) = 2q_a - 1 + 2\left(1 - q_a\right) f_a^*\left(X\right)$. Note that $\bar{f}\left(X|q\right)_a \in [0,1]$ and $\bar{f}\left(X|q\right)_a - q_a = 2\left(1 - q_a\right) f_a^*\left(X\right) - \left(1 - q_a\right) = 2\left(1 - q_a\right)\left(f^*\left(X\right)_a - \frac{1}{2}\right)$. Hence $\bar{f}\left(\cdot|q\right)$ is a non-guessing rule under q. Note that $\bar{f}\left(X|q\right)_a$ is continuous in q whenever $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$. Two claims for the limit as $\min\left\{m_a, m_{a'}\right\}$ goes to infinity follow from the first part of this proof. If $\mu_a^G \geq \mu_{a'}^G + d$ then $Ef^*\left(\bar{X}\right)_a$ converges uniformly to 1 and hence so does $E\bar{f}\left(\bar{X}|q\right)_a$. If $\mu_a^G \leq \mu_{a'}^G - d$ then $Ef^*\left(\bar{X}\right)_a$ converges uniformly to 0 and hence $E\bar{f}\left(\bar{X}|q\right)_a$ converges to $2q_a - 1$.

In the following we split up the data set and sequentially apply this construction a finite number of times. Assume that $m = k \cdot z$ where k and z are positive integers. Given $M = \left(a_i, x_i\right)_{i=1}^m$ let $M_i = \left(\left(a_j, x_j\right)_{j=(i-1)z+1}^{iz}\right)$ for $i = 1, .., k$. We consider a sequence $\left(f^{(i)}\right)_{i=1}^s$ of s rules such that $f^{(i)} = \bar{f}\left(M_i | f^{(i-1)}\right)$ where $f^{(0)} = q$ and $M_0 = \emptyset$.

To simplify the exposition of the argument we will first consider z to be very large. Assume $\mu_a^G \geq \mu_{a'}^G + d$. Then $f_a^{(i)}$ will approach 1 for all i. Now consider the case where $\mu_a^G \leq \mu_{a'}^G - d$. Then $f_a^{(1)}$ will approach $2q_a - 1$. Note that $q_a - \left(2q_a - 1\right) = 1 - q_a > 0$. So $f_a^{(i)}$ will only stay above $\frac{1}{2}$ for finitely many steps. Thereafter it will fall below $\frac{1}{2}$. Once it falls below $\frac{1}{2}$ the roles of actions a and a' are swapped and thereafter $f_{a'}^{(i)}$ will approach

1. Note that we need $q_a - (k-1)(1-q_a) < \frac{1}{2}$ so we can choose $k = \left\lceil \frac{2q_a - 1}{2(1-q_a)} \right\rceil + 1$ steps.

Now consider what is really happening for large but finite z. We cannot rule out any values of $f^{(\iota)}(M_i)$. Yet with arbitrarily high probability we can ensure that the $f_a^{(k)}$ is close to 1 if $\mu_a \geq \mu_{a'} + d$ and close to 0 if $\mu_a^G \leq \mu_{a'}^G - d$. This is done by choosing z sufficiently large and using the uniform convergence that we proved for the case where $q_a = q_{a'} = \frac{1}{2}$. Consequently the dominant linear non-guessing rule will learn approximately which action is best when $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$. ■

The proof above reveals for any given $\alpha > 0$ that the value of z can be chosen such that the statement in the proposition holds whenever $q_a, q_{a'} \geq \alpha$.

5.4 Finite Sample Learning

In the following we show a close connection between the ability of a rule to learn which action is best and the non-guessing criterion. Clearly no rule will be able to guarantee to choose the best action in the true environment when the two actions are distributed very similarly. However we do expect that they can choose the best action with a high probability provided the two actions are sufficiently different. The following definition provides a way to quantify this from the perspective of the decision-maker. Let $a^*(G) \in \arg\max\left\{\mu_a^G, \mu_{a'}^G\right\}$ which uniquely defines the best action under any G with $\mu_a^G \neq \mu_{a'}^G$.

Definition 18 *We call $\psi \leq 1$ the confidence level of rule f for identifying the best action under the pseudo prior q and the minimal effect size $d > 0$ if (i) $E_G\left(f\left(\bar{X}\right)\right)_{a^*G} \geq \psi$ holds whenever $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$, and (ii) for any $\psi' > \psi$ there exists G such that $E_G\left(f\left(\bar{X}\right)\right)_{a^*G} < \psi'$ and $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$.*

Some notes are in place. As the decision-maker does not know the true distribution the property they identify for the true distribution has to hold for all distributions. As it is not possible to learn which action is best when the two actions are arbitrarily similar the specification of how well the rule is for learning has to be conditional on there being some minimal difference between the two actions in the true distribution. As the decision-maker only cares about means we quantify the difference between the two distributions by the difference between their respective means.

Proposition 19 *(i) For any given pseudo prior q the confidence level of the best linear non-guessing rule under a given minimal effect size weakly increases if more observations are added to the data set.*

(ii) Consider a balanced sample. The best linear non-guessing rule under the uninformative pseudo prior attains the highest confidence level under any given minimal effect size. In particular, any Bayesian rule attains a weakly lower confidence level for the same minimal effect size.

So we find that the best linear non-guessing rule f^* under the uninformative pseudo prior is best at learning which action is better when the sample is balanced. Its performance improves when more observations are added and the data set becomes unbalanced. We do not know which rule is best when the sample is unbalanced. However the monotonicity identified in (i) has the following implication. For a given unbalanced data set the minimal probability of the best linear rule f^* is bounded below by that of f^* in the largest balanced data set that is contained the unbalanced data set considered. It is also bounded above that in by the smallest balanced data set that contains the unbalanced data set considered.

Next we investigate in more detail the relationship between the minimal effect size and the confidence level.

5.4.1 Balanced data sets under the uninformative pseudo prior

Consider a balanced data set and the uninformative pseudo prior. Let $\psi^* (m_a, m_{a'}, d, q_a)$ denote the confidence level of the best linear rule given sample sizes m_a and m_a under the minimal effect size d and the weight q_a put by the pseudo prior on action a. Let $d^* (m_a, m_{a'}, \psi, q_a)$ denote its inverse when considering ψ^* as a function of d, so $\psi^* (m_a, m_{a'}, d^* (m_a, m_{a'}, \psi, q_a)) = \psi$.

We consider the relationship between minimal effect size and confidence levels. Let $B (i, m, p) = \binom{m}{i} p^i (1 - p)^{m-i}$.

Proposition 20 *The confidence level ψ^* of the best linear rule in a balanced sample under the uninformative prior is given by*

$$\psi^* \left(m_a, m_a, d, \frac{1}{2} \right) = \sum_{i=m/2+1}^{m} B \left(i, m, \frac{1}{2} (1 + d) \right) + \frac{1}{2} B \left(m/2, m, \frac{1}{2} (1 + d) \right).$$

The probability of choosing the best action is equal to this value of ψ when $\mu_a^G = \frac{1}{2} (1 + d)$ and $\mu_{a'}^G = \frac{1}{2} (1 - d)$.

The proof uses results in (Schlag, 2006) given the close connection to minimax regret.

In Figure 1 we graph $\psi^*\left(10,10,\cdot,\frac{1}{2}\right)$. We observe that $\psi^*\left(10,10,\cdot,\frac{1}{2}\right)$ is approximately equal to $\frac{1}{2}+\frac{7}{4}\cdot d$ for ψ^* below 0.65 and to $1-(1-d)^{\frac{20}{3}}$ for ψ^* above 0.9. Numerical calculations show that the minimal effect size needed to support a given confidence level drops approximately by 29.1% when doubling the sample size. This means that $d^*(m_a,m_{a'},\psi)\approx 0.709\cdot d^*(m,\psi)$, numerical calculations show that the error of this approximation is below 0.005 when $10\leq m\leq 200$. Specifically, we find that d^* is multiplicatively separable in ψ and m_a and can be written as

$$d^*(m,\psi)\approx d^*(40,40,\psi)\cdot\left(\frac{m_a}{40}\right)^{\frac{\ln(0.709)}{\ln 2}} \quad \text{if } m_a=m_{a'} \text{ and } q_a=\frac{1}{2}. \tag{3}$$

Here we present the approximation in a way such that it is exact when $m_a=m_{a'}=40$.

Following (3), one can linearly rescale the x achsis in Figure 1 to obtain the relationship between minimal effect size and confidence level for other balanced samples. For example, when $m_a=m_{a'}=100$ then multiply each of the tick marks 0.2, 0.4, 0.6, ... with $\left(\frac{200}{20}\right)^{\frac{\ln(0.709)}{\ln 2}}=0.319\,05$, so 0.2 turns into 0.06381.

5.4.2 Unbalanced Data Sets under the Uninformative Pseudo Prior

We now turn to unbalanced samples. We no longer have an exact formula for the confidence level. The choice probabilities are easily determined for any pair of means given the formula for the best linear rule given in Theorem 13. However, a formal identification of the worst case is missing. Never-the-less we obtain similar results. The mean of the worst case distribution is for all practical purposes essentially in the middle, namely if $q_a<\frac{1}{2}$ then the worst case distribution for minimal effect size d is attained when $\mu_a^G\approx\frac{1}{2}(1-d)$ and $\mu_{a'}^G\approx\frac{1}{2}(1+d)$. The error by evaluating ψ^* as if means are in the middle is below 0.002. Given this approximate worst case we can compute the confidence levels using the formula in Theorem 13. This has a simple approximate representation, specifically,

$$d^*(m_a,m_{a'},\psi)\approx\frac{d^*\left(40,40\cdot\frac{m_{a'}}{m_a},0.9\right)}{d^*(40,40,0.9)}d^*(40,40,\psi)\cdot\left(\frac{m_a}{40}\right)^{\frac{\ln(0.709)}{\ln 2}} \quad \text{for } m_{a'}\geq m_a \text{ and } q_a=\frac{1}{2}.$$

So d^* is multiplicatively separable in m_a, $\frac{m_{a'}}{m_a}$ and ψ, here the formula is written in a way that it is exact when $m_a=40$ and $\psi=0.9$.

It seems inefficient to learn from an unbalanced sample. We present a means to measure this inefficiency. Assume that there is a balanced sample and you wish to add (an even number of) k more observations. Seems most efficient to allocate them evenly to the two actions. This can be shown formally by following Schlag (2006), which is however beyond the scope of this paper. What if they are all added as observations of a'? Then the resulting confidence level could have also been reached with a balanced sample by adding less observations. Let $\hat{m}$ be this balanced sample, so $\psi^*\left(m_a, m_a + k, d, \frac{1}{2}\right) = \psi^*\left(\hat{m}_a, \hat{m}_a, d, \frac{1}{2}\right)$. Then $2m_a + k - 2\hat{m}_a$ observations were needlessly added. The percentage of added observations wasted is hence $\frac{2m_a+k-2\hat{m}_a}{k} * 100$. In Figure 2 we plot this against the percentage of observations added as given by $\frac{k}{m_a} * 100$. Due to integer constraints, $\hat{m}$ is typically computed as random combination of two adjacent sample sizes. For example, start with a balanced sample of 40 observations of each action and add 16 observations of action a'. The confidence level reached with this unbalanced sample is the same as in the balanced sample with 46.34 observations of each action. The percentage of additional observations wasted is then equal to $\frac{96-2*46.34}{16} * 100 = 20.75\%$. Note that the percentage waisted would equal 0 if no observations are saved by allocating equally. It would be 100% if only data that comes in pairs improves the confidence level. Given (3) the value of $\hat{m}_a$ will approximately not depend on the minimal effect size d. Observe that learning is reasonably efficient even in very unbalanced samples as only roughly 1/3 of the added observations are wasted when one sample is doubled.

5.4.3 Balanced Data Sets under General Pseudo Priors

Finally we turn to the inefficiencies driven by using a pseudo prior that is not uninformative, so $q_a \neq \frac{1}{2}$. We investigate this in balanced data sets. We find that the mean of the worst case distribution is in the middle and we once again have a simple numerical representation, namely,

$$d^*\left(m_a, m_a, \psi, q_a\right) \approx d^*\left(40, 40, \psi, q_a\right) \cdot \left(\frac{m_a}{40}\right)^{\frac{\ln(0.709)}{\ln 2}}.$$

In Figure 3 we plot the confidence level as a function of the minimal effect size for different pseudo priors. Note the inefficiency in learning introduced by the asymmetric pseudo prior. By the non-guessing condition, the expected weight put on action a will increase whenever it is the better action. However, this increase will be small as lots of evidence in favor of a will be needed to substantially increase its weight. Moreover, in the worst case the weight put on the better action starts at $\min\{q_a, q_{a'}\}$ which is lower than under the uninformative prior where it starts at $\frac{1}{2}$. The worst case is attained when action that receives less weight under the pseudo prior is the better action.

To get a better feel for the inefficiencies introduced by the asymmetric pseudo prior we calculate the percentage of samples that were wasted (when concerned only with the confidence level) under an unequal pseudo prior as compared to when using the efficient uninformative pseudo prior. We show the results for $m_a = m_{a'} = 40$ in Figure 4. So for $q_a = \frac{1}{3}$ confidence level of 0.9 (which requires $d \approx 0.191$) we find approximately 44% wasted, for $q_a = \frac{1}{10}$ the value is approximately 75%.

Remember that the pseudo prior might be the result of earlier learning. One expects that action a is better when $q_a > \frac{1}{2}$ but this need not be true as it might be that one obtained untypical realizations of the payoffs. The findings above show how likely it will be to learn that action a' is better even $q_a > \frac{1}{2}$. The pseudo prior might be also be driven by the circumstances surrounding the decision-making. We observe above that learning abilities are substantially impeded when not choosing the uninformative pseudo prior. This is expected as the pseudo prior reflects some beliefs or understanding of the environment that is not captured in our measure of learning performance as given by the confidence level. We present a more fitting definition of learning for settings in which the pseudo prior reflects beliefs. Assume that the value q_a given by the pseudo prior captures the probability that action a will be the best action. This is a type of belief. It can result from asking experts which action is to choose and then setting q_a equal to the proportion who recommend action a. One might then assume that one of the experts is correct with each expert being equally like to be correct. Consequently we let the true environment reflect this belief and assume q_a represents the true probability that action a is the best action. This motivates the following definition.

Definition 21 *We call $\psi^w \leq 1$ the weighted confidence level of rule f for identifying the best action under the pseudo prior q and the minimal effect size $d > 0$ if (i) $q_a E_{G^{(a)}} \left(f\left(\bar{X}\right)\right)_a + q_{a'} E_{G^{(a')}} \left(f\left(\bar{X}\right)\right)_{a'} \geq \psi^w$ holds whenever $\mu_a^{G^{(a)}} \geq \mu_{a'}^{G^{(a)}} + d$ and $\mu_{a'}^{G^{(a')}} \geq \mu_a^{G^{(a')}} + d$, and (ii) for any $\psi' > \psi^w$ there exist distributions $G^{(a)}$ and $G^{(a')}$ such that $q_a E_{G^{(a)}} \left(f\left(\bar{X}\right)\right)_a + q_{a'} E_{G^{(a')}} \left(f\left(\bar{X}\right)\right)_{a'} < \psi'$, $\mu_a^{G^{(a)}} \geq \mu_{a'}^{G^{(a)}} + d$ and $\mu_{a'}^{G^{(a')}} \geq \mu_a^{G^{(a')}} + d$.*

It follows that the weighted confidence level is at least $q_a^2 + (1 - q_a)^2$. This is because with probability q_a ($q_{a'}$) action a (a') is the better action in which case the probability of choosing action a (a') is at least q_a ($q_{a'} = 1 - q_a$). So the minimal value is higher when the pseudo prior is more extreme. This does not mean that the better performance is realized by choosing a more extreme pseudo prior as in this scenario the pseudo prior is given by the understanding of the environment or the recommendations of experts and not the result of a strategic choice. With this new definition we reproduce the analysis that led to Figures 3 and 4. In Figure 5 we illustrate the tradeoff between minimal effect size and weighted confidence level in balanced samples using $m_a = m_{a'} = 40$. Compared to the uninformative pseudo prior, observe that the unequal pseudo priors are more efficient for small effect sizes and slightly less efficient for large ones. In Figure 6 we evaluate the percentage of observations waisted or excessive (when value is negative) as compared to balanced sample with uninformative pseudo prior when $m_a = m_{a'} = 40$.

We observe how much more efficient (in terms of saving observations) it is to follow unequal pseudo prior as opposed to using the uninformative pseudo prior, unless the weighted confidence level is very high.

5.5 Non-Randomized Rules

In this section we briefly show that evidence can also be extracted from the data set with a non-randomized rule. To do so, sufficient evidence is needed. Evidence refers here to how to get better payoffs than under the pseudo prior. Sufficient evidence means that the two actions should not perform too similar.

Our analysis above shows that randomization is needed to ensure non-guessing. The proof of Theorem 13 shows that this is due to the environments where both actions are almost equally good. Once there is a minimal difference between the means of the two actions then non-randomized rules exist that are non-guessing. In that case there is enough evidence that randomization is no longer needed. We present such rules. They are constructed from the dominant linear non-guessing rule f^* as follows. However they are no longer linear. Consider $\kappa_a, \kappa_{a'} \in (0,1)$ such that $\kappa_{a'} = 1 - \kappa_a$. Let f^d be the rule such that $f^d(X)_a = 1$ if $f^*(X)_a \geq \kappa_a$ and $f^d(X)_{a'} = 1$ if $f^*(X)_a > \kappa_{a'}$ (which means that $f^d(X)_a = 0$ if $f^*(X)_a < \kappa_a$). So action a is chosen if and only if the probability of choosing a under the dominant linear non-guessing rule is sufficiently high, action a' is chosen otherwise. In the appendix we show that a sufficient condition for this rule to be non-guessing under q is that

$$E_G\left(f^*\left(\bar{X}\right)\right)_{\bar{a}} \geq (1-\kappa_{\bar{a}})\, q_{\bar{a}} + \kappa_{\bar{a}}$$

holds for all $\bar{a} \in A = \{a, a'\}$ and $G \in \mathcal{G}$. This condition will hold if $\left|\mu_a^G - \mu_{a'}^G\right|$ is sufficiently large. For given sample sizes m_a and $m_{a'}$ the free parameter κ_a is easily found that makes the required minimal difference between the two actions as small as possible. For example, when $q_a = \frac{1}{2}$ and $\min\{m_a, m_{a'}\} \geq 23$ it is best to set $\kappa_a = \frac{1}{2}$ and the non-guessing condition holds when $\left|\mu_a^G - \mu_{a'}^G\right| \geq 0.1$. Note that this rule is a particular ε non-guessing rules for $\varepsilon = 0.05$ as we do not know whether it can extract evidence when $\left|\mu_a^G - \mu_{a'}^G\right| < 0.1$, at most it loses $\frac{1}{2} \cdot 0.1 = 0.05$ is lost when comparing to q.

The maximal regret of these "de-randomized" rules can be bounded from above, using the techniques provided in Schlag (2006b). This bound is equal to $\frac{1}{\min\{\kappa_a, \kappa_{a'}\}}$ times the maximal regret of the dominant linear non-guessing rule. In the example above, the regret of the derandomized rule is bounded by $2 \cdot 0.02512 \leq 0.0503$.

5.6 Almost No Guessing

We briefly expand on a weaker concept. The starting point is the insight from Proposition 4 that actions that are do not receive any weight under the pseudo prior will not be chosen, regardless of how much data is gathered and how attractive they look later. This result relies on the unwillingness to accept any reduction in payoffs, regardless of how small it might be. We slightly relax our definition to tolerate small reductions.

Definition 22 *Given $\varepsilon > 0$, the rule f is an ε* almost non-guessing rule under prior *q if $E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G \geq q \cdot \mu^G - \varepsilon$ holds for all $G \in \mathcal{G}$.*

For any given $\varepsilon > 0$ we construct a rule that is ε almost non-guessing and that puts weight on actions that are not in the support of q. We do this by selecting a mixed action $\bar{q}$ that has full support and that is sufficiently close to q so that payoffs never fall below $q \cdot \mu^G - \varepsilon$. The full support condition on $\bar{q}$ ensures that the rule also puts weight on all actions and not only on those in the support of q. Thus, innovation is possible for any data set that contains information on all actions if the decision-maker is willing to perform slightly worse than the pseudo prior.

We present the parameters of the rule suggested and compute its maximal regret in some examples. Consider the case of two actions where $A = \{a, a'\}$ and assume that $q_a = 1$. We search for a rule f such that $E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G - \mu_a^G \geq -\varepsilon$ and $f\left(X\right)_{a'} > 0$ for some data sets X. Clearly, without making any assumptions on f, $E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G - \mu_a^G \geq 0$ holds when $\mu_{a'}^G \geq \mu_a^G$. So we aim to find a rule f such that $\mu_a^G - E_G\left(f\left(\bar{X}\right)\right) \cdot \mu^G \leq \varepsilon$ holds when $\mu_a^G > \mu_{a'}^G$. This means that the regret of not choosing a (even though a is the best action) should be at most ε. We construct such a rule by choosing a rule f that is non-guessing under an appropriately chosen pseudo prior $\bar{q}$. Let $f^{(q)}$ denote a non-guessing rule under q. Let $\bar{q}_a$ be the smallest value of q_a such that $\mu_a^G - E_G\left(f^{(q)}\left(\bar{X}\right)\right) \cdot \mu^G \leq \varepsilon$ holds when $\mu_a^G > \mu_{a'}^G$. We show the maximal regret of this rule when $\varepsilon = 0.01$ for a few balanced sample sizes. These values of maximal regret are attained when a' is best as the maximal regret when instead a is best is bounded by $\varepsilon = 0.01$.

$\varepsilon = 0.01$	(k_1, k_2)	$\bar{q}_a$	$r \leq$
	$(6, 6)$	0.845	0.156
	$(12, 12)$	0.79	0.0873
	$(25, 25)$	0.722	0.0484
	$(145, 145)$	0.5	0.01

We observe for $k_1 = k_2 = 25$ that the non-guessing rule under $\bar{q}$ with $\bar{q}_a = 0.722$ is ε almost non-guessing under q with $q_a = 1$ when $\varepsilon = 0.01$. The wish of the decision-maker to benchmark to choosing a with a tolerance of payoff loss (under the pseudo prior) of at most 0.01 requires 25 observations of each action to guarantee regret below 0.05. Had they started with the uninformative pseudo prior then, as observed above, only 6 observations of each action would have been necessary.

Note also the connection to minimax regret. A rule f that guarantees that regret is below $\bar{r}$ has the property that the increase in payoffs when choosing the best action is at most $\bar{r}$. So the possible increase when choosing any action is also at most $\bar{r}$. Thus such a rule is ε almost non-guessing under any pseudo prior under $\varepsilon = \bar{r}$. As already noted in Section 5.1 after Corollary 14 the randomized extension of the random empirical success rule, which is the dominant linear non-guessing rule under the uninformative pseudo prior, attains minimax regret. Hence, the table entry above for $(k_1, k_2) = (145, 145)$ follows from the respective table entry in Section 5.4.

We hasten to point out that there is no claim that this rule minimizes maximum regret among all ε almost non-guessing rules. A selection among the ε almost no guessing rules is beyond the scope of this paper.

6 Including Covariates

In the following we briefly explain how one can include covariates. We closely follow Manski (2004). There is a finite covariate space. Payoffs depend on the underlying covariates. Choice is conditional on the covariates. In our setting this means that we can allow for pseudo priors that depend on the covariates. A rule is now a choice that conditions on the covariates and has as input a data set. Each element of the data set now has a covariate vector in addition to an action and payoff. We construct rules just as in (Manski, 2004) by separating inference across all the different possible covariate configurations. One only considers those data points that have the same covariates. To these one applies a non-guessing rule. Consequently, the combined rule is a non-guessing rule.

Note how the data set used for a given specification of the covariates gets small as more covariates are added. Here we observe the advantage of our concept of non-guessing rules as it has good properties regardless of the sample size.

7 Connecting to Social Learning and Imitation

Schlag (1998) introduced a setting for selecting rules for social learning. In this model there is an infinite population of individuals. Each individual has to choose an action. Between choices each individual observes performance of others selected at random from the population. Based on this observation the individual chooses her next action. So a rule is a function of choices and payoffs of own and of others. In Schlag (1998) each individual observes one other, in Schlag (1999) two others and in Hofbauer and Schlag (2000) a finite number of other individuals. The objective is to find an improving rule under which, when all use this rule, the average payoff attained in the population is weakly higher in the next round for any underlying distribution of payoffs.

First note that the improving criterion is evaluated ex-ante, prior to observing the choices of the others. However, as it has to hold for any distribution of choices in the population, it has to also hold conditional on who is observed. More specifically, we can assume without loss of generality that a group of individuals sees each other. So in the social learning model, if an individual observes $n-1$ others then the group consists of n individuals. The improving condition holds conditional in the sense that among these n individuals with their actions, ex-ante to observing anyone's payoffs, their expected payoff has to be higher in the next round. This insight uncovers a close connection between improving rules and non-guessing rules under the pseudo prior in which probabilities equal frequencies in the data set.

We explain in more detail. Consider an improving rule. We will construct a non-guessing rule when $q_a = \frac{m_a}{m}$ for all $a \in A$. To do this, randomly select (with equal probability) an element of the data set, act as if the associated action was your own choice and the others in the data set were the choices observed. We then apply the social learning rule. It is as if we are evaluating the average payoff in the data set and looking for this to be always larger than their average payoff in the last round. As the social learning rule is improving, this construction has led to a non-guessing rule under the pseudo prior that chooses the probability of an action equal to its frequency in the data set.

Consider now a non-guessing rule under the pseudo prior that chooses each action according to the frequencies this action appears in the data set. Consider an individual in the social learning context who observes $n-1$ others. Then combine those observations together with her own to define a data set. Apply the rule to this data set to determine the choice probabilities of each action. Now determine switching probabilities in the social learning setting such that the choice probabilities of the rule emerge. This generates an improving rule.

We summarize these findings.

Proposition 23 *Consider a data set X and assume $q_a = \frac{m_a}{m}$ for all $a \in A$. Any improving rule can be used to create a non-guessing rule under q, and vice versa, any non-guessing rule under q can be used to create an improving rule. The expected payoff of the individual rule is equal to the average expected payoff of the associated improving rule if each data point would be associated to an individual who is observing the performance of the others belonging to the data set.*

In particular we observe a direct connection between the insight that choice probabilities of actions not observed may not change (Proposition 4) and the insight that improving rules must be imitating (see Lemma 1 in (Schlag, 1998)).

We can utilize this connection to uncover more about the structure of rules. Schlag (1999) shows that there is no dominant rule for learning from two others when $n = 2$. This means for the setting of this paper that there is no dominant rule when $m_a = 1$, $m_{a'} = 2$ and $n = 2$. So our limitation to linear rules comes with a loss of generality unless $m_a = m_{a'} = 1$.

Finally we note that we have uncovered for the case of two actions a social learning rule that generates a payoff-monotone selection dynamics (Samuelson & Zhang, 1992) in any game where this dynamics approximates the best response dynamics if sufficiently many others are observed and the two actions are not too similar. This follows from Proposition 17 and the above insights.

8 Connecting to Unbiased Hypothesis Testing

Consider the case where there are only two actions. In the following we show that there is a close connection to statistical hypothesis testing. In short, choosing an action is like saying which action is better, it is as if accepting or rejecting the null hypothesis that one action has a larger mean than the other. This leads to a close relationship between a dominant linear non-guessing rule and UMPU test for binary valued data. This relationship reveals a simple interpretation of the dominant linear non-guessing rule as the randomized extension of a permutation test.

We expand. First we show that a rule is non-guessing if and only if it is associated to an unbiased test for comparing the means of the two actions. Consider the null hypothesis $H_0 : \mu_a \leq \mu_{a'}$, or more formally, $H_0 = \left\{G : \mu_a \leq \mu_{a'}^G\right\}$. Then a randomized test ϕ specifies a rejection probability of H_0 given the data. So $\phi\left(\{(a_i, x_i)\}_{i=1}^m\right)$ is the rejection probability of H_0 given $X = (a_i, x_i)_{i=1}^m$. The test ϕ has level α if $E_{G_0}(\phi) \leq \alpha$ whenever G_0 is in H_0, so when $\mu_a^{G_0} \leq \mu_{a'}^{G_0}$. Without loss of generality we can assume that $\alpha = \sup_{G_0 \in H_0} E_{G_0}(\phi)$. The test ϕ is unbiased if $E_{G_1}(\phi) \geq E_{G_0}(\phi)$ whenever G_1 is not in H_0 and G_0 is in H_0.

Assume that ϕ is unbiased. Then it follows that $E_{G_1}(\phi) \geq \alpha \geq E_{G_0}(\phi)$ for $G_1 \notin H_0$ and $G_0 \in H_0$. Consequently, $(E_G(\phi) - \alpha)\left(\mu_a^G - \mu_{a'}^G\right) \geq 0$. This is equivalent to $E_G(\phi) \cdot \mu_a^G + (1 - E_G(\phi))\,\mu_{a'}^G \geq \alpha\mu_a^G + (1-\alpha)\,\mu_{a'}^G$. Thus we have shown that if ϕ is an unbiased test then the rule $f = (\phi, 1-\phi)$ is non-guessing under $q = (\alpha, 1-\alpha)$.

For the converse, consider a non-guessing rule f under $q = (\alpha, 1-\alpha)$. From the equations above it follows that $(E_G(f_a) - q_a)\left(\mu_a^G - \mu_{a'}^G\right) \geq 0$ and hence $E_{G_1}(f_a) \geq q_a \geq E_{G_0}(f_a)$ whenever $G_1 \notin H_0$ and $G_0 \in H_0$. So this means that f_a is an unbiased test of H_0 with level q_a.

Consider now binary valued data. Note that there is a uniformly most powerful test of the above null hypothesis (Tocher, 1950). It is a permutation test with test statistic equal to the number of successes among the samples of action a. It rejects the null hypothesis if the proportion of permutations with more successes among action a than in the observed data set is at most α with appropriate randomization close to the threshold. Note that this is in fact the dominant linear non-guessing rule that is non-guessing. This is no coincidence given our above presentation.

We now formally connect our dominance property to the UMPU property. Consider rules f and g that both are non-guessing rules under q. Then f dominates g if $E_G(f) \cdot \mu^G \geq E_G(g) \cdot \mu^G$. This is equivalent to $\left(E_G(f)_a - E_G(g)_a\right) \cdot \left(\mu_a^G - \mu_{a'}^G\right) \geq 0$. Hence, $E_G(f)_1 \geq E_G(g)_a$ for all $G \notin H_0$ which means that f_a as a test is uniformly more powerful than g_a. However, we can also conclude that $E_G(f)_a \leq E_G(g)_a$ for all $G \in H_0$. This means that f_a is less likely to reject the null hypothesis when it is true. This is a property that a uniformly more powerful test does not need to fulfill according to its definition.

Conversely, consider two tests ϕ and ϕ' with level α where ϕ is uniformly more powerful than ϕ'. Then $(E_G(f)_a - E_G(g)_a) \cdot (\mu_a^G - \mu_{a'}^G) \geq 0$ when $\mu_a^G > \mu_{a'}^G$. However, the definition of uniformly more powerful does not allow us to conclude that $(E_G(f)_a - E_G(g)_a) \cdot (\mu_a^G - \mu_{a'}^G) \geq 0$ also holds when $\mu_a^G < \mu_{a'}^G$.

We can however easily mend this problem. Namely, we can apply the same methodology to testing $H_0 = \{G : \mu_a^G \geq \mu_{a'}^G\}$ with size $1 - \alpha$. Essentially we have swapped the null and the alternative hypothesis. The same UMPU test emerges, except that labels are swapped, which allows us to conclude that $(E_G(f)_a - E_G(g)_a) \cdot (\mu_a^G - \mu_{a'}^G) \geq 0$ holds when $\mu_a^G < \mu_{a'}^G$.

To conclude, we obtain an equivalence between the dominant linear non-guessing rule for binary valued data and the UMPU test for comparing means of two Bernoulli distributions. Note that linear rules are uniquely identified on binary valued data. Comparing linear rules is like comparing rules based on binary data. Hence our dominant linear non-guessing rule is equivalent to a uniformly most powerful test within class of unbiased and linear tests for comparing the means of two variables with known bounds on their support.

On the side the insights above reveal an alternative proof of Theorem 13, by utilizing the UMPU properties and the randomization trick.

We add a comment on the methodology of statistical hypothesis testing. It is hard to argue in favor of the property of a test being unbiased apart from the ease that this adds to construct tests and find uniformly most powerful ones. Given what we write above, we have uncovered the following insights related to statistical hypothesis testing. Consider a statistician who will most likely (with probability $1 - \alpha$) choose action a' and is only willing to change his opinion if the data gives conclusive evidence that he can do better. Then this statistician should choose the action prescribed by the UMPU test. Among all reactions based on conclusive evidence, it not only maximally increases the likelihood of correctly rejecting the null, it also maximally decreases the likelihood of incorrectly rejecting the null. We hasten to point out one disadvantage regarding classical statistical hypothesis testing. The recommendation of our dominant linear non-guessing rule for interval data is typically randomized. Actions are recommended with probabilities. One cannot find rules in our setting that are nonrandom (however do note the results in Section 5.5).

9 Measuring Evidence in Data Sets

Assume that there are two actions to choose from and you wish to make a probabilistic statement, based on the data, about which action has the highest mean. Take our dominant linear non-guessing rule under the uninformative pseudo prior and interpret choice of action a as saying that a has a higher mean than the other action. To apply this non-guessing rule in this way has the property that its expected probability of stating the truth is at least $\frac{1}{2}$. We put this into context. Recalling from elementary statistics, the average observed payoff of action a is an unbiased estimate of the mean of action a, and the expected difference in the average observed payoffs of action a and action a' is positive if and only if a has the higher mean. Yet what we have found in this paper that stating that the action with the higher observed average is also the one with the higher mean can be worse than a uniform random guess. The methodology in this paper shows how to make statements that are always at least as true as a uniform random guess. The dominant linear non-guessing rule under the uninformative pseudo prior is the best such rule in the sense that it is more correct (i.e., it has a higher probability placed on the best action) than any other linear rule that always outperforms the uniform random guess.

10 A Comparison

We summarize the similarities and differences between SEU maximization and non-guessing. Both start with some personal initial understandings about the environment they are facing. The Bayesians think about which environments they might be facing. The non-guessing decision-makers think directly about which actions they might be choosing. They both use probabilities to combine these initial understandings and update this combined understanding after observing the data. The updated combined understanding can then be used as an initial combined understanding for the next data set.

Both kinds of decision-makers can be so convinced in their understanding that no data set can change this understanding. This happens for a Bayesian who has a degenerate prior on a single distribution and for a non-guessing decision-maker who has a non degenerate pseudo prior. On the other hand, both will essentially learn which action is best in large data sets if the true environment is reflected in one of their initial understandings. The only problem with SEU maximization is that their description of the environment is extremely detailed. The set of possible understandings is an infinitely dimensional set. Any prior that comes to mind will not have all possible distributions in its support. Consequently, the best action among those initially considered need not be the best in the true environment. Discretizing the support to make it finitely dimensional does not help in organizing the large set of possible distributions which is needed before assessing a prior. On the other hand, non-guessing decision-makers can easily include all possible actions in their initial understanding (when n is not too large) to ensure that they find the best action in the true environment when the data set is sufficiently large. Large sets of actions may still be difficult to handle, but they do typically come with a clear structure (like prices).

The understanding of the Bayesian as captured by the prior is a sufficient statistic of the data sets observed in the past. The updated prior stores all relevant information when facing a new data set. In contrast, the pseudo prior underlying any non-guessing rule does not necessarily store all relevant information from past data sets observed. A non-guessing decision-maker is often better off remembering all past data sets when facing the next one (as shown in Section 5.1) instead of just remembering the updated pseudo prior.

The understanding of the Bayesian might change dramatically after observing the data set. This happens when observing some data set that cannot be generated under one of the distributions in the support of their prior. After observing this data set that distribution is ruled out in their posterior and receives no longer any weight. Things look a bit different for non-guessing rules. When the data set does not contain both extreme payoffs then the support of the pseudo prior does not change after observing the data set. On the other hand, if both extreme payoffs are in the data set then the understanding possibly changes drastically. In other words, Bayesians may jump to conclusions and rule out understandings they previously had. Non-guessing decision-makers will not do so if the data set does not contain both extreme payoffs.

Bayesians do not care if they might perform worse with the data set than without it under the true distribution as their prior makes them comfortable in trading off performance across different understandings. On the other hand, a non-guessing rule has the defining property that performance with the data set is always superior than without it, regardless of which underlying distribution is the true one.

In the setting of this paper, Bayesians condition their behavior on a prior whose possible support is infinitely dimensional. The underlying set of environments is so rich that an uninformative (i.e., uniform) prior does not exist. Priors are easily criticized as being arbitrary or too complicated. Given the richness, priors can be hard to identify or to describe to others. Moreover, simple comparative statics on the understanding of the decision-maker do not exist as there are many ways in which the prior can be changed. On the other hand, non-guessing rules rely on a much simpler object, namely on a mixed strategy. This makes it is easier to convey to others. The uninformative pseudo prior in which each action is chosen equally likely is focal. Comparative statics are easy to perform. For Bayesians, the richness of the set of priors makes it hard to understand how much data is needed to establish good learning properties. In contrast, non-guessing rules come with simple bounds on how performance depends on the sample sizes (see Section 5.4).

Information has value for Bayesians, in the sense of it being preferred to no information, as Bayesians treat the world as if the distribution they face has been drawn from their prior. This value is personal and hypothetical as it only emerges when averaging across different environments, using weights that they have assigned. In some distributions the value of information under their rule can be negative. There is always some other Bayesian who will recommend them not to use the data. On the other hand, information has value for non-guessing rules regardless of which environment is the true one. Everyone can agree that the decision-maker is better off, so there is no personal element in this assertion. Thus, in some sense, only a non-guessing rule extracts objective information from the data set.

11 Conclusion

to be added

12 Appendix

12.1 Proof of Theorem 13

Without loss of generality assume $m_a \leq m_{a'}$. Following Proposition 11 it is enough to focus on distributions that yield payoffs in $[0,1]$. So all properties that refer to the distributions only depend on their means. Let d_{ij} be the additional probability put on action a on top of q_a when observing i winners of action a and j winners of action a', so $d_{ij} \in [-q_a, q_{a'}]$.

Let $F_a = f_a - q_a$ be the expected increase of play of action a as a function of the means of the two actions μ_a and $\mu_{a'}$. So

$$F_a(\mu_a, \mu_{a'}) = \sum_{i=0}^{m_a} \sum_{j=0}^{m_{a'}} d_{ij} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a - i} \binom{m_{a'}}{j} \mu_{a'}^j (1-\mu_{a'})^{m_{a'}-j} .$$

In particular,

$$\begin{aligned} F_a(\mu_a, \mu_a) &= \sum_{i=0}^{m_a} \sum_{j=0}^{m_{a'}} d_{ij} \binom{m_a}{i} \binom{m_{a'}}{j} \mu_a^{i+j} (1-\mu_a)^{n-i-j} \\ &= \sum_{r=0}^{n} \left(\sum_{i=\max\{0, r-m_{a'}\}}^{\min\{m_a, r\}} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{j} \right) \mu_a^r (1-\mu_a)^{n-r} . \end{aligned}$$

Following Proposition 5, $F_a(\mu_a, \mu_a) = 0$ holds for all $\mu_a \in [0,1]$. In particular, $d_{0,0} = d_{m_a, m_{a'}} = 0$. The identity theorem for polynomials implies that

$$\sum_{i=\max\{0, r-m_{a'}\}}^{\min\{m_a, r\}} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} = 0$$

holds for all $r \in \{1, ..., n-1\}$.

Note that

$$\begin{aligned} F_a \;=\; & \sum_{r=1}^{m_a} \sum_{i=0}^{r} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\ & + \sum_{r=m_a+1}^{m_{a'}} \sum_{i=0}^{m_a} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\ & + \sum_{r=m_{a'}+1}^{n-1} \sum_{i=r-m_{a'}}^{m_a} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \end{aligned}$$

Accordingly we will distinguish between two cases for r.

Assume $r \in \{1, .., m_a\}$. Let F_a^r be be the expected increase of play of action a conditional on r successes. We use the fact that using that

$$\sum_{i=0}^{r-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} d_{i,r-i} + \binom{m_a}{r} d_{r,0} = 0$$

to obtain

$$\begin{aligned}
F_a^r &= \sum_{i=0}^{r} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\
&= \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\
&\quad + d_{r,0} \binom{m_a}{r} \mu_a^r (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}} \\
&= \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \mu_a^i (1-\mu_a)^{m_a-i} \binom{m_{a'}}{r-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\
&\quad - \sum_{i=0}^{r-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} d_{i,r-i} \mu_a^r (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}} \\
&= \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} \left(\begin{array}{c} \mu_a^i (1-\mu_a)^{m_a-i} \mu_{a'}^{r-i} (1-\mu_{a'})^{m_{a'}-(r-i)} \\ -\mu_a^r (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}} \end{array} \right) \\
&= \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} \left(\begin{array}{c} ((1-\mu_a)\mu_{a'})^{r-i} \\ -((1-\mu_{a'})\mu_a)^{r-i} \end{array} \right) \mu_a^i (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}-(r-i)}
\end{aligned}$$

Using the fact that

$$((1-\mu_a)\mu_{a'})^{r-i} - ((1-\mu_{a'})\mu_a)^{r-i} = (\mu_{a'} - \mu_a) \sum_{j=0}^{r-i-1} ((1-\mu_a)\mu_{a'})^j ((1-\mu_{a'})\mu_a)^{r-i-1-j}$$

we obtain

$$F_a^r = -(\mu_a - \mu_{a'}) \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} \left(\begin{array}{c} \left(\sum_{j=0}^{r-i-1} ((1-\mu_a)\mu_{a'})^j ((1-\mu_{a'})\mu_a)^{r-i-1-j}\right) \\ \mu_a^i (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}-(r-i)} \end{array} \right)$$

Setting

$$w_{i,r-i} = \left(\sum_{j=0}^{r-i-1} ((1-\mu_a)\mu_{a'})^j ((1-\mu_{a'})\mu_a)^{r-1-j} \right) (1-\mu_a)^{m_a-r} (1-\mu_{a'})^{m_{a'}-r}$$

we can shorten the expression for F_a^r to obtain

$$F_a^r = -(\mu_a - \mu_{a'}) \sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} w_{i,r-i}.$$

In the following we aim to minimize

$$\sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} w_{i,r-i}$$

subject to the constraints that

$$-q_a \le d_{r,0} = -\frac{1}{\binom{m_a}{r}} \sum_{i=0}^{r-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} d_{i,r-i} \le q_{a'}$$

and $-q_a \le d_{i,r-i} \le q_{a'}$ for $i \le r-1$.

We will show that this is solved by $d_{i,r-i} = -q_a$ for $i < s$, $d_{i,r-i} = q_{a'}$ for $i > s$ and by $d_{s,r-s} \in [-q_a, q_{a'}]$ solving

$$\sum_{i=0}^{s-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} (-q_a) + \binom{m_a}{s} \binom{m_{a'}}{r-s} d_{s,r-s} + \sum_{i=s+1}^{r} \binom{m_a}{i} \binom{m_{a'}}{r-i} q_{a'} = 0.$$

The Lagrangian is given by

$$\begin{aligned} L &= -\sum_{i=0}^{r-1} d_{i,r-i} \binom{m_a}{i} \binom{m_{a'}}{r-i} w_{i,r-i} - \sum_{i=0}^{r-1} \lambda_i \left(d_{i,r-i} - q_{a'}\right) \\ &\quad -\lambda_r \left(-\frac{1}{\binom{m_a}{r}} \sum_{i=0}^{r-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} d_{i,r-i} - q_{a'} \right) \\ &\quad -\sum_{i=0}^{r-1} \tau_i \left(-q_a - d_{i,r-i}\right) - \tau_r \left(-q_a + \frac{1}{\binom{m_a}{r}} \sum_{i=0}^{r-1} \binom{m_a}{i} \binom{m_{a'}}{r-i} d_{i,r-i} \right). \end{aligned}$$

where

$$\begin{aligned} \frac{dL}{d\left(d_{i,r-i}\right)} &= -\binom{m_a}{i} \binom{m_{a'}}{r-i} w_{i,r-i} - \lambda_i + \lambda_r \frac{1}{\binom{m_a}{r}} \binom{m_a}{i} \binom{m_{a'}}{r-i} \\ &\quad +\tau_i - \tau_r \frac{1}{\binom{m_a}{r}} \binom{m_a}{i} \binom{m_{a'}}{r-i}. \end{aligned}$$

The solution is given by $\lambda_s = \tau_s = 0$, $\lambda_i = 0$ for $i < s$ and $\tau_i = 0$ for $i > s$, $\lambda_r = \binom{m_a}{r} w_{s,r-s}$,

$$\tau_i = \binom{m_a}{i} \binom{m_{a'}}{r-i} \left(w_{i,r-i} - w_{s,r-s}\right)$$

for $i < s$ and

$$\lambda_i = \binom{m_a}{i} \binom{m_{a'}}{r-i} \left(w_{s,r-s} - w_{i,r-i}\right)$$

for $s < i < r$. Note that $\tau_i \ge 0$ for $i < s$ and $\lambda_i \ge 0$ for $i > s$ as $w_{i,r-i}$ is decreasing in i.

(ii) Assume $r \in \{m_a + 1, ..., n-1\}$. We repeat the arguments above.
Using that

$$\sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} \binom{m_a}{i}\binom{m_{a'}}{r-i} d_{i,r-i} + \binom{m_{a'}}{r-m_a} d_{m_a,r-m_a} = 0$$

we obtain

$$\begin{aligned}
F_a^r &= \sum_{i=\max\{0,r-(n-m)\}}^{m} d_{i,r-i}\binom{m_a}{i}\mu_a^i\left(1-\mu_a\right)^{m_a-i}\binom{m_{a'}}{r-i}\mu_{a'}^{r-i}\left(1-\mu_{a'}\right)^{m_{a'}-(r-i)} \\
&= \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\mu_a^i\left(1-\mu_a\right)^{m_a-i}\binom{m_{a'}}{r-i}\mu_{a'}^{r-i}\left(1-\mu_{a'}\right)^{m_{a'}-(r-i)} \\
&\quad + d_{m_a,r-m_a}\binom{m_{a'}}{r-m_a}\mu_a^{m_a}\mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{n-r} \\
&= \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\mu_a^i\left(1-\mu_a\right)^{m_a-i}\binom{m_{a'}}{r-i}\mu_{a'}^{r-i}\left(1-\mu_{a'}\right)^{m_{a'}-(r-i)} \\
&\quad - \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} \binom{m_a}{i}\binom{m_{a'}}{r-i} d_{i,r-i}\mu_a^{m_a}\mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{n-r} \\
&= \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\binom{m_{a'}}{r-i}\left(\begin{array}{c} \mu_a^i\left(1-\mu_a\right)^{m_a-i}\mu_{a'}^{r-i}\left(1-\mu_{a'}\right)^{m_{a'}-(r-i)} \\ -\mu_a^{m_a}\mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{n-r} \end{array}\right) \\
&= \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\binom{m_{a'}}{r-i}\left(\begin{array}{c} \left(\left(1-\mu_a\right)^{m_a-i}\mu_2^{m_a-i} - \mu_a^{m_a-i}\left(1-\mu_{a'}\right)^{m_a-i}\right) \\ \mu_a^i\mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{m_{a'}-(r-i)} \end{array}\right) \\
&= -\left(\mu_a - \mu_{a'}\right)\cdot \\
&\quad \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\binom{m_{a'}}{r-i}\left(\begin{array}{c} \left(\sum_{j=0}^{m_a-i-1}\left(\left(1-\mu_a\right)\mu_{a'}\right)^j\left(\left(1-\mu_{a'}\right)\mu_a\right)^{m_a-1-j}\right) \\ \mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{m_{a'}-r} \end{array}\right)
\end{aligned}$$

Let

$$w_{i,r-i} = \left(\sum_{j=0}^{m_a-i-1}\left(\left(1-\mu_a\right)\mu_{a'}\right)^j\left(\left(1-\mu_{a'}\right)\mu_a\right)^{m_a-1-j}\right)\mu_{a'}^{r-m_a}\left(1-\mu_{a'}\right)^{m_{a'}-r}.$$

So we aim to minimize

$$\sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} d_{i,r-i}\binom{m_a}{i}\binom{m_{a'}}{r-i} w_{i,r-i}$$

such that

$$-q_a \leq d_{m_a,r-m_a} = -\frac{1}{\binom{m_{a'}}{r-m_a}} \sum_{i=\max\{0,r-m_{a'}\}}^{m_a-1} \binom{m_{a'}}{r-i} d_{i,r-i} \leq q_{a'}$$

and $-q_a \leq d_{i,r-i} \leq q_{a'}$ for $i \leq m_a - 1$. This is solved by $d_{i,r-i} = -q_a$ for $i < s$, $a_{i,r-i} = q_{a'}$ for $i > s$ and $d_{s,r-s} \in [-q_a, q_{a'}]$ solving

$$\sum_{i=\max\{0,r-m_{a'}\}}^{s-1} \binom{m_a}{i}\binom{m_{a'}}{r-i}(-q_a) + \binom{m_a}{s}\binom{m_{a'}}{r-s} d_{s,r-s} + \sum_{i=s+1}^{m_a} \binom{m_a}{i}\binom{m_{a'}}{r-i} q_{a'} = 0$$

Setting up the Lagrangian as above and taking the derivative we obtain

$$\begin{aligned} \frac{dL}{da_{i,r-i}} &= -\binom{m_a}{i}\binom{m_{a'}}{r-i} w_{i,r-i} - \lambda_i + \lambda_r \frac{1}{\binom{m_{a'}}{r-m_a}} \binom{m_a}{i}\binom{m_{a'}}{r-i} \\ &\quad + \tau_i - \tau_r \frac{1}{\binom{m_{a'}}{r-m_a}} \binom{m_a}{i}\binom{m_{a'}}{r-i}. \end{aligned}$$

This is solved by $\lambda_s = \tau_s = 0$, $\lambda_i = 0$ for $i < s$ and $\tau_i = 0$ for $i > s$, $\lambda_r = \binom{m_{a'}}{r-m_a} w_{s,r-s}$, $\tau_i = \binom{m_a}{i}\binom{m_{a'}}{r-i}(w_{i,r-i} - w_{s,r-s}) \geq 0$ for $i < s$ and $\lambda_i = \binom{m_a}{i}\binom{m_{a'}}{r-i}(w_{s,r-s} - w_{i,r-i}) \geq 0$ for $i > s$. This completes the proof.

12.2 Parameters of the Dominant Rule in Small Samples

We present the values of s and θ for the dominant linear non-guessing rule. We first do this for the uninformative pseudo prior.

n	2	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5
m_a	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	2	2
r	1	1	2	1	2	3	1	2	3	1	2	3	4	1	2	3	4
s_r	0	0	1	0	1	1	0	1	1	0	0	1	1	0	1	1	2
θ_r	0	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{3}$	1	$\frac{2}{3}$	0	$\frac{2}{3}$	0	$\frac{3}{8}$	$\frac{1}{6}$	$\frac{5}{6}$	$\frac{5}{8}$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{5}{6}$

Next we present the parameters when the pseudo prior assigns probabilities that are equal to the frequencies in the data set.

n	2	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5
m_a	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	2	2
r	1	1	2	1	2	3	1	2	3	1	2	3	4	1	2	3	4
s_r	0	0	1	0	1	1	0	1	1	0	1	1	1	0	1	1	2
θ_r	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{2}{3}$

On the side, note that the values of s_r are the same as under the uninformative pseudo prior except when $n = 5$, $m_a = 1$ and $r = 2$. This is due to the coarseness of the threshold for small sample sizes.

13 Proposition and Proof for Section 5.5

Proposition 24 *(i) If* $E_G\left(f^*\left(\bar{X}\right)\right)_{\bar{a}} \geq (1-\kappa_{\bar{a}})\, q_{\bar{a}} + \kappa_{\bar{a}}$ *holds for* $\bar{a} \in \{a, a'\}$ *then* $E\left(f^d\left(\bar{X}\right)\right) \cdot \mu^G \geq q \cdot \mu^G$.

(ii) For every $d \in (0,1)$ *there exists* $\bar{m}$ *such that if* $m_a, m_{a'} \geq \bar{m}$ *and* $\left|\mu_a^G - \mu_{a'}^G\right| \geq d$ *then* $E_G\left(f^d\left(\bar{X}\right)\right) \cdot \mu^G \geq q \cdot \mu^G$.

(iii) $\max\left\{\mu_a^G, \mu_{a'}^G\right\} - E_G\left(f^d\left(\bar{X}\right)\right) \cdot \mu^G \leq \frac{1}{\min\{\kappa_a, \kappa_{a'}\}}\left(\max\left\{\mu_a^G, \mu_{a'}^G\right\} - E_G\left(f^*\left(\bar{X}\right)\right) \cdot \mu^G\right).$

Proof. Consider first some random variable Y with distribution P_F and some $z < 1$. Then

$$EY = \int_{y>z} y dP_F(y) + \int_{y \leq z} y dP_F(y) \leq P_F(Y > z) + z\left(1 - P_F(Y > z)\right)$$

and hence

$$P_F(Y \geq z) \geq P_F(Y > z) \geq \frac{1}{1-z}(EY - z).$$

Now consider part (i). Assume $\mu_a^G > \mu_{a'}^G$. Let $dP^G(X) = \prod_{i=1}^m dG_{a_i}(x_i)$. From the above we obtain

$$E\left(f^d\left(\bar{X}\right)\right)_a = P^G\left(f^*(X)_a \geq \kappa_a\right) \geq \frac{1}{1-\kappa_a}\left(E_G\left(f^*(X)\right) - \kappa_a\right).$$

So $E_G\left(f^d\left(\bar{X}\right)\right) \cdot \mu^G \geq q \cdot \mu^G$ if

$$E_G\left(f^d\left(\bar{X}\right)\right)_a - q_a \geq \frac{1}{1-\kappa_a}\left(E_G\left(f^*(X)\right) - \kappa_a\right) - q_a \geq 0$$

if

$$E_G\left(f^*\left(\bar{X}\right)\right)_a \geq (1-\kappa_a)\, q_a + \kappa_a.$$

The analogous condition is found similarly for a' which completes the proof of part (i).

We do not have to prove part (ii) as it follows from part (i) and Proposition 17.

For part (iii) assume without loss of generality that $\mu_a^G > \mu_{a'}^G$. Then

$$\begin{aligned} E_G\left(f^d\left(\bar{X}\right)\right) \cdot \mu^G &= E_G\left(f^d\left(\bar{X}\right)\right)_a \mu_a^G + E_G\left(f^d\left(\bar{X}\right)\right)_{a'} \mu_{a'}^G \\ &= E_G\left(f^d\left(\bar{X}\right)\right)_a \left(\mu_a^G - \mu_{a'}^G\right) + \mu_{a'}^G \\ &\geq \frac{1}{1-\kappa_a}\left(E_G\left(f^*(X)\right)_a - \kappa_a\right)\left(\mu_a^G - \mu_{a'}^G\right) + \mu_{a'}^G \end{aligned}$$

so

$$\begin{aligned}
\mu_a^G - E_G\left(f^d\left(\bar{X}\right)\right)\cdot\mu^G &\leq \mu_a^G - \mu_{a'}^G - \frac{1}{1-\kappa_a}\left(E_G\left(f^*\left(X\right)\right)_a - \kappa_a\right)\left(\mu_a^G - \mu_{a'}^G\right) \\
&= \frac{1}{\kappa_{a'}}\left(\mu_a^G - \mu_{a'}^G\right)E_G\left(f^*\left(X\right)\right)_{a'} \\
&= \frac{1}{\kappa_{a'}}\left(\mu_a^G - E_G\left(f^*\left(\bar{X}\right)\right)\cdot\mu^G\right)
\end{aligned}$$

which completes the proof. ■

References

[1] Bergemann, D., and S. Morris (2005). Robust Mechanism Design. Econometrica, 73(6), 1771–1813.

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

[3] Cesa-Bianchi, N., and G. Lugosi (2006), *Prediction, Learning, and Games*, Cambridge University Press.

[4] Chen, H. and P. Guggenberger (2025), A Note on Minimax Regret Rules with Multiple Treatments in Finite Samples, Econometric Theory (forthcoming).

[5] Fabijan, A., P. Dmitriev, H. Holmstrom Olsson and J. Bosch (2018), Online Controlled Experimentation at Scale: An Empirical Survey on the Current State of A/B Testing. 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Prague, Czech Republic, 68-72.

[6] Gossner, O., and K.H. Schlag (2013), Finite-Sample Exact Tests for Linear Regressions with Bounded Dependent Variables, J. Econometrics 177, 75-84

[7] Hofbauer, J., and K.H. Schlag (2000) Sophisticated Imitation in Cyclic Games. *J. Evol. Econ.* 10, 523–543.

[8] Huber, P. (1972), "The 1972 Wald Lecture Robust Statistics: A Review." Ann. Math. Statist. 43(4) 1041-1067, August, 1972.

[9] Lakshmivarahan, S., and M. Thathachar (1973), Absolutely Expedient Learning Algorithms for Stochastic Automata, *IEEE Transactions on Systems, Man and Cybernetics* 3, 281-286.

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

[11] Merriam-Webster Dictionary (2024), Merriam-Webster, https://www.merriam-webster.com.

[12] Rustichini, A. (1999), Optimal Properties of Stimulus-Response Learning Models, Games and Economic Behavior, 29(1–2), 244-273.

[13] Samuelson, L. J. Zhang (1992), Evolutionary Stability in Asymmetric Games, J. Economic Theory 57(2), 363-391.

[14] Savage, L.J. (1951). The Theory of Statistical Decision. *J. American Statistical Association*, 46(253), 55-67.

[15] Schlag, K.H. (1998) Why Imitate, and If So, How?: A Boundedly Rational Approach to Multi-Armed Bandits, *J. Economic Theory* 78(1), 130-156

[16] Schlag, K.H. (1999), Which One Should I Imitate? *J. Mathematical Economics* 31(4) 493-522.

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

[18] Schlag, K.H., (2006a), Designing Non-Parametric Estimates and Tests for Means. Economics Working Papers ECO 2006/26, European University Institute, Florence.

[19] Schlag, K.H. (2006b), ELEVEN Tests needed for a Recommendation, EUI Working Paper 2006/2.

[20] Stoye, J (2009), Minimax Regret Treatment Choice with Finite Samples, *J. Econometrics* 151, 70-81.

[21] Taguchi, G., Phadke, M.S. (1989). Quality Engineering through Design Optimization. In: Dehnad, K. (eds) Quality Control, Robust Design, and the Taguchi Method (ch 5). Springer, Boston, MA., 77-96.

[22] Tetenov, A. (2012), Statistical Treatment Choice Based on Asymmetric Minimax Regret Criteria, J. Econometrics 166, 157-165.

[23] Tocher, K.D. (1950), Extension of the Neyman-Pearson Theory of Tests to Discontinuous Variates, *Biometrika* 37, 130-144.

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