European University Institute

Department of Economics

EUI Working Papers

ECO No. 2007/1

Distribution-Free Learning

Karl H. Schlag

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Distribution-Free Learning

KARL H. SCHLAG

EUI Working Paper **ECO** No. 2007/1

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

Distribution-Free Learning

Karl H. Schlag

European University Institute[1]

December 30, 2006

[1]Economics Department, European University Institute. Via della Piazzuola 43, 50133 Florence, Italy, schlag@eui.eu

Abstract

We select among rules for learning which of two actions in a stationary decision problem achieves a higher expected payoff when payoffs realized by both actions are known in previous instances. Only a bounded set containing all possible payoffs is known. Rules are evaluated using maximum risk with maximin utility, minimax regret, competitive ratio and selection procedures being special cases. A randomized variant of fictitious play attains minimax risk for all risk functions with ex-ante expected payoffs increasing in the number of observations. Fictitious play itself has neither of these two properties. Tight bounds on maximal regret and probability of selecting the best action are included.

Keywords: fictitious play, nonparametric, finite sample, matched pairs, foregone payoffs, minimax risk, ex-ante improving, selection procedure.

JEL classification numbers: D83, D81, C44.

1 Introduction

How can we learn from a given finite sample or learn over time from own previous experience without relying on some prior? We investigate this issue in a stationary decision problem with two ex-ante identical actions, each generating a random payoff from an unknown distribution, in which one knows what each action would have realized in some previous instances. The objective is to choose the action that yields the higher expected payoff. Examples include invest or not invest, buy or not buy, forecasting which of two states occurs next, comparing well-being before and after a treatment, and cooperate or defect when acting as if opponents are non strategic. We present a "universal" nonparametric distribution-free learning rule when there is a known bounded set that contains any payoff.

Learning is traditionally associated to the process of updating some prior when new information arrives using Bayes' rule. Anscombe and Aumann (1963) present axioms that imply that preferences are based on some subjective prior. This prior is deduced from preferences but not from the primitives of the model. Different priors can lead to different choices. The flexibility of the model results in ambiguous predictions. One cannot speak of learning in terms of inference. Justification of choice in front of others can be problematic as the subjective prior cannot be deduced from observables.

In this paper we would like to investigate individual learning that does not rely on such a prior. We focus on the extreme case in which the two actions are ex-ante identical so that there is no initial information that can be used to compare the two actions. Three alternative decision making criteria that do not involve priors have been axiomatized: *maximin utility*, *minimax regret* and *relative minimax*. The first two, due to Wald (1950) and Savage (1951) respectively, have both been axiomatized by Milnor (1954) and recently also by Stoye (2006). The Symmetry Axiom plays a central role in ensuring that choice only depends on observables as it postulates that choices are not allowed to depend on labels. Relative minimax has been introduced and axiomatized by Terlizzese (2006) and also satisfies the Symmetry Axiom. An alternative is to focus on the probability of selecting the best action as in the literature on selection procedures starting with Sobel and Huyett (1957). One may wish to capture learning in a repeated decision making context by postulating that expected payoffs should be increasing between rounds conditional on the previous round (*ab-*

solute expediency, Lakshmivarahan and Thathachar, 1973) or from an ex-ante point of view before making the first choice (*ex-ante improving*, Schlag, 2002, cf. Börgers et al. 2004). Learning without priors is very common in machine learning, artificial intelligence and computer science. Classical statistics and a large part of statistical decision theory is based on making choices without assessing priors.

To put our work in perspective we review some concepts. Our model is *distribution-free* as we make no assumptions on the underlying distributions apart from specifying conditions on its support. For instance one can impose the minimal condition that payoffs belong to a given bounded interval. In this case our approach is also *nonparametric* (Fraser, 1957) as the set of possible distributions is then infinitely dimensional. While *social learning* includes gathering information from others, we focus on *individual learning* that only involves using information from own previous experience. Following Rustichini (1999), *partial information* refers to learning from own previous payoffs while *full information* assumes that the payoff of each action is observed in each realization (also called *matched pairs* or *paired data* in statistics). In the context of repeated decision making, the full information setting is also called learning from *foregone payoffs*.

Only little is known about how to make choices when learning without priors, in particular if the "how" is assessed according to some formal decision criterion. Specifically we are interested in stationary decision problems: games against nature as opposed to games against other players. One may also choose to treat games against others as games against nature by ignoring the strategic aspect of opponent's play. There has been some research concerned with learning in the limit where average payoffs are evaluated once an infinite amount of data has been gathered (Robbins, 1952, Rustichini, 1999), some studying rates of convergence (e.g. Lai and Robbins, 1985). It is astonishing how much attention has been given to such studies on learning in the limit without even incorporating concern for uniform convergence in view of the fact that typically data is either limited and the decision maker has a minimal degree of impatience. We are interested in learning from finite samples as well as in learning over time from own previous experience.

Some results on learning in stationary decision problems are available for the partial information setting. Canner (1970) shows how to choose under minimax regret when facing a sample in which each action has been sampled equally often. Schlag (2006a) derives a rule that attains minimax regret for a given sample size when the

rule also specifies which action to sample. Berry and Fristedt (1985) and Auer et al. (2002) derive upper bounds on the minimax regret of a patient decision maker learning over time. Schlag (2003) identifies rules involving commitment that attain minimax regret when the discount factor is not too large. Börgers et al. (2004) select among the absolutely expedient learning rules that have a single round of memory.

For the full information setting, results have only been obtained for learning in the limit. *Fictitious play* (Brown, 1949), that specifies to choose an action that achieved the highest empirical average payoff, will almost surely select which action is best (in terms of achieving the highest expected payoff) in the long run. Other prominent rules with this property are the exponential adjustment process of Rustichini (1999) and regret matching of Hart and Mas-Colell (2000).

This is the first paper that investigates distribution-free learning within the full information setting beyond the case of learning in the limit. Note that the full information setting is more tractable than the partial information setting as there is no trade-off between exploration and exploitation. The information gathered does not depend on previous choices.

The formal criterion for selection in this paper is *minimax risk* (Wald, 1950) where the underlying loss function is assumed to be symmetric and to only depend on the expected payoffs. All common distribution-free decision making criteria are included: competitive ratio (Borodin and El-Yaniv, 1998), relative minimax, minimax regret, maximin utility as well as maximizing the minimal probability of selecting the best action conditional on a minimal difference between the two means (Sobel and Huyett, 1957). We also evaluate whether rules are ex-ante improving or absolutely expedient.

Assume first that payoffs are restricted to be binary valued, e.g. an outcome can only either be a success or a failure. Then we find that fictitious play attains minimax risk for any given number of observations and that fictitious play is ex-ante improving. This is the first formal foundation of fictitious play for finite samples without using priors.[1] However once payoffs can equal one of three different values then fictitious play looses its nice properties due to its ignorance of small differences in payoffs. There is a loss function, defined by the objective of not choosing the worse action, such that fictitious play does not attain minimax risk for any $n \in \mathbb{N}$. Moreover fictitious play is no longer ex-ante improving. In fact, for any given number m we

[1] The only previous result for finite samples is due to Fudenberg and Kreps (1990) who show that fictitious play is a best response to a Dirichlet prior.

present an example with support on three outcomes in which the expected weight under fictitious play on the better action is strictly decreasing until m choices have been made. Of course, as mentioned above, fictitious play almost surely chooses the best action in the limit.

We are genuinely interested in making choices when each action can yield more than two possible payoffs. We assume that there is a known bounded interval that contains all payoffs that can be generated by either action. Knowledge on the underlying support can be included and our results apply when two conditions are fulfilled. (a) The two actions are ex-ante identical. (b) For any distribution there is always one with the same means that only has support on the extreme payoffs in the given bounded interval. The rule we select is a variant of fictitious play we call *binomial fictitious play*. According to this rule, observed payoffs are first randomly transformed into a binary sequence before choosing according to fictitious play. We show that binomial fictitious play attains minimax risk for any underlying loss function. Moreover, binomial fictitious play achieves lower maximal risk than fictitious play among all distributions that have the same underlying means. In addition, binomial fictitious play is ex-ante improving and almost surely chooses the better action in the limit. One may argue that these results justify to call binomial fictitious play a universal learning rule.

Mathematically our selection result combines two findings. Fictitious play is a best response to any symmetric prior when payoffs are binary valued and there are only two actions. The randomization technique demonstrated in Schlag (2006b) that first independently appeared in Cucconi (1968) in a statistical application and in Schlag (2003) for decision making shows how to transforms exact results for the binary case into exact results for the nonparametric setting.[2]

We also provide some insights into the class of rules that attain minimax risk when only a bounded interval containing all payoffs is known. Some decision making criterion such as maximin utility have no prediction power. Choosing each action equally likely and thus ignoring information given by the sample attains maximin utility. More valuable insights are gained for loss functions that *reflect learning when learning matters* in the sense that loss is only above the minimum level if both actions do not yield the same expected payoff.[3] The minimax regret criterion is based on a

[2] "Exact" is a term from statistics that refers to claims that are not based on asymptotic results.

[3] We do not allow for concern for lower variance provided means do not differ. Note that ax-

loss function that has this property. For such loss functions we show that in order for an alternative rule to perform similarly well as binomial fictitious play it must behave like (binomial) fictitious play when facing a sample consisting only of binary valued payoffs in which empirical averages of the two actions do not coincide.

Performance of binomial fictitious play is illustrated in two settings. The value of minimax regret is presented for a given sample size and in the repeated decision making setting, building on results in Schlag (2006a). The maximal probability of guaranteeing selection of the best action conditional on the two means being sufficiently different is cited from Schlag (2006b). We show that the rule that turns out to attain minimax risk in these two settings, the *binomial average rule*, is not surprisingly dominated by binomial fictitious play as the binomial average rule is designed for partial information and its minimax risk property under full information is only a side product in these two papers.

The power of our results stems from the fact that we limit attention to a very simple decision problem. Once there are more than two actions then fictitious play is no longer a best response to any symmetric prior over binary valued payoff distributions. It is only a best response when actions are known to be independent. With the techniques used in this paper one can only prove that binomial fictitious play achieves minimax regret when payoff distributions associated to actions are known to be independent.

The presentation proceeds as follows. In Section 2 we present the setup. In Section 3 we present binomial fictitious play and compare it to fictitious play. Section 4 contains the analysis of minimax risk. In Section 5 we consider the two-armed bandit setting and derive the value of minimax regret for two alternative specifications of time preferences. In Section 6 we briefly consider more than two actions, in Section 7 we conclude. In the appendix we describe the connection between minimax risk and hypothesis testing.

---

iomatized criteria such as minimax regret and maximin utility capture any concern for minimizing dispersion by first translating outcomes into utility. These utilities should then be identified with what we call payoffs.

2 Choice between Two Actions

Consider two actions $i = 1, 2$ where action i is associated to a random variable Y_i and $Y = (Y_1, Y_2)$ is distributed according to the distribution P. Assume that outcomes realized by the two random variables Y_1 and Y_2 are known to be contained in $[\alpha, \omega] \subset \mathbb{R}$ with $\alpha < \omega$, but that the joint distribution $P \in \Delta [\alpha, \omega]^2$ is otherwise unknown. ΔA denotes the set of all distributions with support in A. P^s will denote the distribution that results from permuting the labels of the actions in P, for finite support this means that $P(y_1, y_2) = P^s(y_2, y_1)$ for all $y \in [\alpha, \omega]^2$.

Consider a decision maker facing a single decision, to choose either action 1 or action 2. Choice of action i yields a payoff that is drawn from P_i where P_i is the marginal distribution of P with respect to component i, $i = 1, 2$. We assume that the decision maker would like to choose the action that has the highest mean, such an action will be called *best*. However this is not possible as we assume that P is unknown. Let $\mu_i = E_P(Y_i)$ denote the mean payoff generated by action i under distribution P, $i = 1, 2$. None of our results will be affected if payoffs are transformed affinely, hence we assume without loss of generality that payoffs are known to be contained in $[0, 1]$ (transform y_i into $\frac{y_i - \alpha}{\omega - \alpha}$). Let $[1]$ denote the index of action with higher mean when $\mu_1 \neq \mu_2$ so $\mu_{[1]} = \mu_i$ if $\mu_i > \mu_{3-i}$. We say that a payoff y is *binary valued* if $y \in \{0, 1\}$ and that P is *binary valued* if $P \in \Delta \{0, 1\}^2$.

In a more formal choice setting, actions would generate random outcomes where outcomes would belong to some given set of outcomes. Outcomes would then be transformed into utilities and we would assume that these belong to $[\alpha, \omega]$. Utilities would then be identified with payoffs.

Assume that the decision maker has observed n independent realizations of Y denoted by $y^1, .., y^n$ so $y^k = \left(y_1^k, y_2^k\right) \in [\alpha, \omega]^2$ is an independent random drawn of a pair of outcomes based on the distribution P. It is important to stress that each data point y^k consists of the payoff that each action achieved, in statistics one also speaks of *matched pairs* or of *paired data*. Let $y^{1,n} = (y^1, .., y^n)$ denote the sequence of observations. n will be called the *sample size*. Rustichini (1999) calls this the full information case. In the sequential choice setting we consider later this is called learning from *foregone* payoffs.

The decision maker can condition her choice on the observed realizations of Y.

Her rule or *strategy* σ is hence a function $\sigma : [0,1]^{2n} \to \Delta\{1,2\}$ where $\sigma_i(y^{1,n})$ is the probability of choosing action i after observing $y^{1,n}$. A strategy is called *symmetric* if its behavior does not depend on how actions are labelled. The set of all strategies is denoted by Σ. Let $E_P\sigma_i(y^{1,n})$ denote the ex-ante expected probability of choosing action i based on n independent observations to be gathered. Let $E_P\pi(\sigma, y^{1,n}) = E_P\sigma_1(y^{1,n})\mu_1 + E_P\sigma_2(y^{1,n})\mu_2$ denote the associated ex-ante expected payoff.

3 Two Strategies

Two particular strategies will be of interest for our analysis, fictitious play and binomial fictitious play.

The strategy σ^f called *fictitious play* satisfies $\sigma_i^f(y^{1,n}) = 1$ if $\frac{1}{n}\sum_{k=1}^n y_i^k > \frac{1}{n}\sum_{k=1}^n y_{3-i}^k$ and $\sigma_i^f(y^{1,n}) = \frac{1}{2}$ if $\frac{1}{n}\sum_{k=1}^n y_1^k = \frac{1}{n}\sum_{k=1}^n y_2^k$ or if $n = 0$. In words, the decision maker chooses the action that performed best in the past, choosing each action equally likely in round one or when both actions achieved the same average payoffs. Notice that we choose a particular representative of fictitious play. (i) There are is no initial weights on either action and hence only observed payoffs influence future behavior. (ii) The most popular tie breaking rule is assumed, namely each action is chosen equally likely whenever the empirical averages are equal.

The strategy σ is called *binomial* if $\sigma(y^{1,n}) = \sigma\left(\left(t\left(y^k\right)\right)_{k=1}^n\right)$ for some (random) transformation $t : [0,1]^2 \to \Delta\{0,1\}^2$ that is *mean preserving* in the sense that it satisfies $y_i = \Pr(t_i(y) = 1)$ for $i = 1, 2$. The strategy σ^b is called *binomial fictitious play* if it is the binomial strategy that chooses each action equally likely in round one and coincides with fictitious play whenever only binary valued payoffs have been observed. Thus binomial fictitious play is characterized by the underlying mean preserving transformation t.

Notice that any mean preserving transformation is equal to the identity on $\Delta\{0,1\}^2$. Two particular mean preserving transformations will play a role in our later analysis. The *independent transformation* t^I is the unique) mean preserving transformation that transforms the outcome of each action independently, so $t^I(y) = t_1^I(y_1) * t_2^I(y_2)$. The *correlated transformation* t^C is characterized by being the unique mean preserving transformation that preserves the ranking within the pair of outcomes in the sense that $\Pr\left(t^C(y) = (1,0)\right) = 0$ if $y_1 \leq y_2$ and $\Pr\left(t^C(y) = (0,1)\right) = 0$ if $y_1 \geq y_2$.

Uniqueness follows as these two conditions imply

$$\begin{aligned} \Pr\left(t^C(y) = (0,0)\right) &= 1 - \max\{y_1, y_2\} \\ \Pr\left(t^C(y) = (1,1)\right) &= \min\{y_1, y_2\} \ . \end{aligned}$$

An alternative characterization of the correlated transformation is that it maximizes $\Pr(t(y) = (1,1))$ among all mean preserving transformations. This follows from from the next two statements.

$$\Pr(t(y) = (1,1)) \leq \min\{\Pr(t_1(y) = 1), \Pr(t_2(y) = 1)\} = \min\{y_1, y_2\}$$

holds for all mean preserving transformations t. Moreover, it is easily verified that the correlated transformation is the unique mean preserving transformation t that satisfies $\Pr(t(y) = (1,1)) = \min\{y_1, y_2\}$. Similarly it is easily verified that the correlated transformation is the unique mean preserving transformation that maximizes the covariance of $t_1(y)$ and $t_2(y)$ for $y_1, y_2 \in (0,1)$.

When we wish to identify that binomial fictitious play is based on one of the two transformations t^I or t^C we will write σ^{bI} or σ^{bC} respectively.

In the following two subsections we investigate the behavior of binomial fictitious play and compare this to that of fictitious play.

3.1 Conditional Choice

We illustrate how sensitive binomial fictitious play is to empirical success measured in terms of average payoffs realized is incorporated. This is in the light of the fact that fictitious play, by definition, puts all weight on the empirically more successful action.

Assume $n = 1$. We find

$$\begin{aligned} \sigma_1^b(y) &= \Pr(t(y) = (1,0)) + \frac{1}{2}\Pr(t_1(y) = t_2(y)) \\ \sigma_1^b(y) - \sigma_2^b(y) &= \Pr(t(y) = (1,0)) - \Pr(t(y) = (0,1)) = y_1 - y_2 \end{aligned}$$

and hence

$$\sigma_1^b(y) = \frac{1}{2} + \frac{1}{2}(y_1 - y_2) \tag{1}$$

holds independent of the specific mean preserving transformation. Binomial fictitious play σ^b is more likely to select the action that yielded the higher payoff (than the

one that yielded the lower payoff). The probability placed on the empirically more successful action is increasing in the difference between the success of the two actions. So behavior under binomial fictitious play and fictitious play can be very different, in particular it is easily verified that $\sup_y \left|\sigma_1^b(y) - \sigma_1^f(y)\right| = 1/2$.

Consider now $n = 2$ with binomial fictitious play based on the independent transformation. Assume that $y^{1,2} = ((\eta, 0), (\eta, 2\eta))$ for some $\eta \in (0, 1/2)$ so $y_1^1 + y_1^2 = y_2^1 + y_2^2$. It is easily verified that $\sigma_1^{bI}(y^{1,2}) = \frac{1}{2} - \frac{1}{2}\eta^2(1 - 2\eta)$. So we find that while both actions were empirically equally successful, the decision maker using σ^{bI} is more likely to choose action 2 than action 1.

Assume instead correlated transformation. Clearly if $y_1^n \geq y_2^n$ for $n = 1, 2$ then $\sigma_1^{bC}(y^{1,2}) \geq \frac{1}{2}$. Assume that $y_1^1 > y_2^1$ and $y_1^2 < y_2^2$. Then

$$\begin{aligned}\sigma_1^{bC}\left(y^{1,2}\right) &= \left(y_1^1 - y_2^1\right)\left(1 - \frac{1}{2}\left(y_2^2 - y_1^2\right)\right) + \frac{1}{2}\left(1 - \left(y_1^1 - y_2^1\right)\right)\left(1 - \left(y_2^2 - y_1^2\right)\right) \\ &= \frac{1}{2} + \frac{1}{2}\left(y_1^1 + y_1^2 - y_2^1 - y_2^2\right).\end{aligned}$$

Consequently, $\sigma_i^{bC}(y^{1,2}) \geq \frac{1}{2}$ holds for all $y^{1,2}$ such that $y_i^1 + y_i^2 \geq y_{3-i}^1 + y_{3-i}^2$. The empirically more successful action of the first two rounds is chosen more likely in the third round.

We investigate correlated transformation when $n = 3$. It is easily verified when $y^{1,3} = ((\eta, 1), (\eta, 0), (1 - 2\eta, 0))$ for some $\eta \in (0, 1/2)$ that $\sigma_1^{bC}(y^{1,3}) = \frac{1}{2} - \frac{1}{2}\eta^2(1 - 2\eta)$ which means that action 2 is chosen more likely despite the fact that both actions are empirically equally successful.

We summarize.

Remark 1 *Binomial fictitious play puts more weight on the empirically more successful action (i) under the independent transformation when $n = 1$ and (ii) under the correlated transformation when $n \leq 2$. Statements (i) and (ii) are not necessarily true for other values of n.*

Finally we briefly investigate whether one of these two rules is absolutely expedient. A rule is called *absolutely expedient* if expected payoffs conditional on the mixed action in the previous round are increasing. Assume that $(1, 0)$ occurred in the first round. Then fictitious play and binomial fictitious play specify to choose action 1 in round 2. As it is possible that action 1 is the best action and the weight on the best

action has to be weakly increasing, an absolute expedient rule will choose action 1 also in all later rounds. Thus neither fictitious play nor binomial fictitious play are absolutely expedient.

3.2 Unconditional Choice

We investigate expected behavior of the two rules where expectations are calculated ex-ante before making any observations.

Assume $n = 1$. It follows directly from (1) that

$$E_P\sigma_1^b(y) = \frac{1}{2}(1 + \mu_1 - \mu_2) \tag{2}$$

and hence that

$$\begin{aligned} E_P\pi\left(\sigma^b, y\right) &= \frac{1}{2}(1 + \mu_1 - \mu_2)\mu_1 + \frac{1}{2}(1 + \mu_2 - \mu_1)\mu_2 \\ &= \frac{1}{2}(\mu_1 + \mu_2) + \frac{1}{2}(\mu_1 - \mu_2)^2. \end{aligned}$$

Now consider fictitious play facing P such that $P(0, x) + P(1, x) = 1$ with $1/2 < \mu_1 < x = \mu_2 < 1$. Then $E_P\sigma_1^f(y) = \mu_1$ which implies that more weight is put on the worse action. As

$$E_P\pi\left(\sigma^f, y\right) = \frac{1}{2}(\mu_1 + \mu_2) + \left(\frac{1}{2} - \mu_1\right)(\mu_2 - \mu_1) \tag{3}$$

we find that fictitious play performs worse than the rule that prescribes to choose each action equally likely.

Consider now $n = 2$. Let $p_{10} = \int \Pr(t(y) = (1, 0))\, dP(y)$ and $p_{01} = \int \Pr(t(y) = (0, 1))\, dP(y)$. Hence $p_{10} - p_{01} = \mu_1 - \mu_2$. Then

$$\begin{aligned} E_P\sigma_1^b\left(y^{1,2}\right) &= p_{10} + (1 - p_{10} - p_{01})\left(p_{10} + \frac{1}{2}(1 - p_{10} - p_{01})\right) \\ &= \frac{1}{2} + \frac{1}{2}(p_{10} - p_{01})(2 - p_{10} - p_{01}) \\ &= \frac{1}{2} + \frac{1}{2}(\mu_1 - \mu_2)(2 - p_{10} - p_{01}) \\ &= \frac{1}{2} + \frac{1}{2}(\mu_1 - \mu_2)(2 - \mu_1 - \mu_2 + 2p_{11}) \end{aligned} \tag{4}$$

as $p_{01} + p_{10} = \mu_1 + \mu_2 - 2p_{11}$. In particular, the correlated transformation maximizes the probability of choosing the best action and the expected payoff in round 3 among all mean preserving transformations. Note that

$$E_P\sigma_1^b\left(y^{1,2}\right) = E_P\sigma_1^b(y) + (1 - p_{10} - p_{01})\frac{1}{2}(\mu_1 - \mu_2) \tag{5}$$

as there is only value to observing payoffs in round 2 if either $(0,0)$ or $(1,1)$ is realized in round 1. We obtain

$$E_P\pi\left(\sigma^b, y^{1,2}\right) = \frac{1}{2}\left(\mu_1+\mu_2\right) + \frac{1}{2}\left(\mu_1-\mu_2\right)^2 + \left(1-\mu_1-\mu_2+2p_{11}\right)\frac{1}{2}\left(\mu_1-\mu_2\right)^2.$$

As specific formulae for $n \geq 3$ are involved, we demonstrate qualitative properties in terms of how ex-ante expected payoffs change in the number of observations available.

Proposition 1 *(ia) Fictitious play is not ex-ante improving, specifically for any $m \in \mathbb{N}$ there exists $P \in \Delta[0,1]^2$ such that $E_P\pi\left(\sigma^f, y^{1,n}\right)$ is strictly decreasing in n for $n \leq m$.*

(ib) Binomial fictitious play is ex-ante improving, specifically $E_P\pi\left(\sigma^b, y^{1,n+2}\right) > E_P\pi\left(\sigma^b, y^{1,n}\right)$ if $\mu_1 \neq \mu_2$.

(ii) For any $d > 0$ both fictitious play and binomial fictitious play are uniformly consistent estimators of the best action if $|\mu_1 - \mu_2| \geq d$, specifically for any $\varepsilon > 0$ there exists n_0 such that $\mu_i \geq \mu_{3-i} + d$ and $n \geq n_0$ implies $E_P\sigma_i^b\left(y^{1,n}\right) \geq 1-\varepsilon$.

Proof. Part (ia). Given the properties of a mean preserving transformation it is sufficient to restrict attention to $P \in \Delta\{0,1\}^2$. Fix any binary valued P and consider a rational decision maker with prior $\bar{Q} \in \Delta\left(\Delta\{0,1\}^2\right)$ that puts equal weight on P and P^s. Then it is easily verified using Bayes rule that fictitious play is a best response to $\bar{Q}$. As more information cannot be harmful and $E_Q\pi\left(\sigma^b, y^{1,n}\right) = E_P\pi\left(\sigma^b, y^{1,n}\right)$ it follows that $E_P\pi\left(\sigma^b, y^{1,n}\right)$ is weakly increasing in n.

Consider n odd. If $\mu_1 \neq \mu_2$ then two more observations can change action 1 from being the unique best response to making it strictly worse than action 2 when facing $\bar{Q}$. Hence, $E_P\pi\left(\sigma^b, y^{1,n+2}\right) > E_P\pi\left(\sigma^b, y^{1,n}\right)$ if $\mu_1 \neq \mu_2$.

Part (ib). For any given $m \in \mathbb{N}$ choose $x < \frac{1}{m}$ and consider P such that $P(0,x) + P(1,x) = 1$ and $1/2 < \mu_1 < x = \mu_2$. For $n \leq m$ we then obtain $\sigma_1^f\left(y^{1,n}\right) = 1-(1-\mu_1)^n$ which is strictly increasing in n.

Part (ii). The law of large numbers implies that $\lim_{n\to\infty} E_P\sigma_1^b\left(y^{1,n}\right) = \lim_{n\to\infty} E_P\sigma_1^f\left(y^{1,n}\right) = 1$ whenever $\mu_i > \mu_{3-i}$. ■

Part (ii) shows that both fictitious play and binomial fictitious play are uniformly consistent estimators of the best action provided the two means differ by at least d.

Parts (ia) and (ib) show that binomial fictitious play is ex-ante improving while fictitious is not where a rule σ is called *ex-ante improving* (Schlag, 2002, cf. Börgers et al., 2004) if $E_P\pi(\sigma, y^{1,n+1}) \geq E_P\pi(\sigma, y^{1,n})$ holds for all n and all P.

Notice that binomial fictitious play does not induce strictly increasing ex-ante payoffs when $\mu_1 \neq \mu_2$ as $E_P\pi\left(\sigma^b, y^{1,n+1}\right) = E_P\pi\left(\sigma^b, y^{1,n}\right)$ holds when n is odd and $P(\{(1,0),(0,1)\}) = 1$. This is because any best response against $\bar{Q}$ after an odd number of rounds remains a best response when one more observation is added ($\bar{Q}$ defined in the proof of Proposition 1 (ia)).

We present some differences between fictitious play and binomial fictitious play in terms of the probability of choosing the best action.

Corollary 1 *Assume $\mu_1 \neq \mu_2$.*
(i) $E_P\sigma^b_{[1]}(y^{1,n}) > \frac{1}{2}$ for $n \geq 1$.
(ii) For any $n \geq 1$ there exists $P \in \Delta[0,1]^2$ such that $E_P\sigma^f_{[1]}(y^{1,n}) \leq e^{-1}$.[4]

Proof. Part (i) follows directly from (2) and Proposition 1(ia).

Part (ii). In the example used to prove Proposition 1(ib) we find that $\lim_{\mu_1 \to x \to 1/m} \sigma^f_1(y^{1,m}) = 1 - \left(1 - \frac{1}{m}\right)^m > 1 - e^{-1}$ where 1 is the worse action which proves the statement. ■

4 Minimax Risk

We now present a methodology for selecting a strategy. The methodology enables to model a decision maker that knows more about the possible underlying distributions.

Let $\mathcal{P} \subseteq \Delta[0,1]^2$ be the closure of the set of all distributions that the decision maker cannot rule out based on her a priori knowledge of the choice setting. Thus, by definition $\mathcal{P}$ is closed. We make the following two additional assumptions on $\mathcal{P}$. (a) $\mathcal{P}$ is symmetric in the sense that if $P \in \mathcal{P}$ then $P^s \in \mathcal{P}$. (b) For every $P \in \mathcal{P}$ there exists $P^B \in \mathcal{P} \cap \Delta\{0,1\}^2$ such that $\mu\left(P^B\right) = \mu(P)$.

Property (a) reflects that the two actions are ex-ante identical up to their labelling. Even if the decision maker should perceive the two actions as different, this property ensures that the choices do not depend on this a priori perception. There is the following underlying principle: if the actions are really different then the data should

[4] $e^{-1} \approx 0.368$.

reflect this, not the structural assumptions. Property (b) holds automatically if $\mathcal{P}$ is described only in terms of the underlying means.

We say that a mean preserving transformation t is $\mathcal{P}$ *invariant* if for every $P \in \mathcal{P}$ there exists $P^B \in \mathcal{P}$ such that $P^B(t(y)) = P(y)$ for all $y \in [0,1]^2$. Following property (b) imposed above on $\mathcal{P}$, such a $\mathcal{P}$ invariant transformation always exists. Note that if $\mathcal{P}$ consists of all independent distributions, so $\mathcal{P} = (\Delta\{0,1\})^2$, then only the independent transformation t^I is $\mathcal{P}$ invariant. On the other hand, if $\mathcal{P}$ consists of all distributions, so $\mathcal{P} = \Delta\{0,1\}^2$, then any mean preserving transformation is $\mathcal{P}$ invariant. In the following we restrict attention to representatives of binomial fictitious play that are based on some $\mathcal{P}$ invariant transformation.

In applications there is typically some set $\mathcal{Y} \subseteq \mathbb{R}$ that contains all payoffs that can be generated by either action. In this paper we assume that $\mathcal{Y}$ is bounded. By applying an affine transformation, we can assume without loss of generality that $\inf \mathcal{Y} = 0$ and $\sup \mathcal{Y} = 1$. Then $\mathcal{P} \subseteq \Delta\mathcal{Y}^2$. If there are no further restrictions then $\mathcal{P} = \Delta\mathcal{Y}^2$. If one additionally knows that actions are independent then $\mathcal{P} = \Delta\mathcal{Y} \times \Delta\mathcal{Y}$. One may know additional information about how the payoffs generated by the two actions depend on each other. For instance, if one only knows that one of the two actions can yield a strictly positive payoff then $\mathcal{P} = \{P \in \Delta\mathcal{Y}^2 : P_1 P_2 = 0\}$ which of course satisfies conditions (a) and (b). Or one may be interested in correctly forecasting which of two states is more likely to occur. If at most one of these two states can occur then $\mathcal{P} = \Delta\{(0,0),(0,1),(1,0)\}$ which implies that $\mu_1 + \mu_2 \leq 1$.

Let $g(i,P)$ measure the *loss* of choosing action i when facing distribution P. We assume that g satisfies the following two conditions. (i) Loss only depends on the distribution via a symmetric continuous function of the means, so there exists a continuous function $g_0 : \{1,2\} \times [0,1]^2 \to \mathbb{R}$ such that $g(i,P) = g_0(i, \mu(P))$ and $g_0(1,\mu_1,\mu_2) = g_0(2,\mu_2,\mu_1)$ for $i = 1,2$. (ii) Choice of an action with higher expected payoff yields lower loss, so $g(i,P) < g(j,P)$ if and only if $\mu_i > \mu_j$.

Given (i) we can assume without loss of generality that g is non negative with $\inf_P g(1,P) = 0$. We then say that a loss function g *reflects learning when learning matters* if each action yields vanishing loss whenever the two means are arbitrarily close, formally if $\lim_{k\to\infty} P^k = P^\infty$ and $\mu_1(P^\infty) = \mu_2(P^\infty)$ implies
$\lim_{k\to\infty} \max\left\{g\left(1,P^k\right), g\left(2,P^k\right)\right\} = 0$.

We also add a nontriviality assumption on the pair $\mathcal{P}$ and g by requiring that there is some $P \in \mathcal{P}$ such that $g(1,P) \neq g(2,P)$.

Risk of choosing strategy σ when facing distribution P is defined as the expected loss, so $g(\sigma, P) = \sum_{i=1}^{2} E_P \sigma_i (y^{1,n})\, g(i, P)$. Risk when facing prior $Q \in \Delta\mathcal{P}$ is measured in terms of expected risk, so $g(\sigma, Q) = \int g(\sigma, P)\, dQ(P)$.

We will adapt a worst case approach which goes back to Wald (1950) and search for a strategy σ^* that attains *minimax risk* in the sense that $\sigma^* \in \arg\min_{\sigma\in\Sigma} \sup_{P\in\mathcal{P}} g(\sigma, P)$. $\min_{\sigma\in\Sigma} \sup_{P\in\mathcal{P}} g(\sigma, P)$ will be called the *value of minimax risk*. Q^* is called a *least favorable prior* if $\inf_{\sigma\in\Sigma} g(\sigma, Q^*) = \max_{Q\in\Delta\mathcal{P}} \inf_{\sigma\in\Sigma} g(\sigma, Q)$. The interpretation is that a rational decision maker endowed with a least favorable prior is worse off in terms of risk than with any other prior.

Loss can be measured in terms of *regret* defined as $g(i, P) = \max\{\mu_1, \mu_2\} - \mu_i$. The criterion of minimax regret is due to Savage (1951). If a strategy attains minimax risk when loss is a translation of negative payoffs, so $g(i, P) = 1 - \mu_i(P)$, then we say that it attains *maximin utility*. The maximin utility criterion was first defined by Wald (1950). Both minimax regret and maximin utility were first axiomatized by Milnor (1954) and recently also by Stoye (2006). The *relative minimax* criterion axiomatized by Terlizzese (2006) is similar to the minimax regret criterion except that loss of opportunity is now measured in relative terms. $\mathcal{P}$ must satisfy $\mu_1(P) \neq \mu_2(P)$ for all $P \in \mathcal{P}$ and loss is defined by

$$g(i, P) = 1 - \frac{\mu_i - \min\{\mu_1, \mu_2\}}{\max\{\mu_1, \mu_2\} - \min\{\mu_1, \mu_2\}} = 1_{\{\mu_i < \mu_{3-i}\}}.$$

Note that the three criteria have in common that they have been axiomatized and are invariant to any positive affine transformations of the payoffs. Neither of these properties holds for the *competitive ratio* criterion (Borodin and El-Yaniv, 1998) that is very popular in the computer science literature. It only can be used when $\alpha > 0$ and is defined by

$$g(i, P) = 1 - \frac{\alpha + (\omega - \alpha)\,\mu_i}{\alpha + (\omega - \alpha)\max\{\mu_1, \mu_2\}}.$$

Apart from the lack of a normalization competitive ratio is identical to the relative minimax criterion. One may also choose as in the literature on selection procedures (Sobel and Huyett, 1957), without reference to any axioms, to be only interested in selecting the better action whenever the two means differ by at least d where $d \in (0, 1)$ is given. Loss g is then defined by $g(i, P) = 1_{\{\mu_i \leq \mu_{3-i} - d\}}$. A rule then attains minimax risk if it maximizes the minimum probability of correct selection. Notice that this last loss function can also be used to find a (randomized) test for the null hypothesis that

$\mu_1 \geq \mu_2 + d$ against the alternative hypothesis that $\mu_1 \leq \mu_2 - d$. Choosing action 1 or action 2 is identified with not rejecting or rejecting the null hypothesis respectively. A rule that attains minimax risk under this loss function yields a test (provided a saddle point defined below exists) that has equal type I and II errors such that there is no alternative test with the same type I error (size) that has a strictly lower type II error (see appendix).

Note that all examples presented above apart from maximin utility and relative minimax reflect learning when learning matters.

Following von Neumann Morgenstern (see also Savage, 1954) we derive minimax risk by finding a saddle point. (σ^*, Q^*) is a *saddle point* if

$$\max_{Q \in \Delta\mathcal{P}} g(\sigma^*, Q) = g(\sigma^*, Q^*) = \min_{\sigma \in \Sigma} g(\sigma, Q^*).$$

In other words, (σ^*, Q^*) is a Nash equilibrium of a imaginary zero sum game between the decision maker and nature in which the decision maker aims to minimize risk while nature aims to maximize the risk of the decision maker.

If (σ^*, Q^*) is a saddle point then σ^* attains minimax risk and Q^* is a least favorable prior. Whenever a saddle point exists then any pair consisting of a minimax risk strategy and a least favorable prior constitutes a saddle point. In particular, if a saddle point exists then minimax risk is consistent with rational decision making as any minimax risk strategy is a best response to a least favorable prior. Formally, if (σ^*, Q^*) is a saddle point then $\sigma^* \in \arg\min_{\sigma \in \Sigma} g(\sigma, Q^*)$ and hence $\sigma^* \in \arg\max_{\sigma \in \Sigma} \int_P \sum_{i=1}^{2} E_P \sigma_i(y^{1,n}) \mu_i(P) \, dQ(P)$.

We now turn to our results. We compare risk of binomial fictitious play to that of other symmetric rules and establish the minimax risk properties of binomial fictitious play when comparing to any other rule.[5] We also provide some necessary conditions for performing as well as binomial fictitious play.

To simplify presentation we say that a rule σ is *observationally equivalent to fictitious play* σ^f *when facing* $P^B \in \Delta\{0,1\}^2$ if for all $y^{1,n}$ that are realized with positive probability when facing P^B, $\sigma_i(y^{1,n}) > 0$ implies $\sigma_i^f(y^{1,n}) > 0$ which is equivalent to requiring $\sigma(y^{1,n}) = \sigma^f(y^{1,n})$ when $\sum_{k=1}^{n} y_1^k \neq \sum_{k=1}^{n} y_2^k$.

[5] Note that there is always a rule that outperforms binomial fictitious play when facing a given distribution P, namely one of the two rules that chooses the same action regardless of the sample.

Proposition 2 *(i) For any symmetric rule* σ^s, $g\left(\sigma^b, P^B\right) \leq g\left(\sigma^s, P^B\right)$ *for all* $P^B \in \Delta\{0,1\}^2$ *with strict inequality holding if* $\mu_1\left(P^B\right) \neq \mu_2\left(P^B\right)$ *and* σ^s *is* ***not*** *observationally equivalent to fictitious play when facing* P^B.

(ii) Binomial fictitious play σ^b *attains minimax risk for any* $\mathcal{P}$ *and there exists a least favorable prior that has support in* $\Delta\{0,1\}^2$.

(iii) If loss reflects learning when learning matters and σ^* *attains minimax risk then* σ^* *is observationally equivalent to fictitious play when facing any* $P^B \in \Delta\{0,1\}^2$ *contained in the support of a least favorable prior.*

Proof. Part (ii). We first show that σ^b attains minimax risk and do this by finding a saddle point.

$g\left(\sigma^b, P\right)$ attains its maximum on $\mathcal{P} \cap \Delta\{0,1\}^2$ as g is continuous in μ. Following arguments in Schlag (2003) and Schlag (2006a), $\max_{P\in\mathcal{P}} g\left(\sigma^b, P\right) = \max_{P\in\mathcal{P}\cap\Delta\{0,1\}^2} g\left(\sigma^b, P\right)$. Notice that it is here that we use the existence of a $\mathcal{P}$ invariant transformation. Choose $P^* \in \arg\max_{P\in\mathcal{P}\cap\Delta\{0,1\}^2} g\left(\sigma^b, P\right)$ and let Q^* be the prior that puts equal weight on P^* and P^{*s}.

We verify that $\left(\sigma^b, Q^*\right)$ is a saddle point. Since $P^*, P^{*s} \in \arg\max_{P\in\mathcal{P}} g\left(\sigma^b, P\right)$ we obtain that $Q^* \in \arg\max_{Q\in\Delta\mathcal{P}} g\left(\sigma^b, Q\right)$. It is easily verified that σ^b is a best response against Q^* in the sense that $\sigma^b \in \arg\max_{\sigma\in\Sigma}\left(E_{P^*}\pi\left(\sigma, y^{1,n}\right) + E_{P^{*s}}\pi\left(\sigma, y^{1,n}\right)\right)$. Hence $\sigma^b \in \arg\min_{\sigma\in\Sigma} g\left(\sigma, Q^*\right)$.

Since $\left(\sigma^b, Q^*\right)$ is a saddle point, σ^b attains minimax risk and Q^* is a least favorable prior which by construction has support in $\Delta\{0,1\}^2$.

Part (i). Let $\mathcal{P} = \left\{P^B, P^{Bs}\right\}$. Then Q^* puts equal weight on P^B and P^{Bs}. The inequality in part (i) then follows from the fact that $g\left(\sigma^s, Q^*\right) = g\left(\sigma^s, P^B\right)$ as σ^s is symmetric. If $g\left(\sigma^s, P^B\right) = g\left(\sigma^b, P^B\right)$ then σ^s attains minimax risk and hence σ^s is a best response to Q^*. So if $\mu_1\left(P^B\right) \neq \mu_2\left(P^B\right)$ then σ^s has to be observationally equivalent to fictitious play when facing P^B.

Part (iii). For general $\mathcal{P}$, if loss reflects learning when learning matters then $\mu_1\left(P^*\right) \neq \mu_2\left(P^*\right)$ if P^* is in the support of a least favorable prior. Following our proof of part (i), we obtain that σ^* is observationally equivalent to fictitious play when facing P^*. ■

We expand minimally on Proposition 2(i) and evaluate performance when facing general distributions.

Corollary 2 *For any symmetric rule σ^s and any $\bar{\mu} \in [0,1]^2$, $\max_{P:\mu(P)=\bar{\mu}} g\left(\sigma^b, P\right) \leq \max_{P:\mu(P)=\bar{\mu}} g\left(\sigma^s, P\right)$ with strict inequality holding if $\bar{\mu}_1 \neq \bar{\mu}_2$ and σ^s is **not** observationally equivalent to fictitious play when facing P^B.*

Assume that loss reflects learning when learning matters. Following Proposition 2(i) and Corollary 2, if a rule wants to perform as well as binomial fictitious play then it must be behaviorally equivalent to fictitious play when facing any binary valued distribution. Apart from fictitious play, the other rules suggested in the literature for learning under foregone payoffs in various environments (e.g. stochastic fictitious play and regret matching) perform worse than binomial fictitious play in our setting. Only binomial fictitious play and fictitious play remain to be compared.

We present some risk properties of fictitious play. Part (i) represents the first foundation of fictitious play that is not based on priors.

Corollary 3 *(i) Fictitious play attains minimax risk if $\mathcal{P} \subseteq \Delta\{0,1\}^2$.*

(ii) Fix $n \in \mathbb{N}$. There exists $\bar{\mu} \in [0,1]^2$ such that $\max_{P:\mu(P)=\bar{\mu}} g\left(\sigma^b, P\right) < \max_{P:\mu(P)=\bar{\mu}} g\left(\sigma^f, P\right)$ holds for all g. There is some $\mathcal{P}$ such that fictitious play does not attain minimax risk for any loss function g.

(iii) If $\mathcal{P} = \Delta[0,1]^2$ and loss g is measured by $g(i,P) = 1_{\{\mu_i < \mu_{3-i}\}}$ then fictitious play does not attain minimax risk for any $n \in \mathbb{N}$.

Proof. Part (i) follows directly from Proposition 2.

Part (ii). Consider a distribution $\bar{P}$ that satisfies Corollary 1(ii). Facing $\bar{P}$ the risk of fictitious play is strictly higher than that of the strategy to choose each action equally likely which by Corollary 1(i) is strictly higher than the maximal risk attained by binomial fictitious play.

Part (iii). The distribution $\bar{P}$ used in the proof of part (ii) shows that fictitious play does not attain minimax risk. ■

In the following we briefly investigate minimax risk in more detail when $\mathcal{Y} = [0,1]$ under each of the three criteria that have been axiomatized (maximin utility, minimax regret, relative minimax) and for selecting which action is best.

4.1 Maximin Utility

Consider the maximin utility criterion when $\mathcal{P} = \Delta[0,1]^2$ or when $\mathcal{P} = (\Delta[0,1])^2$. Proposition 2(ii) does not apply as loss underlying maximin utility does not reflect

learning when learning matters.

It follows easily that any strategy attains maximin utility. This is because the distribution under which both actions always yield payoff 0 is a least favorable distribution for each strategy. In particular, rules that ignore all information can attain maximin utility. Analogous results were attained for the case of a single unknown action by Manski (2005) and for the partial information setting by Schlag (2006a).

4.2 Minimax Regret

Consider the minimax regret criterion and $\mathcal{P} = \Delta[0,1]^2$. Schlag (2006a) shows that the binomial average rule attains minimax regret in the setting of this paper even though it uses less information.

The *binomial average rule*, denoted here by σ^a, is the pendent of binomial fictitious play for the partial information setting. For even sample sizes each action is sampled equally often, payoffs in $(0,1)$ are transformed independently in $\{0,1\}$ as in this paper and then the empirically most successful action is chosen, mixing equally likely when there is a tie. Thus, behavior of the binomial average rule in the partial information setting based on $2n$ observations facing distribution P is identical to that of binomial fictitious play under the independent transformation in the full information setting based on n observations facing the distribution $\hat{P}$ that has the same marginals as P and where actions yield independent payoffs. To obtain the definition of the binomial average rule for odd sample sizes the following adjustment is made. If the transformed payoff in the last round is 0 then drop this observation and add an observation of payoff 1 to the alternative action. Then proceed as in the definition for even sample sizes (note that a tie is not possible when n is odd).

In the following we show that binomial fictitious play outperforms the binomial average rule for any loss function. Formally speaking, this shows that the binomial average rule is not *admissible* for the full information setting when $n \geq 2$.[6]

Proposition 3 *For any loss function, $r\left(\sigma^b, P\right) \leq r\left(\sigma^a, P\right)$ for all $P \in \Delta[0,1]^2$ with strict inequality if $\mu_1 \neq \mu_2$, $\mu_1 + \mu_2 \neq 1$ and $n \geq 2$. Moreover, the statement holds with strict inequality when $\sigma^b \in \left\{\sigma^{bI}, \sigma^{bC}\right\}$ if and only if $\mu_1 \neq \mu_2$, $P \notin \Delta\{(0,1),(1,0)\}$ and $n \geq 2$.*

[6] A rule is admissible if there is no other rule that always yields lower risk with strictly lower risk in some enviornments. In game theoretic terms, a rule is admissible if it is not weakly dominated.

Proof. Fix a representative of binomial fictitious play. Fix any distribution P. Then there exists $P^B \in \Delta\{0,1\}^2$ such that $\mu(P) = \mu\left(P^B\right)$ and $r\left(\sigma^b, P\right) = r\left(\sigma^b, P^B\right)$. Following Proposition 2(i), $r\left(\sigma^b, P^B\right) \leq r\left(\sigma^a, P^B\right)$. Since $r(\sigma^a, P) = r\left(\sigma^a, P^B\right)$ we obtain that $r\left(\sigma^b, P\right) \leq r(\sigma^a, P)$.

It follows from Proposition 2(i) that $r\left(\sigma^b, P^B\right) < r\left(\sigma^a, P^B\right)$ if $n \geq 2$, $\mu_1 \neq \mu_2$ and $P^B \notin \Delta\{(0,1),(1,0)\}$. Note that $P^B \notin \Delta\{(0,1),(1,0)\}$ holds if $\mu_1 + \mu_2 \neq 1$ as $\mu_1 + \mu_2 < 1$ implies $P^B(0,0) > 0$ and $\mu_1 + \mu_2 > 1$ implies $P^B(1,1) > 0$.

It is easily verified that binomial fictitious play based on either the independent or on the correlated transformation will yield a strictly lower risk than the binomial average rule if and only if $\mu_1 \neq \mu_2$, $P \notin \Delta\{(0,1),(1,0)\}$ and $n \geq 2$. ■

The reason why the binomial average rule can attain minimax regret under full information is that Schlag (2006a) shows when $\mathcal{P} = \Delta[0,1]^2$ that there is a least favorable prior in $\Delta\{(0,1),(1,0)\}$. If instead $\mathcal{P} \cap \{P : \mu_1 + \mu_2 = 1\} = \emptyset$ and $n \geq 2$ then it follows from the proposition above that the binomial average rule can no longer attain minimax risk.

Given $\mathcal{P} = \Delta[0,1]^2$ and following Schlag (2006a), the value of minimax regret is the same in the partial and in the full information setting. It equals $1/2$ for $n = 0$, $1/8$ for $n = 1$, and is approximately equal to $0.17/\sqrt{n + 0.8}$ when $n > 1$ and n is odd. For n even the value of minimax regret is equal to that of the preceding round, hence approximately equal to $0.17/\sqrt{n - 0.2}$.

We expand minimally on earlier results to find that absolute expediency and minimax regret are not compatible.

Corollary 4 *If $\mathcal{P} = \Delta[0,1]^2$ then an absolutely expedient rule that attains minimax regret for $n = 0$ does not attain minimax regret for $n \geq 1$.*

Proof. As loss reflects learning when learning matters then there is some $P^* \in \Delta\{0,1\}^2$ such that $\mu_1(P^*) > \mu_2(P^*)$ and P^* is contained in the support of a least favorable prior. Assume that σ^* attains minimax regret and is absolutely expedient. Then σ^* is observationally equivalent to fictitious play when facing P^*. Since $(1,0)$ can occur in round 1 when facing P^*, as argued at the end of Section 3.1, an absolutely expedient rule chooses action 1 in all rounds $n \geq 2$. In order for σ^* to be observationally equivalent to fictitious play it follows that $P^*(1,0) = 1$. However it is

easily verified (see also Schlag, 2006a) that P^* is contained in the support of a least favorable prior only if $n = 0$. ■

We briefly investigate the performance of fictitious play in small samples. Considering regret when facing the distribution $\hat{P} \in \Delta\{(0,x),(1,x)\}$ for $x < 1/n$ and $x < \mu_1 = z$ we obtain

$$\max_P r\left(\sigma^f, P\right) \geq \sup_{0<x<\min\{z,1/n\}} (z-x)(1-z)^n = \frac{n^n}{(n+1)^{n+1}}.$$

It is easily verified that this lower bound on the maximal regret of fictitious play is strictly above the value of minimax regret when $n \leq 4$. Thus fictitious play does not attain minimax regret when $1 \leq n \leq 4$. In particular, when $n = 1$ then we find that the maximal regret of fictitious play is at least $1/2$ where $1/2$ is the value of minimax regret without any observation. Whether or not fictitious play attains minimax regret for $n > 5$ and $\mathcal{P} = \Delta[0,1]^2$ remains an open question.

4.3 Relative Minimax

Consider relative minimax and $\mathcal{P} = \Delta[0,1]^2 \setminus \{P : \mu_1 = \mu_2\}$. Notice that this specification does not fit our setting as $\mathcal{P}$ is not closed. Notice also that relative minimax cannot be extended to a continuous loss function on $\Delta[0,1]^2$. Thus we cannot build on our above results and instead perform some explicit calculations. We will show that, similar to the maximin utility criterion, rules that ignore all information can attain relative minimax for any sample size n.

Let $\bar{\sigma}$ be the strategy that specifies to choose each action equally likely regardless of which payoffs are observed in the sample. $\bar{\sigma}$ is the unique symmetric rule that does not depend on the sample. It is easily verified that $\sup_{P:\mu_1\neq\mu_2} r(\bar{\sigma}, P) = -1/2$. Now consider a prior $\hat{Q}_\varepsilon$ that puts equal weight on $\hat{P}$ and $\hat{P}^s$ where $\mu(P) = (0,\varepsilon)$ with $\varepsilon > 0$ and ε small. Then with high probability both actions always yield payoff 0 in the sample and hence $\lim_{\varepsilon\to 0} r\left(\sigma, \hat{Q}_\varepsilon\right) = -1/2$ for any strategy σ. Thus the value of relative minimax is equal to $-1/2$, in particular $\bar{\sigma}$ attains relative minimax.

4.4 Finding the Best Action

Consider $\mathcal{P} = \Delta[0,1]^2$ and the objective of finding the best action conditional on the difference between the two means being at least equal to some given distance $d \in (0,1)$. So loss g is given by $g(i,P) = 1_{\{\mu_i \leq \mu_{3-i} - d\}}$. Schlag (2006b) considers this

loss function in the partial information setting and shows as in the case of minimax regret (Schlag, 2006a) that the binomial average rule attains minimax risk and that the value of minimax risk in the partial information setting is equal to that in the full information setting. We cite the value of minimax risk provided in Schlag (2006a, 2006b):

$$\min_{\sigma}\max_{P} r_{2m}(\sigma,P)=\min_{\sigma}\max_{P} r_{2m-1}(\sigma,P)=\left(\frac{1-d}{2}\right)^{2m-1}\sum_{k=0}^{m-1}\binom{2m-1}{k}\left(\frac{1+d}{1-d}\right)^{k}.$$

For example, consider $n = 29$ observations and $d = 0.296$. Then the value of minimax risk is equal to 0.05. So under binomial fictitious play the two means have to be at least 0.296 apart in order for there to be a rule (e.g. binomial fictitious play) that is able to select the best action with probability at least 0.95. Moreover the performance can only be improved in terms of minimizing maximal risk if at least **two** more samples are observed. In terms of testing consider the hypothesis $\mu_1 \geq \mu_2 - d$ verses $\mu_1 \leq \mu_2 - d$ for $d = 0.296$. Then there is no test that has type I and type II error below 5% when the sample size n is smaller than 28 but there is when $n = 29$. For more information on this test see the appendix. Other values are easily calculated, e.g. one cannot guarantee to find the best action more than 70.7% of the time if $d = 0.1$ and $n = 29$, both type I and type II error can be pushed below 5% for $d = 0.1$ if and only if $n \geq 269$.

5 Two-Armed Bandit

In the following we consider the two-armed bandit setting with discounting when foregone payoffs are observable. The decision maker has to repeatedly choose either action 1 or action 2, always facing the same but unknown distribution P, each choice of action i yields an independent payoff drawn from P_i. The strategy σ of a decision maker is now given by a function $\sigma : \cup_{n=0}^{\infty} [0,1]^{2n} \to \Delta\{1,2\}$. Loss g is now a function of the sequence of choices so $g : \{1,2\}^{\infty} \times \mathcal{P} \to \mathbb{R}$. To keep the presentation simple we assume that g is additive so $g((i_n)_n, P) = \sum g^n(i_n, P)$ for some $g^n : \{1,2\} \times \mathcal{P} \to \mathbb{R}$ where g^n satisfies conditions (i) and (ii) from Section 4 for each n and where we require that g is bounded.

As the choice of the decision maker does not influence the observed outcomes the decision maker can solve each round separately. This proves the following.

Corollary 5 *Binomial fictitious play σ^b attains minimax risk in the two-armed bandit setting.*

5.1 Minimax Regret

In the following we measure loss in terms of regret. Consider first the case where payoffs are discounted so there exists $\delta \in (0,1)$ such that

$$g\left((i_n)_n, P\right) = \max\{\mu_1, \mu_2\} - (1-\delta)\sum_{n=1}^{\infty} \delta^{n-1}\mu_{i_n}.$$

Proposition 4 *Assume $\mathcal{P} = \Delta[0,1]^2$.*

(i) The value of minimax regret is equal to $\frac{1}{2}(1-\delta)$ if and only if $\delta \leq \frac{1}{2}\left(\sqrt{5}-1\right)$, for $\frac{1}{2}\left(\sqrt{5}-1\right) < \delta < 1$ it is equal to

$$(1-\delta)\max_{x\in[0,1]}\left[\frac{x}{2} + (1+\delta)\,x\sum_{m=1}^{\infty}\delta^{2m-1}\sum_{k=0}^{m-1}\binom{2m-1}{k}\left(\frac{1}{2}(1+x)\right)^k\left(\frac{1}{2}(1-x)\right)^{2m-1-k}\right] \tag{6}$$

which is strictly larger than $\frac{1}{2}(1-\delta)$.

(ii) Fictitious play fails to attain minimax regret if $\frac{1}{2} < \delta \leq \frac{1}{2}\left(\sqrt{5}-1\right)$.

Proof. Part (i). Fix $x \in (0,1]$ and let $\mathcal{P}_x = \left\{P \in \Delta[0,1]^2 : |\mu_1 - \mu_2| = x\right\}$. The arguments used in Schlag (2006a) show that the binomial average rule attains minimax regret among all distributions in $\mathcal{P}_x$ when foregone payoffs are observable with least favorable prior given by $P((1,0)) = \frac{1}{2}(1+x) = 1 - P((0,1))$. The analysis in Schlag (2006a) then shows that the value of maximal regret conditional on $2m-1$ or $2m$ observations is equal to

$$\sum_{k=0}^{m-1}\binom{2m-1}{k}\left(\frac{1}{2}(1+x)\right)^k\left(\frac{1}{2}(1-x)\right)^{2m-1-k}$$

for $m \in \mathbb{N}$. Using the fact that regret in round 1 is equal to $\frac{1}{2}x$ it follows that the value of maximal regret in the two armed bandit is equal to

$$(1-\delta)\left[\frac{x}{2} + \delta\,(1+\delta)\,x\sum_{m=1}^{\infty}\delta^{2m-1}\sum_{k=0}^{m-1}\binom{2m-1}{k}\left(\frac{1}{2}(1+x)\right)^k\left(\frac{1}{2}(1-x)\right)^{2m-1-k}\right] \tag{7}$$

and consequently (6) specifies the value of maximal regret when any distribution in $\Delta[0,1]^2$ is allowed.

Derive an upper bound on the maximal regret under binomial fictitious by adjusting behavior by assuming that the decision maker chooses the action chosen in round 7 for ever. Formally this means that we discard the terms in (7) with $m > 3$ and add

$$x\delta^7 \sum_{k=0}^{2} \binom{5}{k} \left(\frac{1}{2}(1+x)\right)^k \left(\frac{1}{2}(1-x)\right)^{5-k}.$$

Taking the derivative with respect to x it is easily shown that regret of this rule is maximized for $x = 1$ taking value $\frac{1}{2}(1-\delta)$ if $\delta \leq \frac{1}{2}\left(\sqrt{5-1}\right)$. This means that $\frac{1}{2}(1-\delta)$ is also an upper bound on maximal regret of σ^b for $\delta \leq \frac{1}{2}\left(\sqrt{5-1}\right)$. This value is actually obtained for $x = 1$ so that we have proven the if statement. Concerning the only if statement we derive the expression in (7) to x and enter $x = 1$ to obtain $\frac{1}{2}(1-\delta)\left(1-\delta-\delta^2\right)$ which means that $x = 1$ is not the maximizer if $\delta > \frac{1}{2}\left(\sqrt{5}-1\right)$.

Part (ii). Consider P such that $P((1,0)) = 1 - P((0,x)) = z$ for some given $x < 1$ and z close to 0. Then

$$\pi \leq (1-\delta)\frac{1}{2} + (1-\delta)\delta\left(z^2 + (1-z)^2\right) + (1-\delta)\delta^2\left(\left(1-(1-z)^2\right)z + (1-z)^3\right) + \delta^3(1-z)$$

and hence

$$r\left(\sigma^f, P\right) \geq 1 - z - \left(\begin{array}{c} (1-\delta)\frac{1}{2} + (1-\delta)\delta\left(z^2 + (1-z)^2\right) \\ + (1-\delta)\delta^2\left(\left(1-(1-z)^2\right)z + (1-z)^3\right) + \delta^3(1-z) \end{array} \right). \quad (8)$$

Notice that the right hand side equals $\frac{1}{2}(1-\delta)$ if $z = 0$. Taking the derivative of the right hand side with respect to z and evaluating this at $z = 0$ we obtain $-(1-\delta)(1-2\delta)(1+\delta)$ which is strictly positive if $\delta > 1/2$. Given part (i), these two findings prove part (ii). ■

We use (6) to plot the value of minimax regret for $\delta > \frac{1}{2}\left(\sqrt{5}-1\right) \approx 0.618$ in Figure 1.[7]

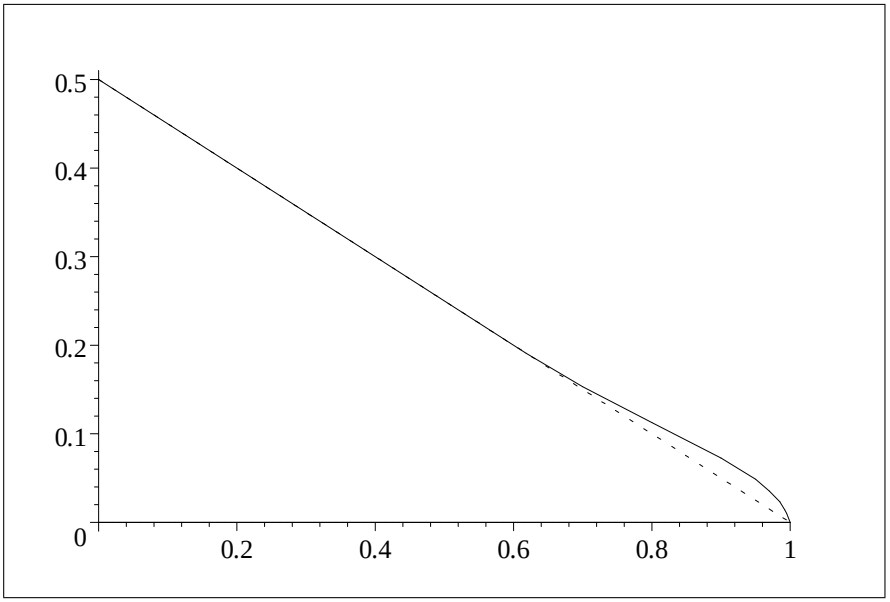

Figure 1: Value of minimax regret as a function of δ with $\frac{1}{2}(1-\delta)$ added as dotted line.

Remark 2 *Assume $\mathcal{P} = \Delta[0,1]^2$. We verify numerically that fictitious play does not attain minimax regret for $\frac{1}{2}\left(\sqrt{5}-1\right) < \delta \leq 0.79935$. For this we compare the lower bound for regret of σ^f in (8) to the value of minimax regret found in Proposition 4.*

Consider now the following alternative specification of regret that is common in the machine learning literature. Fix N and let loss be defined by

$$g\left(\left(i_n\right)_n, P\right) = \max\left\{\mu_1, \mu_2\right\} - \frac{1}{N}\sum_{n=1}^{N}\mu_{i_n}.$$

Proposition 5 *Assume $\mathcal{P} = \Delta[0,1]^2$. The value of minimax regret is equal to*

$$\frac{1}{N}\max_{x\in[0,1]}\left(\frac{x}{2} + x\sum_{m=1}^{N}\sum_{k=0}^{\lfloor (N+1)/2\rfloor -1}\binom{2\lfloor (N+1)/2\rfloor -1}{k}\left(\frac{1}{2}(1+x)\right)^k\left(\frac{1}{2}(1-x)\right)^{2\lfloor (N+1)/2\rfloor -1-k}\right).$$

[7] The first sum is evaluated for $m \leq 35$.

Proof. The same arguments as used in the proof of Proposition 4 apply. ■

We plot the value of minimax regret for $N \leq 22$ and note that $r \approx 5\%$ if $N = 30$ and for $3 \leq N \leq 80$ that $r \approx \frac{0.265}{\sqrt{N-1}}$ up to the third decimal point different from 0.

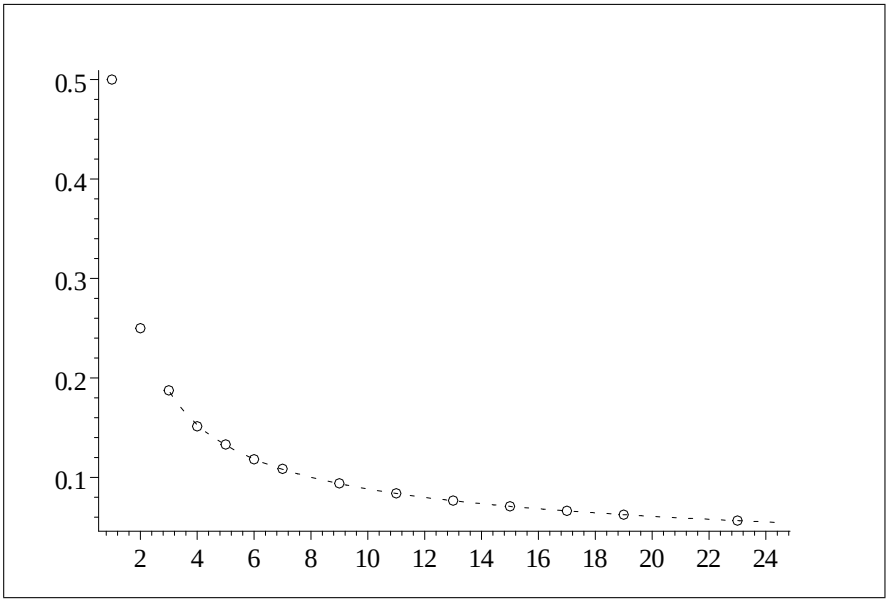

Figure 2: Value of average minimax regret as a function of N with approximation $0.265/\sqrt{N-1}$ added as dotted line.

6 Choice among Three and More Actions

In the following we briefly consider choice among $I \geq 3$ ex-ante identical actions. The definition of binomial fictitious play is easily extended, maintaining the assumption that the decision maker chooses equally likely among the empirically most successful actions after first binomially transforming the data.

Proposition 6 *When actions are known to be independent then binomial fictitious play σ^{bI} based on the independent transformation attains minimax risk conditional on n observations for any n and hence also attains minimax risk in the multi-armed bandit setting for any δ.*

Proof. The proof is analogous to that of Proposition 2. The only adjustment is that Q^* now puts equal weight on the $I!$ distributions that emerge from permuting

the labels of P^*. Independence of actions allows us to conclude immediately that choosing the most successful action is a best response to Q^*. ■

Next we show by example that fictitious play is no longer always a best response to symmetric priors over binary valued distributions.

Example 1 *Let $e^i, v^i \in \{0,1\}^I$ be such that e^i is the i-th unit vector and $v^i_j = 0$ if and only if $i = j$. Let $\bar{P}$ be such that $\bar{P}(e_1) = 1 - \bar{P}(v_1) \in (1/2, 1)$ so $\mu_1 > \mu_i$ for $i \geq 2$ and consider the symmetric prior $\bar{Q}$ that puts weight $1/I!$ on each of the distributions that emerges from $\bar{P}$ by permuting the labels of the actions. For any n the best response to $\bar{Q}$ is to choose action i if and only if either e_i or v_i was observed in the first round. In particular note that fictitious play is not a best response to $\bar{Q}$ as action i is never chosen when only v^i has been observed.*

Note that the above example does not preclude fictitious play from attaining minimax risk for binary valued distributions. It only shows that the proof technique used above for the case of two actions does not extend to the setting with three or more actions.

7 Conclusion

We illustrate how the findings of this paper can be used in different disciplines. In statistical decision theory one can now derive policy recommendations based on paired data for small sample sizes as maximal regret is below 5% when there are at least 11 observations. One can investigate the effectivity of a treatment by comparing well being before and after the treatment of n subjects. From a more statistical perspective, we have shown that binomial fictitious play is also the selection procedure that yields the highest minimal probability of selecting whether the mean well being was higher before or after the treatment. One can use binomial fictitious play to forecast which of two states will occur next. In game theory one can investigate learning in games when players believe that their opponents are using a stationary strategy. In economics one can now relax the rationality of agents away from the standard subjective expected utility model without assigning ad-hoc behavior. For instance, in industrial organization one could now investigate a finite number of rational firms independently setting stationary random prices who are repeatedly competing for consumers who have no priors.

The only previous justification in the literature for using fictitious play is the fact that it is very popular, very simple to implement and has nice limit properties. We add a formal justification for choice between two actions under binary valued payoff distributions based on finite samples. Fictitious play is ex-ante improving and attains minimax risk for any loss function that only depends on the underlying means. However neither of these two properties carry over once payoffs are only known to belong to a bounded interval. Instead, its variant introduced in this paper called binomial fictitious play has these properties.

Choice between two actions in a stationary decision problem with full information is an extremely simple setting. This allows us to select binomial fictitious play for general loss functions. Our analysis provides a possible starting point for many future investigations. One may want to compare the different representatives of binomial fictitious play, as a function of the underlying mean preserving transformation. For applications to data the correlated transformation seems most promising as it maximizes the covariance between the two transformed outcomes. Higher correlation will tend to reduce the variance in the final choice as it did for the correlated binomial average rule in the partial information setting (Eozenou et al., 2006). More general settings need to be investigated, such as choice when different actions have different outcomes or when there are three or more dependent actions.

We point out the advantage of choosing according to minimax risk. Most important there are axiomatic foundations that yield the particular representatives maximin utility, minimax regret and relative minimax (Milnor, 1954, Stoye, 2006, Terlizzese, 2006). Both the maximin utility and the relative minimax criteria are too weak to generate nice limit properties without restricting the set of environments. The symmetric rule that ignores the sample attains maximin utility and relative minimax. Thus we put additional emphasis in this paper on minimax regret. The connection to statistics also makes the objective of selecting the best action interesting.

The concepts of ex-ante improving and absolute expediency are appealing but do not have an axiomatic foundation. Characterizations are difficult to obtain as these concepts are defined by an infinite number of constraints (cf. Schlag, 1998, 1999 and Börgers et al., 2004). It is interesting to note that the rule we found that attains minimax risk for all loss functions is also ex-ante improving. On the other hand we find that minimax regret and absolute expediency are incompatible.

Choice without priors means that the same selected rule is applicable to many different problems. Thus, the decision-maker does not have to reoptimize whenever facing a new decision problem. It is the lack of priors that seems to cause strategies selected to also be simple (see also Schlag, 1998, 2003, 2006a).

We finally point out some connections to the partial information setting.

Schlag (2006a, 2006b) shows when loss is either equal to regret or to the probability of choosing the worse action that the value of minimax risk under partial information, when a given number n of observations can be gathered by the decision maker, is the same as it is under full information. We select a rule that outperforms the rule selected for the partial information setting as it achieves strictly lower risk except in very particular environments.

In the partial information setting, the results in Schlag (2006a, 2006b) are useless for sequential decision making. The binomial average rule is designed to perform best in round $n+1$ with choices beforehand not influencing loss. In a sequential decision making problem with discounting the choice in each round not only provides information but directly enters the loss function. Exploration and exploitation have to be traded off. Schlag (2003) considers the sequential decision making problem but is only able to derive rules that attain minimax regret for small and moderate discount factors if the decision maker is able to commit to these rules. In the full information setting, there is no issue of time consistency or commitment. As nature chooses the distribution before round 1 it is only justified to evaluate rules according to ex-ante payoffs. However, whether payoffs starting round 1 or starting round k are considered, the same rule is selected. This is because with full information there is no choice of how to gather information.

The role that is played by fictitious play in our paper is played by the empirical success rule in Schlag (2006a). In fact, these two rules are identical if n observations of action 1 and n of action 2 are identified with n observations of pairs consisting of one observation of each action.

A Testing Symmetric Hypotheses

In the following we briefly show how the findings of this paper can be used to design tests.

Fix some $d \in (0,1)$. Assume that we wish to test the null hypothesis that $\mu_1 \geq \mu_2 + d$ against the alternative hypothesis that $\mu_1 \leq \mu_2 - d$ where μ_1 and μ_2 are unknown apart from that $\mu_1, \mu_2 \in [0,1]^2$.[8] Note that hypotheses are *symmetric* in the sense that the null and the alternative hypothesis are interchanged when the labels of the two variables are interchanged.

In the following we will show how to find a test with the following two properties. The test has equal type I and II errors and there is no alternative test with the same type I error (size) that has a strictly lower type II error. Notice that the second property is stronger than that of simply being most powerful. The test will not be unbiased and it will be randomized in the sense that it will produce some probability with which the null hypothesis can be rejected where this probability will typically be in $(0,1)$. The binomial fictitious play will be such a test by identifying choice of action 1 and of action 2 with not rejecting and rejecting the null hypothesis respectively.

Consider the loss function $g(i,P) = 1_{\{\mu_i \leq \mu_{3-i} - d\}}$ and find a rule σ^* that attains minimax risk, e.g. σ^* could be binomial fictitious play. We prove that σ^* has the two properties mentioned above. Assume that there is a test $\hat{\sigma}$ with a lower or equal type I error, so $\max_{P:\mu_1 \geq \mu_2 + d} g(\hat{\sigma}, P) \leq \max_{P:\mu_1 \geq \mu_2 + d} g(\sigma^*, P)$, that has a strictly lower type II error, so $\max_{P:\mu_1 \leq \mu_2 - d} g(\hat{\sigma}, P) < \max_{P:\mu_1 \leq \mu_2 - d} g(\sigma^*, P)$. Since there is a saddle point (see proof of Proposition 2), a least favorable prior Q^* will have a distributions with $\mu_1 \geq \mu_2 + d$ as well as with $\mu_1 \leq \mu_2 - d$ in its support. Consequently, $\max_{P:\mu_1 \geq \mu_2 + d} g(\sigma^*, P) = g(\sigma^*, Q^*) = \max_{P:\mu_1 \leq \mu_2 - d} g(\hat{\sigma}, P)$ which means that the type I and type II errors are equal for σ^* and that $g(\hat{\sigma}, Q^*) < g(\sigma^*, Q^*)$ where the latter contradicts the fact that σ^* attains minimax risk.

[8] The range where $\mu_2 - d < \mu_1 < \mu_2 + d$ is called the *indifference zone* as the decision maker does not care whether or not the null hypothesis is rejected for distributions with means that fall within this range.

References

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

[2] Auer, P., N. Cesa-Bianchi and P. Fischer (2002), "Finite-Time Analysis of the Multiarmed Bandit Problem," *Mach. Learning* **27**, 235–256.

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

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

[5] Borodin, A. and R. El-Yaniv (1998), *Online Computation and Competitive Analysis*, Cambridge: Cambridge University Press.

[6] Brown, G.W. (1951), "Iterative Solution of Games by Fictitious Play," in: T.C. Koopmans (Ed.), *Activity Analysis of Production and Allocation*, Wiley, NewYork, 374–376.

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

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

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

[10] Fraser, D.A.S. (1957), *Nonparametric Methods in Statistics*, New York: John Wiley and Sons.

[11] Fudenberg, D. and D. Kreps (1990), *Lectures on Learning and Equilbirum in Strategic-Form Games*, Mimeo, CORE Lecture Series.

[12] Hart, S. and A. Mas-Colell (2000), "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica* 68, 1127–1150.

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

[14] Lakshmivarahan, S. and M. A. L. Thathachar (1973), "Absolutely Expedient Learning Algorithms for Stochastic Automata," *IEEE Trans. Syst., Man. Cybernetics*, vol. **SMC-3**, 281–286.

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

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

[17] Rustichini, A. (1999), "Optimal Properties of Stimulus–Response Learning Models," *Games Econ. Beh.* **29**, 244–273.

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

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

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

[21] Schlag, K.H. (1999), "Which One Should I Imitate," J. Math. Econ. **31(4)**, 493–522.

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

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

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

[25] Schlag, K.H. (2006b), *Nonparametric Minimax Risk Estimates and Most Powerful Tests for Means*, Unpublished Manuscript, European University Institute.

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

[27] Stoye, J. (2006), *Statistical Decisions under Ambiguity*, Unpublished Manuscript, New York University.

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

[29] Wald, A. (1950), *Statistical Decision Functions*, New York: John Wiley & Sons.