

Alvaro Sandroni *

Department of Managerial Economics and Decision Sciences
Kellogg School of Management
Northwestern University
2001 Sheridan Road Evanston IL 60208
sandroni@kellogg.northwestern.edu

September 8, 2004

Abstract

A decision maker wants to know if an expert has relevant information about a stochastic process. The expert either perfectly knows the probability of all future events or knows nothing about these probabilities. The expert must announce the probability distribution of all future events before any data is observed. An empirical test is called good if it can determine the expert type based on a arbitrarily large data set. In a recent contribution Dekel and Feinberg (2004) propose a test and demonstrate strong results indicating that the test is good. In contrast, I show that a good empirical test does not exist.

*I thankfully acknowledge financial support from the National Science Foundation grant SES 0109650. I also thank Eddie Dekel, Yossi Feinberg, Nabil Al-Najjar and Rakesh Vohra for useful comments. All errors are mine.

1. Introduction

A decision maker, named Alice, wants to know if an expert has relevant information about a stochastic process. The expert is one of two possible types. Either he perfectly knows the probability of all future events or he knows nothing. Before any data is observed the expert announces the probability distribution of all future events (i.e., the expert announces a probability measure on the space of infinite sequences of 0's and 1's). Then, the data unfolds. The central question in this paper is whether there exists an empirical test that Alice could use to determine the type of the expert.

Assume that the probability of 1 is known to be fixed. The expert announces that 1 has probability p . Alice tests the expert as follows: she observes the empirical frequency of 1 and if, after sufficiently many periods, this frequency does not remain sufficiently close to p then she concludes that the expert knows nothing. Only experts who know the relevant probability (or are lucky enough to make an announcement close to the truth) pass this empirical test. Now assume that the probability of 1 may not be fixed. Alice could still test the expert as follows: she checks to see if the empirical frequency of 1 is close to p in the periods that the expert predicted 1 with probability close to p . This is a calibration test. Dawid (1982) shows that the calibration test will be passed (almost surely) by an expert who announces the true data generating process. However, Foster and Vohra (1998) shows that the expert who knows nothing can also pass the calibration test. They show that the uninformed expert has a randomized strategy that will produce calibrated forecasts with probability one (according to the randomized strategy) on any sequence of outcomes. Given that it requires almost no knowledge to be passed, the calibration test is commonly perceived as weak. A *good* empirical test (i.e., a test capable of determining the expert type) must return a passing grade to the informed expert and the uninformed expert must find it near impossible to pass.

A good empirical test would reveal an observable feature of probability measures that cannot be obtained without relevant knowledge. Fudenberg and Levine (1999), Lehrer (2001), Hart and Mas-Colell (2001), Sandroni, Smorodinsky and Vohra (2003) considered stronger forms of calibration tests that are performed on many subsequences. They show that an uninformed expert can produce forecasts that will meet these requirements on any sequence of outcomes. Hence, these tests are not good. They return the same verdict to experts who know the data generating process and to experts who know nothing (except for the test

itself). Sandroni (2003) shows a large class of tests based on sequential forecasts (not necessarily calibration tests) that are not good. However, Sandroni's result is severely limited because it is assumed that Alice will only have bounded data sequence to conduct the test. In particular, the number of data points that will be available to Alice is known to the expert. These results motivate the search for a good empirical test. A result in this direction was obtained in a recent thought provoking contribution by Dekel and Feinberg (2004).

Dekel and Feinberg (2004) note that given any probability measure there exists a first category set that has high probability.¹ Let's call this set the *associated set*. The Dekel and Feinberg (2004) test rejects the probability measure if and only if the observed sequence of data does not belong to the associated set. By definition, the informed expert can pass the test, with high probability, by announcing the correct data generating process. The uninformed expert, on the other hand, faces a serious difficulty. First category sets are small in a topological sense. So, when the uninformed expert chooses a probability measure he commits to a single topologically small set without any knowledge of which sequence of outcomes will occur. It seems intuitively impossible for the uninformed expert to pass this test. Dekel and Feinberg (2004) formalize this intuition by showing that for each data first category set A , there exists only a small set of probability measures (i.e., a first category set in the space of probability measures) that assign high probability to A .

The Dekel and Feinberg (2004) test has not been considered in the aforementioned literature. In this literature the expert is only required to forecast the probability of next period's events given the past observed string of data. To construct the associated set it is necessary to know next period's forecasts given any possible finite history (not just the observed history). Moreover, the Dekel and Feinberg (2004) test is clearly not a calibration test and it is not based on bounded information. Alice can observe as many data points as she wants before deciding on the expert's knowledge. The contribution of this paper is the demonstration that, in spite of Dekel and Feinberg (2004) strong results, the uninformed expert can pass this test.

Let an *empirical test* be an arbitrary function mapping that takes as an input a probability measure and a sequence of outcomes and returns, as an output, either

¹A set is of *first category* if it is a countable union of sets that are nowhere dense. A set is called *nowhere dense* when the interior of its closure is empty.

a pass or a reject. No assumptions are placed on it.² Let's say that *rejection can be avoided with probability $1 - \varepsilon$* if the uninformed expert can select probability measures (possibly at random) such that, for any sequence of outcomes, the realized probability measure will not be rejected in any finite time with probability $1 - \varepsilon$ (according to the randomization used by the uninformed expert). I show that if the test does not reject correct probability measures with probability $1 - \varepsilon$ then rejection can be avoided with probability $1 - \varepsilon$. Hence, no empirical test is always capable of distinguishing the expert who knows the entire data generating process and the expert who knows nothing about the data generating process. This result shows that a good empirical test does not exist.

The main result, combined with an intuitive and informal explanation of the basic argument, is presented in section 2. A brief conclusion and a formal proof follow in the next two sections.

2. Model and Result

Each period either 0 or 1 is observed.³ Let $\Omega = \{0, 1\}^\infty$ be the set of all infinite sequences of outcomes. Let N be the set of natural numbers. Let $\{0, 1\}^t$, $t \in N$, be the t -Cartesian product of $\{0, 1\}$. For every finite history $s_t \in \{0, 1\}^t$, $t \in N$, a *cylinder* with base on s_t is the set $C(s_t) = \{s \in \Omega \mid s = (s_t, \dots)\}$ of all infinite histories whose t initial elements coincide with s_t . Let \mathfrak{F}_t be the σ -algebra that consists of all finite unions of cylinders with base on $\{0, 1\}^t$. Let \mathfrak{F} is the σ -algebra generated by the algebra $\mathfrak{F}^0 \equiv \bigcup_{t \in N} \mathfrak{F}_t$, i.e., \mathfrak{F} is the smallest σ -algebra

which contains \mathfrak{F}^0 . Let $(\Omega, \mathfrak{F}, P)$ be the measure space, where P is a probability measure. Let $\Delta(\Omega)$ be the set of probability measures on (Ω, \mathfrak{F}) .

Before any data is observed, the expert must announce a probability distribution $P \in \Delta(\Omega)$. Based on a sequence of outcomes $w \in \Omega$, an outsider must decide whether or not the expert is informed about the governing data generating process.

Definition 1. *An empirical test is a function $T : \Omega \times \Delta(\Omega) \rightarrow \{0, 1\}$.*

²However, it is assumed that the expert knows the test (which could be a combination of different tests). I refer the reader to Lehrer (2001) for some results in which the expert is uncertain about how he will be tested.

³It is immediate to extend the results to the case where one outcome, out of finitely many possibilities, is observed each period.

When the test returns a 0 the test rejects the probability measure based on the outcome sequence. Conversely, when the test returns a 1 the test does not reject the probability measure based on the outcome sequence. Fix $\varepsilon \geq 0$.

Definition 2. *An empirical test T does not reject the truth with probability $1 - \varepsilon$ if for any $P \in \Delta(\Omega)$*

$$P \{w \in \Omega \mid T(w, P) = 1\} \geq 1 - \varepsilon.$$

A good empirical test must not reject the truth with high probability. Otherwise it will be difficult to interpret the meaning of a rejection. Calibration tests and the test in Dekel and Feinberg (2004) are examples of empirical tests that do not reject the truth.

Definition 3. *Given a test T , let $T_\infty : \Omega \times \Delta(\Omega) \rightarrow \{0, 1\}$ be a function such that $T_\infty(w, P) = 0$ if and only if there exists some finite history $s_t \in \{0, 1\}^t$ and cylinder $C(s_t)$ such that $w \in C(s_t)$ and $T(w', P) = 0$ for all $w' \in C(s_t)$.*

That is, if $T_\infty(w, P) = 0$ then P will be rejected in some finite time (when $w \in \Omega$ is realized). Conversely, if $T_\infty(w, P) = 1$ then P will not be rejected in any finite time (when $w \in \Omega$ is realized).

Definition 4. *Rejection can be avoided with probability $1 - \varepsilon$ if there exists a measure space $(\Delta(\Omega), \mathcal{G}, \zeta)$, where \mathcal{G} is a σ -algebra and ζ is probability measure on $(\Delta(\Omega), \mathcal{G})$, such that for every infinite sequence of outcomes $w \in \Omega$*

$$\zeta \{P \in \Delta(\Omega) \mid T_\infty(w, P) = 1\} \geq 1 - \varepsilon.$$

The measure space $(\Delta(\Omega), \mathcal{G}, \zeta)$ may depend on the test T , but no prior knowledge of the data generating process is required for its construction. So, if it is possible to avoid rejection with probability $1 - \varepsilon$ then the uninformed expert can randomly select probability measures such that, with probability $1 - \varepsilon$ (according to the randomization used by the expert), will not be rejected in finite time, no matter which sequence of outcomes is realized.

Definition 5. *An empirical test T is good if, for some $\varepsilon \in [0, 1]$, T does not reject the truth with probability $1 - \varepsilon$ and it is not the case that rejection can be avoided with probability $1 - \varepsilon$.*

A good empirical test always rejects the uninformed expert with higher probability than the informed expert.

Proposition 1. *A good empirical test does not exist.*

Proposition 1 shows that no empirical test can always determine whether or not an expert has relevant knowledge over the data generating process.

The proof of Proposition 1 assumes that T does not reject the truth with probability $1 - \varepsilon$ and then shows the existence of the measure space $(\Delta(\Omega), \mathcal{G}, \zeta)$ that allows rejection to be avoided with probability $1 - \varepsilon$. The σ -algebra \mathcal{G} is obtained as an extension of the sets under which T is zero and one. This makes both T and T_∞ are \mathcal{G} -measurable. The probability measure ζ is constructed as follows: Fix an arbitrary period $r \in N$. Consider the two-player zero-sum game between an expert and nature. A mixed strategy for the expert is a probability measure $\zeta \in \Delta(\Delta(\Omega))$. A mixed strategy for nature is a probability measure $\tilde{P} \in \Delta(\Omega)$. The expert's payoff is 1 if the realized probability measure P (from the expert's randomization) and the realized outcome sequence w (from nature's randomization) are such that the measure P will not be rejected based on w , before period r . Otherwise, the expert payoff is zero. For every mixed strategy that nature has, there is one strategy for the expert (to announce the truth) that yields payoff $1 - \varepsilon$. By Fan's (1953) Minimax theorem there exists a mixed strategy for the expert that yields expected payoff $1 - \varepsilon$ for expert strategy of nature. That is, there exists a probability distribution $\zeta_r \in \Delta(\Delta(\Omega))$ such that if $P \in \Delta(\Omega)$ is selected according to ζ_r then, with probability $1 - \varepsilon$, P will not be rejected before period r (no matter which outcome sequence is realized). The heart of the argument is that there exists a limit point ζ of a sub-sequence of $\{\zeta_r, r \in N\}$ with the desired property on the open-ended model in which rejection may occur at any point in time.

3. Conclusion

A good empirical test would determine, from an observable sequence of outcomes, if a given probability measure is the data generating process or if it is produced without any knowledge over the data generating process. The main result in this paper shows that it is always possible to pass an empirical test that does not reject the truth (unless constraints are imposed on which probability measures can be announced). Hence, a good empirical test does not exist.

4. Proof of Proposition 1

I will make the following simplifying assumption:

Assumption 1. *For every probability measure $P \in \Delta(\Omega)$ there exists $w \in \Omega$ such that $T_\infty(w, P) = 0$.*

Note that if assumption 1 is not satisfied then there exists $\bar{P} \in \Delta(\Omega)$ such that $T_\infty(w, \bar{P}) = 1$ for all $w \in \Omega$. Hence, the expert could simply announce \bar{P} and would not be rejected in finite time for all sequences of outcomes $w \in \Omega$. In this case, it is straightforward to prove proposition 1.

Let $\bar{\Omega} \equiv \bigcup_{t \in N} \{0, 1\}^t$ be the set of finite histories. Given a finite history $s_t \in \bar{\Omega}$, a *dual cylinder* with base on s_t is the set $D(s_t) = \{P \in \Delta(\Omega) \mid T(w, P) = 0 \text{ for every } w \in C(s_t)\}$. So, $D(s_t) \subset \Delta(\Omega)$ is the set of probability measures rejected at s_t . Let \mathcal{G}_0 be the collections of sets $\{D(s_t), s_t \in \bar{\Omega}\}$. That is, \mathcal{G}_0 is the collection of all dual cylinders. Let \mathcal{G}^0 be algebra generated by, i.e., \mathcal{G}^0 is the smallest algebra that contains \mathcal{G}_0 . The existence of \mathcal{G}^0 is demonstrated in Halmos (1970), chapter 1 – 5, Theorem A and observation (4). From the fact that \mathcal{G}_0 is countable, it follows that \mathcal{G}^0 is also countable (see Halmos (1970), chapter 1 – 5, Theorem C). Let \mathcal{G} be the σ -algebra generated by \mathcal{G}^0 . Let $(\Delta(\Omega), \mathcal{G}, \zeta)$ be the measure space, where ζ is a probability measure on the measurable space $(\Delta(\Omega), \mathcal{G})$. That is, ζ is an extension (by Carathéodory extension theorem) of a probability measure on the countable algebra \mathcal{G}^0 .⁴ Let $\Delta(\Delta(\Omega))$ be the set of probability measures ζ on the measurable space $(\Delta(\Omega), \mathcal{G})$. Let's say that a sequence $\{\zeta^n \in \Delta(\Delta(\Omega)), n \in N\}$ converges to a limit $\zeta \in \Delta(\Delta(\Omega))$ (i.e., $\zeta^n \xrightarrow[n \rightarrow \infty]{} \zeta$) if and only if $\zeta^n(D) \xrightarrow[n \rightarrow \infty]{} \zeta(D)$ for every $D \in \mathcal{G}^0$. Given that \mathcal{G}^0 is countable, it follows by standard diagonalization argument that

every sequence $\{\zeta^n \in \Delta(\Delta(\Omega)), n \in N\}$ has a convergent subsequence. (4.1)

The set $\Delta(\Omega)$ is endowed with the weak topology. That is, $P^n \xrightarrow[n \rightarrow \infty]{} P$ if and only if $P^n(C) \xrightarrow[n \rightarrow \infty]{} P(C)$ for every $C \in \mathfrak{S}^0$. Given that \mathfrak{S}^0 is countable, it follows that by standard diagonalization argument that every sequence $\{P^n \in \Delta(\Omega)$,

⁴Hence, a priori, the measure space $(\Delta(\Omega), \mathcal{G}, \zeta)$ does not depend of any topology that may be imposed on $\Delta(\Omega)$.

$n \in N\}$ has a convergent subsequence. Hence, it follows from that Kantorovich and Akilov (1982), chapter 1 – 5, theorem 2, that $\Delta(\Omega)$ is a compact space.⁵

Consider two probability measures $P \in \Delta(\Omega)$ and $\tilde{P} \in \Delta(\Omega)$. If $P \neq \tilde{P}$ then there exists $C \in \mathfrak{S}^0$ such that $P(C) \neq \tilde{P}(C)$. This follows because both P and \tilde{P} are extensions (by Carathéodory extension theorem) of probability measures on the countable algebra \mathfrak{S}^0 . Hence, $\Delta(\Omega)$ is a Hausdorff space. These results can be summarized as follows: in the weak topology,

$$\Delta(\Omega) \text{ is a compact, Hausdorff space.} \quad (4.2)$$

Definition 6. Given a test T and period $r \in N$, let $T_r : \Omega \times \Delta(\Omega) \rightarrow \{0, 1\}$ be a function such that $T_r(w, P) = 0$ if and only if there exists some cylinder $C(s_r)$ with base on $s_r \in \{0, 1\}^r$ such that $w \in C(s_r)$ and $T(w', P) = 0$ for all $w' \in C(s_r)$.

That is, if $T_r(w, P) = 0$ then P will be rejected at or before period r (when $w \in \Omega$ is realized).

Definition 7. Rejection can be delayed for $r \in N$ periods, with probability $1 - \varepsilon$, if there exists a forecasting scheme $\bar{\zeta}_r$ such that for every infinite sequence of outcomes $w \in \Omega$

$$\bar{\zeta}_r \{P \in \Delta(\Omega) \mid T_r(w, P) = 1\} \geq 1 - \varepsilon.$$

Lemma 1. Fix a period $r \in N$. If a test T does not reject the truth with probability $1 - \varepsilon$ then rejection can be delayed for r periods with probability $1 - \varepsilon - \frac{1}{r}$.

Proof - Let $E^{\tilde{P}}$ be the expectation operator associated with $\tilde{P} \in \Delta(\Omega)$. Let E^ζ be the expectation operator associated with $\zeta \in \Delta(\Delta(\Omega))$. Let $H : \Delta(\Delta(\Omega)) \times \Delta(\Omega) \rightarrow [0, 1]$ be defined by

$$H(\zeta, \tilde{P}) = E^{\tilde{P}} E^\zeta \{T_r\}.$$

⁵To apply this result one must first show that $\Delta(\Omega)$ is a metric space. This is, however, well known. Consider two probability measures $P \in \Delta(\Omega)$ and $\tilde{P} \in \Delta(\Omega)$. Let a metric d be given by

$$d(P, \tilde{P}) = \sum_{r=1}^{\infty} (0.5)^r \sup_{C \in \mathfrak{S}_r} |P(C) - \tilde{P}(C)|$$

Fix $w \in \Omega$, $w = (s_r, \dots)$. The function $T_r(w, \cdot) : \Delta(\Omega) \rightarrow [0, 1]$ is \mathcal{G} -measurable. This follows because $T_r(w, \cdot)^{-1}(0) = D(s_r) \in \mathcal{G}$ and $T_r(w, \cdot)^{-1}(1) = (D(s_r))^c \in \mathcal{G}$. Hence, H is well-defined. In fact, $E^\zeta\{T_r\} = 1 - \zeta(D(s_r))$. So,

$$H(\zeta, \tilde{P}) = \sum_{s_r \in \{0,1\}^r} \tilde{P}(C(s_r)) (1 - \zeta(D(s_r))). \quad (4.3)$$

Given a probability measure $\tilde{P} \in \Delta(\Omega)$, let $\zeta^{\tilde{P}} \in \Delta(\Delta(\Omega))$ be a probability measure that assigns zero measure to any dual cylinders $D(s_t) \in \mathcal{G}_0$, $s_t \in \bar{\Omega}$ that do not contain \tilde{P} . That is, let $\zeta^{\tilde{P}} \in \Delta(\Delta(\Omega))$ be any probability measure such that

$$\zeta^{\tilde{P}}(D(s_t)) = 0 \text{ whenever } \tilde{P} \notin D(s_t).$$

Note that, by assumption 1, such probability measure $\zeta^{\tilde{P}}$ exists (although it may not be unique) because at least one dual cylinder $D(s_t)$ contains \tilde{P} . By (3.3),

$$H(\zeta^{\tilde{P}}, \tilde{P}) \geq \sum_{s_r \in \{0,1\}^r, \tilde{P} \notin D(s_r)} \tilde{P}(C(s_r)). \quad (4.4)$$

By definition, $\tilde{P} \in D(s_r)$ if and only if $T_r(w, \tilde{P}) = 0$ for every $w \in C(s_r)$. Hence,

$$\tilde{P} \left\{ w \in \Omega \mid T_r(w, \tilde{P}) = 1 \right\} = \sum_{s_r \in \{0,1\}^r, \tilde{P} \notin D(s_r)} \tilde{P}(C(s_r)). \quad (4.5)$$

By 3.4 and 3.5,

$$H(\zeta^{\tilde{P}}, \tilde{P}) \geq \tilde{P} \left\{ w \in \Omega \mid T_r(w, \tilde{P}) = 1 \right\}.$$

By definition if $T_r(w, \tilde{P}) = 0$ then $T_\infty(w, \tilde{P}) = 0$. So,

$$H(\zeta^{\tilde{P}}, \tilde{P}) \geq \tilde{P} \left\{ w \in \Omega \mid T_r(w, \tilde{P}) = 1 \right\} \geq \tilde{P} \left\{ w \in \Omega \mid T_\infty(w, \tilde{P}) = 1 \right\} \geq 1 - \varepsilon.$$

Thus,

$$H(\zeta^{\tilde{P}}, \tilde{P}) \geq 1 - \varepsilon \Rightarrow \min_{\Delta(\Omega)} \sup_{\Delta(\Delta(\Omega))} H \geq 1 - \varepsilon.$$

It is immediate from 3.3 that the function H is continuous on \tilde{P} . This follows because if \tilde{P}^n converges to \tilde{P} then all $\tilde{P}^n(C(s_r))$ converges to $\tilde{P}(C(s_r))$, $s_r \in \{0,1\}^r$, and there are only finitely many such dual cylinders. It follows from the

linearity of the expectation operator that H is linear on both arguments. By 3.2, $\Delta(\Omega)$ is a compact, Hausdorff space. Hence, by Fan's (1953) Minimax Theorem (2),

$$\min_{\Delta(\Omega)} \sup_{\Delta(\Delta(\Omega))} H = \sup_{\Delta(\Delta(\Omega))} \min_{\Delta(\Omega)} H.$$

Thus, there exists $\bar{\zeta}_r \in \Delta(\Delta(\Omega))$ such that for every $\tilde{P} \in \Delta(\Omega)$, $H(\bar{\zeta}_r, \tilde{P}) \geq 1 - \varepsilon - \frac{1}{r}$. However, if \tilde{P} assigns full mass to a single $w \in \Omega$ then

$$H(\bar{\zeta}_r, \tilde{P}) = \bar{\zeta}_r \{P \in \Delta(\Omega) \mid T_r(w, P) = 1\}.$$

This follows because for a given $w \in \Omega$, $E^\zeta\{T_r\} = \zeta \{P \in \Delta(\Omega) \mid T_r(w, P) = 1\}$. Therefore, for every $w \in \Omega$,

$$\bar{\zeta}_r \{P \in \Delta(\Omega) \mid T_r(w, P) = 1\} \geq 1 - \varepsilon - \frac{1}{r}.$$

■

Remark 1. It is possible to topologize $\Delta(\Delta(\Omega))$ making it a compact, Hausdorff space such that H is continuous on ζ (the argument is analogous to the one made for $\Delta(\Omega)$). Thus, Fan's (1953) Minimax Theorem (1) replaces "sup" with "max." It follows that under the assumptions of Lemma 1, rejection can be delayed for r periods with probability $1 - \varepsilon$ (instead of $1 - \varepsilon - \frac{1}{r}$). This stronger result, however, is not needed for the proof of proposition 1.

Proof of Proposition 1 - Assume that the empirical test T does not reject the truth with probability $1 - \varepsilon$. Let $\{\bar{\zeta}_r, r \in N\}$ be a sequence of probability measures on $\Delta(\Omega)$ as defined in Lemma 1. By 3.1, there exists a subsequence, also indexed by r , that converges to $\bar{\zeta}$. Fix an arbitrary sequence of outcomes $w \in \Omega$, $w = (s_t, \dots)$. For any given period $t \in N$, and dual cylinder $D(s_t) \in \mathcal{G}^0$. By definition,

$$\bar{\zeta}_r(D(s_t)) \xrightarrow{r \rightarrow \infty} \bar{\zeta}(D(s_t)).$$

If $r \geq t$ and the first t elements $s_r \in \{0, 1\}^r$ coincide with $s_t \in \{0, 1\}^t$ then, by definition, $D(s_t) \subset D(s_r)$. By Lemma 1, $\bar{\zeta}_r(D(s_r)) \leq \varepsilon + \frac{1}{r}$. It follows that $\bar{\zeta}_r(D(s_t)) \leq \varepsilon + \frac{1}{r}$ if $r \geq t$. Hence, $\bar{\zeta}(D(s_t)) \leq \varepsilon$.

Given that $D(s_t) \subset D(s_{t+1})$, it follows that $D(s_t) \uparrow D$, where $D \in \mathcal{G}$ is the union of all $D(s_t)$, $t \in N$. It follows that $\bar{\zeta}(D) \leq \varepsilon$. By definition, the complement of D is $\{P \in \Delta(\Omega) \mid T_\infty(w, P) = 1\}$. Hence,

$$\zeta \{P \in \Delta(\Omega) \mid T_\infty(w, P) = 1\} \geq 1 - \varepsilon.$$

So, rejection can be avoided with probability $1 - \varepsilon$. ■

References

- [1] Dawid, A. (1982) "The Well Calibrated Bayesian," (with discussion) *Journal of the American Statistical Association* **77**, 605 – 613.
- [2] Dekel and Feinberg (2004) "A True Expert Knows which Question Should be Asked," mimeo.
- [3] Kantorovich, L. and G. Akilov (1982): *Functional Analysis*. Pergamon Press.
- [4] Fan, K. (1953) "Minimax Theorems," *Proceedings of the National Academy of Science U.S.A.*, **39**, 42-47.
- [5] Foster, D. and R. Vohra (1998) "Asymptotic Calibration," *Biometrika*, **85-2**, 379 – 390.
- [6] Fudenberg, D. and D. Levine (1995), "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, **19**, 1065-1089.
- [7] Hart, S. and A. Mas-Colell (2001) "A General Class of Adaptive Strategies", *Journal of Economic Theory*, **98**, 26-54.
- [8] Halmos, P. (1970): *Measure Theory*. Springer-Verlag.
- [9] Lehrer, E. (2001) "Any Inspection Rule is Manipulable," *Econometrica*, **69-5**, 1333 – 1347.
- [10] Sandroni, A., R. Smorodinsky and R. Vohra (2003) "Calibration with Many Checking Rules," *Mathematics of Operations Research* **28-1**, 141-153.
- [11] Sandroni (2003) "The Reproducible Properties of Correct Forecasts," *International Journal of Game Theory*, **32**, 151-159.