A Category Test that Cannot be Manipulated*

Eddie Dekel† and Yossi Feinberg‡

Preliminary, July 2005§

Abstract

In Dekel and Feinberg (2004) we suggested a test for discovering whether a potential expert is informed of the distribution of a stochastic process. This *category test* requires predicting a "small"– category I – set of outcomes. In this paper we show that under the continuum hypothesis there is a category test that cannot be manipulated, i.e. such that no matter how the potential expert randomizes his prediction, there will be realizations where he will fail to pass the test with probability 1. The set of realizations where he fails can be made large – a category II set. Moreover, this result holds for the finite approximations of the category test where the non-expert is failed in finite time and the expert is failed with small probability. **JEL Classification:** K9

1 Introduction

In Dekel and Feinberg (2004) we suggested a test to determine whether a potential expert knows the distribution governing a stochastic process. The tester is completely uninformed and non-Bayesian, in the sense that she does not have a prior distribution over the possible distributions that may govern the stochastic process, nor does she have a prior over the probability that she is facing an expert. We showed that for each predicted probability measure P there exists a (category I) set S_P such that $P(S_P) = 1$ and that, for any such (category I) set, the set of measures that assign positive probability to S_P is small, in the sense that it is a category I set of measures in the space of probability measures.[1] Thus we

*We would like to thank Alvaro Sandroni and Eilon Solan for helpful comments.

†*Department of Economics, Northwestern University, Evanston, USA and Tel-Aviv University, Tel-Aviv, Israel.* e-mail: dekel@northwestern.edu

‡*Stanford Graduate School of Business, Stanford CA 94305-5015.* e-mail: yossi@gsb.stanford.edu

§Please contact authors before quoting.

[1] A category I set of outcomes is a countable union of nowhere dense sets (i.e., sets whose closure has an empty interior).

showed the existence of a class of tests such that "most" predictions—other than predicting the actual distribution (or one of the other members in a category I set of distributions)—will almost surely fail any such test, and the knowledgeable expert will almost surely pass. We also provided a finite approximation for these tests: for any given $\varepsilon > 0$ any prediction can be tested with a set that will fail all but a category I set of predictions in finite time and will fail the expert with probability of no more than ε.

However, it is conceivable that an uninformed expert might still be able to make a randomized prediction that would pass a test on every realization of the process almost surely (with respect to his randomized strategy), i.e., one could potentially *manipulate* the test. In fact, in response to our paper, Sandroni (2005a) argued that there exists a category test that can be manipulated. In this paper we provide a category test that cannot be manipulated: a test from the class of tests presented in Dekel and Feinberg (2004) such that no matter what randomized prediction a non-expert makes, he is guaranteed to fail the test on some realizations. Moreover, the set of realization on which a potential manipulator will fail can be made larger than the set of realizations where the expert passes, specifically it can be guaranteed to be a category II set. It turns out that the finite approximation of this test also is not manipulable: the uninformed expert is assured to fail on a category II set of realizations even though the test must determine the non-expert in finite time while not failing the expert with high probability.

These results distinguish such "category" tests from the well studied calibration tests not only in the formal interpretation of what a test is supposed to accomplish, but also in whether a non-expert can manipulate the test.[2] As was shown by Lehrer (2001), Kalai, Lehrer and Smorodinsky (1999), Fudenberg and Levine (1999), Sandroni, Smorodinsky and Vohra (2003), and Sandroni (2003) large classes of generalizations and variations of calibration tests can be manipulated. Such manipulation results tell us these tests require a large set of outcomes on which a prediction passes the test. This can be seen in Sandroni (2005b) who considers tests that not only reject a non-expert in finite time but also must pass the expert in finite time. This amounts to asking the predictor to pick a finite number of periods and determine the outcome of the test by the prediction up to that period. In particular, the predicted set of outcomes must be an open set, hence is "too large". In all these papers, a randomized strategy by the non-expert can pass a calibration test on *every* realization with probability 1 (with respect to the random prediction). In contrast, our category test guarantees failure on a set of outcomes that is not small no matter what randomized strategy the non-expert uses (obviously without hindering the guarantee that the informed expert will pass the test with high probability).

[2]See Dawid (1982,1985) and Foster and Vohra (1998) for early papers on calibration.

2 An Unmanipulable Test

Consider the set of realizations of a stochastic process $\Omega = \{0,1\}^{\aleph_0}$ governed by a probability distribution. Let $\Delta(\Omega)$ denote the set of probability measures over Ω endowed with the σ-field generated by the finite cylinders. Let $t : \Delta(\Omega) \longrightarrow 2^{\Omega}$ denote a test. The interpretation is that if a predictor proposes the distribution P then he passes the test t if and only if the realization $\omega \in \Omega$ of the process satisfies $\omega \in t(P)$.

A test does not reject the true expert with probability 1 if:

$$P(t(P)) = 1 \text{ for every } P \in \Delta(\Omega), \tag{1}$$

i.e., the probability of passing the test when the distribution that governs the realization is P and the reported prediction is also P must be one.

Calibration tests can be formulated as tests that satisfy the property in (1).[3] The problem with calibration tests is that they are deemed "bad" because a non-expert can manipulate them. That is, a non-expert can randomize among the predictions P such that for *every* realization of the process the test will be passed with probability 1 on the realization (probability 1 with respect to the randomized prediction). Formally:

$$\text{There exists } \mu \in \Delta(\Delta(\Omega)) \text{ such that, for every } \omega \in \Omega,\ \mu\{P|\omega \in t(P)\} = 1. \tag{2}$$

In other words, by randomly predicting P according to $\mu \in \Delta(\Delta(\Omega))$ an uninformed predictor will pass a calibration test t at every realization of the process. We say that a test can be *manipulated* (sometimes referred to as passed by a non-expert) if (2) holds. Conversely, a test t cannot be manipulated if

$$\text{For every } \mu \in \Delta(\Delta(\Omega)) \text{ there exists an } \omega \in \Omega, \text{ such that } \mu\{P|\omega \in t(P)\} < 1. \tag{3}$$

It is plausible to require even more, namely that a test that cannot be manipulated will guarantee failure of a non-expert at some realization:

$$\text{For every } \mu \in \Delta(\Delta(\Omega)) \text{ there exists an } \omega \in \Omega, \text{ such that } \mu\{P|\omega \in t(P)\} = 0. \tag{4}$$

[3] By reporting a distribution P one can construct the answers to a calibration test (which is applied to the conditional empirical distribution of the subsequent period(s) along each realized path). Similarly, any randomization over predictions generates random answers to a calibration test. Conversely, any prediction evaluated by a calibration test specifies a (predicted) distribution of the subsequent period(s) along every history and hence generates an overall distribution P. Similarly, for any sequence of randomized answers to which a calibration test can be applied one can associate a randomization over the set of measures.

But even the requirement in (4) only asks for failure in a single realization which motivates the following definition. A test t *cannot be manipulated on a set of points that is of category II* if

$$\text{For every } \mu \in \Delta(\Delta(\Omega)) \text{ there exists a category II set } S \subset \Omega \quad (5)$$
$$\text{such that for every } \omega \in S \text{ we have } \mu\{P|\omega \in t(P)\} = 0.$$

Proposition 1 *Assume the continuum hypothesis. There exists a test that does not reject the true expert with probability* 1 *and that cannot be manipulated on a set of points that is of category II.*

The implication of the proposition is that there is a test which guarantees that no matter which randomized prediction the predictor employs, he is guaranteed to fail on a "large" (in the sense of category) set of points, while the test still passes the expert with probability 1. Obviously, the expert also passes on a small set of realizations, but this small set has probability 1 when the true distribution is predicted.

The test that cannot be manipulated is a test from the class of tests we defined in Dekel and Feinberg (2004). We now call these *category tests* and they are defined as tests t which satisfy:

$$\text{For every } P \text{ we have that } t(P) \text{ is category I set such that } P(t(P)) = 1 \quad (6)$$

By definition category tests satisfy the property in (1). Recall that a category I set is a countable union of nowhere dense sets, i.e., a countable union of sets such that the interior of their closure is empty. In Dekel and Feinberg (2004) we showed that there exist category tests. Furthermore, we showed that for every category I set H the set of measures P such that $P(H) > 0$ is "small" in the sense that it is a category I set of *measures* in the space of probability measures $\Delta(\Delta(\Omega))$.

Before we prove the proposition we establish some preliminary results.

First we note that any subset of a category I set is also a category I since any subset of a nowhere dense set is nowhere dense. In addition, any countable union of category I sets is a category I set. A set is called a category II set if it is not a category I set. Note, that the complement of any category I set in Ω is a category II set, but the complement of a category II set can be larger than a category I set.

A set L is called *Lusin set* if L is an uncountable set such that every uncountable subset of L is of category II. The existence of a Lusin set in $[0, 1]$ was shown by Lusin (1914) under the continuum hypothesis. In fact, every category II set contains a Lusin set (see Proposition

20.1 in Oxtoby (1980)). In the Appendix we show that there exists a Lusin set L in the space $\Omega = 2^{\aleph_0}$.

Given a randomized prediction $\mu \in \Delta(\Delta(\Omega))$ we define the measure $\bar{\mu} \in \Delta(\Omega)$ as:

$$\bar{\mu}(E) = \int_{\Delta(\Omega)} P(E)d\mu(P) \tag{7}$$

for every measurable set E. This measure is sometimes refereed to as the "center of gravity" of the measure μ. Note that since Ω is a compact metric space so is $\Delta(\Omega)$ in the weak* topology (cf. Theorem 6.4 in Parthasarathy (1967)). By the definition of the weak* topology we have that for every continuous function $f \in C(\Omega)$ the functional $f(P) = \int_\Omega f(\omega)dP(\omega)$ is a continuous functional on $\Delta(\Omega)$. In particular the continuous functionals on $\Delta(\Omega)$ separate points. From the convexity and compactness of $\Delta(\Omega)$ in the weak* topology we have that the generalized integral $\int_{\Delta(\Omega)} Pd\mu(P)$ exists in the sense that for every linear functional Λ on $\Delta(\Omega)$ we have

$$\Lambda(\bar{\mu}) = \int_{\Delta(\Omega)} (\Lambda(P))d\mu(P) \tag{8}$$

and $\bar{\mu}$ is a probability measure. See Theorems 3.27 and 3.28 in Rudin (1991). Since the measure $\bar{\mu}$ must satisfy

$$\int_\Omega f(\omega)d\bar{\mu} = \int_{\Delta(\Omega)} (\int_\Omega f(\omega)dP)d\mu(P) \tag{9}$$

for every continuous function f we have that regularity implies that (7) is well defined.[4] To see this, consider first a closed set E, we have that $\nu(E) = \inf\{\int f d\nu | f \geq \chi_E\}$ where χ_E is the characteristic function of E. In particular, this holds for $\nu = \bar{\mu}$ as well. By regularity $\nu(G) = \sup\{\nu(E) | E \text{ is closed, } E \subset G\}$ for every measurable set G. Hence we have measurability of $P(E)$ for measurable sets E and $\int_{\Delta(\Omega)} P(E)d\mu(P)$ is defined and coincides with $\bar{\mu}$ as required.

Proof of Proposition 1. Fix an arbitrary category test $\bar{t}$ and a Lusin set $L \subset \Omega$. We define the test t as follows:

$$t(P) = (\bar{t}(P) \setminus L) \cup \{\omega \in L | P(\{\omega\}) > 0\}. \tag{10}$$

The test t maps a probability measure P to a set that only contains points from L if these are atoms of the distribution P. We need to show that t as defined in (10) is indeed a category test and that t cannot be manipulated on a set of category II points.

[4]Since Ω is a separable metric space so is $\Delta(\Omega)$ and the Borel probability measures in $\Delta(\Omega)$ and $\Delta(\Delta(\Omega))$ are regular (see Parthasarathy (1967).

First note that $\bar{t}(P) \setminus L$ is a category I set since it is a subset of the category I set $\bar{t}(P)$. Since P has at most a countable number of atoms the set $\{\omega \in L | P(\{\omega\}) > 0\}$ is countable and a union of a category I set with a countable (hence category I) set is also a category I set. We conclude that $t(P)$ is a category I set.

Since $\bar{t}(P) \cap L$ is a category I set included in the Lusin set L we have that $\bar{t}(P) \cap L$ is a countable set. Hence the set $\bar{t}(P) \setminus L = \bar{t}(P) \setminus (\bar{t}(P) \cap L)$ is measurable since it is the set difference of a measurable set and a countable set. We have

$$P(\bar{t}(P)) = P(\bar{t}(P) \setminus L) + P(\bar{t}(P) \cap L) = P(\bar{t}(P) \setminus L) + \sum_{\{\omega \in \bar{t}(P) \cap L | P(\{\omega\}) > 0\}} P(\{\omega\}). \quad (11)$$

since $\bar{t}(P) \cap L$ is countable. Since $\bar{t}$ is a category test we have $P(\bar{t}(P)) = 1$ and so P has no atoms outside $\bar{t}(P)$ which together with (11) implies

$$P(t(P)) = P(\bar{t}(P) \setminus L) + \sum_{\{\omega \in L | P(\{\omega\}) > 0\}} P(\{\omega\}) = P(\bar{t}(P)) = 1. \quad (12)$$

We have shown that for all P the set $t(P)$ is a category I set and $P(t(P)) = 1$ hence t is a category test.

Consider any given randomized prediction $\mu \in \Delta(\Delta(\Omega))$ where we consider $\Delta(\Omega)$ endowed the Borel σ-field generated by the weak* topology. We now show that for every randomized prediction μ there is a category II set of realizations S such that for all $\omega \in S$ we have $\mu(\{P | \omega \in t(P)\}) = 0$. Let $\omega \in L$ be a point in the Lusin set. We first note that the set $\{P | \omega \in t(P)\} \subset \Delta(\Omega)$ is measurable. Since $\omega \in L$ we have that $\omega \in t(P)$ if and only if $P(\{\omega\}) > 0$ by the definition in (10). Hence for every $\omega \in L$

$$\{P | \omega \in t(P)\} = \{P | P(\{\omega\}) > 0\} \quad (13)$$

so $\{P | \omega \in t(P)\}$ is the set of all measures with an atom at ω from the Lusin set.[5]

The randomized prediction μ will pass the test t when the realization is $\omega \in L$ with positive probability if and only if $\mu(\{P | \omega \in t(P)\}) > 0$. From (13) we have

$$\mu(\{P | \omega \in t(P)\}) = \mu(\{P | P(\{\omega\}) > 0\}) \quad (14)$$

[5]This subset of $\Delta(\Omega)$ is measurable in the Borel σ-field generated by the weak* topology since it is the countable union of the sets $\{P | P(\{\omega\}) \geq 1/n\}$, $n = 1, 2, 3...$ Each set $\{P | P(\{\omega\}) \geq 1/n\}$ is a closed set in the weak* topology since if $P_i \underset{i \to \infty}{\longrightarrow} P$ with $\{P_i\}_{i=1}^{\infty} \subset \{P | P(\{\omega\}) \geq 1/n\}$ then $P_i = \alpha_i \delta_\omega + (1 - \alpha_i) Q_i$ is a convex combination of a probability measure and the Dirac measure at ω with $\alpha_i \geq 1/n$. Taking a converging subsequence of both the α_i's and the Q_i's (the latter has a converging subsequence by the compactness of $\Delta(\Omega)$ in the weak* topology) we find a limit with an atom of at least size $1/n$ at δ_ω (note the continuity of the multiplication and addition operators in the weak* topology).

so the randomized prediction will pass the test t at $\omega \in L$ with positive probability only if $\mu(\{P|P(\{\omega\}) > 0\}) > 0$. From the definition of $\bar{\mu}$ in (7) we have that

$$\mu(\{P|P(\{\omega\}) > 0\}) > 0 \text{ implies } \bar{\mu}(\{\omega\}) > 0. \tag{15}$$

The set of realizations ω such that $\bar{\mu}(\omega) > 0$ is countable hence the set $S = L \setminus \{\omega|\bar{\mu}(\omega) > 0\}$ is a category II set and for every $\omega \in S$ we have $\bar{\mu}(\omega) = 0$ which implies that $\mu(\{P|\omega \in t(P)\}) = 0$. We have shown that for every $\mu \in \Delta(\Delta(\Omega))$ there is a category II set satisfying (5) as required. ■

2.1 Finitely determined tests

There is naturally an interest in finitely determined tests.[6] Finitely determined category tests t are those where $t(P)$ is a closed and nowhere-dense set and hence for every realization ω not in $t(P)$ the test will fail in finite time. Obviously, these tests can only satisfy $P(t(P)) > 1-\varepsilon$. We conclude by showing that there is a finitely determined category test that cannot be manipulated on a category II set.

For every category test t and every $\varepsilon > 0$ we defined in Dekel and Feinberg (2004) a test $t_\varepsilon : \Delta(\Omega) \longrightarrow 2^\Omega$ such that $P(t_\varepsilon(P)) \geq 1 - \varepsilon$ and $t_\varepsilon(P)$ is a closed and nowhere-dense set for every $P \in \Delta(\Omega)$. Furthermore, $t_\varepsilon(P) \subset t(P)$ for all P, so for every ε and category test t there is a finitely determined test t_ε. These finitely determined tests are ε approximations of category tests.

Proposition 2 *Assume the continuum hypothesis. There exists a category test t such that for every $\varepsilon > 0$ for every $\mu \in \Delta(\Delta(\Omega))$ the randomized prediction will fail the finitely determined test t_ε on a category II set of points with μ-probability 1. Formally:*

$$\begin{gathered}\textit{For every } \mu \in \Delta(\Delta(\Omega)) \textit{ there exists a category II set } S \subset \Omega \\ \textit{such that for every } \omega \in S \textit{ we have } \mu\{P|\omega \in t_\varepsilon(P)\} = 0.\end{gathered} \tag{16}$$

Proof. Let t be as in Proposition 1, that is, a category test. Since $t_\varepsilon(P) \subset t(P)$ we have for every ω and P that $\omega \in t_\varepsilon(P)$ implies $\omega \in t(P)$. Hence for every ω we find

$$\{P \in \Delta(\Omega)|\omega \in t_\varepsilon(P)\} \subset \{P \in \Delta(\Omega)|\omega \in t(P)\}. \tag{17}$$

Applying (17) for every $\omega \in S$ where S is the category II set corresponding to μ as in

[6] For example, calibration tests are defined as limits of such finitely observed events.

Proposition 1, we have

$$\begin{aligned} \mu(\{P &\in \Delta(\Omega)|\omega \in t_\varepsilon(P)\}) \leq \\ \mu(\{P &\in \Delta(\Omega)|\omega \in t(P)\}) = 0 \end{aligned} \tag{18}$$

where the final equality follows from Proposition 1. ■

3 Appendix

Proof that there exists a Lusin set in Ω. The proof follows from viewing points in $\Omega = 2^{\aleph_0}$ as the dyadic (binary) expansion of points in $[0, 1]$. The dyadic expansion of the points in a Lusin set $L \subset [0, 1]$ must be a Lusin set in $\Omega = 2^{\aleph_0}$.

The dyadic expansion is unique for all but a countable set of points in $[0, 1]$. Assume by contradiction that the set of dyadic expansions of members of L, which we denote by $\bar{L}$, is not a Lusin set in Ω. Then we could find an uncountable category I subset of $\bar{L}$ in $2^{\aleph_0}$. It suffices to show that the inverse of the dyadic expansion maps a closed nowhere dense set in Ω to a closed nowhere dense set in $[0, 1]$ (hence a countable union of such sets will be mapped to at most a countable union of such sets). This will show that a category I set is mapped to a category I set and will contradict L being a Lusin set since the dyadic expansion and its inverse maps uncountable sets to uncountable sets.

Consider a closed set $S \subset \Omega$. Since S is closed under the product topology its map under the inverse of the dyadic expansion is closed; this is because convergence of the dyadic expansion implies convergence in $[0, 1]$. We need to show that if S is nowhere dense in Ω its preimage is nowhere dense in the interval. Consider any point in the interval and any open neighborhood of that point. Since the dyadic open intervals generate the same topology generated by open intervals we can find a dyadic interval in the open neighborhood which contains the point. The dyadic interval is open in Ω and hence contains points outside the nowhere dense set S. Hence these points are mapped in the inverse of the dyadic expansion to points in the dyadic interval. We conclude that every point in $[0, 1]$ has points from outside the image of S in any open neighborhood and the image of S is therefore nowhere dense as required. ■

References

[1] Dawid, A. P. (1982) "The Well-Calibrated Bayesian." *Journal of the American Statistical Association* **77** (379), 605–613.

[2] Dawid, A. P. (1985) "Calibration-Based Empirical Probability." *The Annals of Statistics* **13** (4), 1251–1274.

[3] Dekel, E., and Feinberg, Y. (2004) "Non-Bayesian Testing of an Expert." *Working paper.*

[4] Foster, D. P., and Vohra, R. V. (1998) "Asymptotic Calibration." *Biometrika* **85** (2), 379–390.

[5] Fudenberg, D., and Levine, D. K. (1999) "Conditional Universal Consistency." *Games and Economic Behavior* **29** (1-2), 104–130.

[6] Kalai, E., Lehrer, E., and Smorodinsky, R. (1999) "Calibrated Forecasting and Merging." *Games and Economic Behavior* **29** (1-2), 151–159.

[7] Lehrer, E. (2001) "Any Inspection Rule is Manipulable." *Econometrica* **69** (5) 1333–1347.

[8] Lusin, N. (1914) "Sur un problèmé de M. Baire ." *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris* **158** 1258-1261.

[9] Oxtoby, J. C. (1980) *Measure and category. A survey of the analogies between topological and measure spaces.* Second edition. Springer-Verlag, New York-Berlin.

[10] Parthasarathy, K. R. (1967) *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3. Academic Press, New York-London.

[11] Rudin, W. (1991) *Functional analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill, New York.

[12] Sandroni, A. (2003) "The Reproducible Properties of Correct Forecasts." *International Journal of Game Theory* **32** (1), 151–159.

[13] Sandroni, A. (2005a) "Testing Knowledge." mimeo.

[14] Sandroni, A. (2005b) "A Negative Report on the Existence of a Good Finite Empirical Test." mimeo.

[15] Sandroni, A., Smorodinsky, R., and Vohra, R. V. (2003) "Calibration with Many Checking Rules." *Mathematics of Operations Research* **28** (1), 141–153.