Calibration with Many Checking Rules*

Alvaro Sandroni † Rann Smorodinsky ‡
Rakesh V. Vohra §

August 27, 1999

Abstract

We propose a universal prediction scheme that calibrates countably many history-based and forecast-based checking rules.

*Sandroni and Smorodinsky thank the financial support from the NSF Grant SBR 9730385, BSF Grant 97-00113/1 and the Bergmann Memorial Research Grant

†Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. (sandroni@nwu.edu)

‡Davidson Faculty of Industrial Engineering and Management, Technion Haifa 32000 Israel. (rann@ie.technion.ac.il), and Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208.

§Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. (r-vohra@nwu.edu)

"A good Christian must beware of mathematicians and those soothsayers who make predictions by unholy methods, especially when their predictions come true, lest they ensnare the soul through association with demons." St. Augustine, De genesis ad litteran, Book II.

1 Introduction

In each period, one of finitely many possible states will occur. To simplify matters, assume just two possible states, a and b. We must predict the probability that state a will be realized next period. We must do so without *any* prior knowledge about the underlying process determining the states in each period. Thus, the forecasts must be based on "pure empiricism," i.e., they may depend on past outcomes, but not on any prior understanding of the data. Given sufficiently many observations, how accurate can our forecasts be?

We are not interested in the accuracy of the forecasts with respect to a particular realization of the states because we do not know which sequence of states will be realized. So, we seek guarantees on the accuracy of the forecasting method for *every* possible infinite string of data on $\{a, b\}^{\infty}$. However, it is by no means obvious what guarantees can be offered. So, we first formalize different notions under which the forecasts may (or may not) comply with the data.

Let a *history-based checking rule* be an arbitrary function of finite histories to $\{0, 1\}$. Given a history-based checking rule, and a forecasting scheme, let a *forecast-based checking rule* be a function equal to one if the history-based checking rule is one and the forecast of a is p and equal to zero otherwise. History-based checking rules and forecast-based checking rules are said to be *active* (or *inactive*) if they assume the value one (or zero). The checking rules determine the subsequences in which the time average forecasts will be compared with the empirical frequencies. For example, the history-based checking rule that is active in the odd periods can be used to compare empirical frequencies and time average forecasts in the odd periods. The forecast-based checking rules associated with this history-based checking rule compares empirical frequencies when the forecast of a is p in odd periods. For example, assume that the forecast of a is 0.2 when a occurs and 0.8 when b occurs. The data always alternates aa and bb (the first state is a). So, the time average forecast of a (0.5) is equal to the empirical frequency of a in the entire sequence, but, in the periods where a was forecasted with probability 0.8, the empirical frequency of a is 0.5. In the odd periods, the time average forecast of a is equal to the empirical frequency of a (0.5), but, a never occurred after a forecast of 0.8 (for a) was announced in an odd period.

The *calibration score* of a checking rule (history or forecast based) is the difference between the frequency of a and the time average of the forecasts of a in the periods that the checking rule was active. A forecasting scheme *calibrates* a checking rule if, given any data, the calibration score is eventually zero.

Given this setup three questions are natural. What kinds of checking rules can be calibrated? What sets of checking rules can be calibrated simultaneously? What are the consequences of being able to calibrate many checking rules?

Foster and Vohra (98) demonstrate the existence of a forecasting scheme that calibrates the forecast-based checking rules associated with the always active history-based checking rule. That is,they show that there is a forecasting scheme with the property that, given any data, the frequency of a, in the periods which a was forecasted with probability p, is also p.

Lehrer (98) shows that there exists a forecasting scheme which simultaneously calibrates countably many history-based checking rules. However, Lehrer's (98) forecasting scheme does not calibrate the associated forecast-based checking rules. So, the time average forecast of a and the average frequencies of a may become close in the entire sequence, in the odd periods, in the even periods, after a, b, or ab has occurred, etc. However, the average frequency of a may be substantially different from p in the subsequence in which a was forecasted with probability p.

Our main result is the existence of a forecasting scheme which simultaneously calibrates countably many history-based checking rules and all the associated forecast-based checking rules.

Can stronger results be obtained? Note that to calibrate a checking rule means that, at some point in the future, if we look backwards, we will see that the time average of the forecasts are close to the empirical frequencies. This does not guarantee that from this point onwards the forecasts will be almost correct. In fact, we show that there is no forecasting scheme that can guarantee this. We also show that no forecasting scheme can calibrate all checking rules simultaneously. In this sense, calibrating countably many history-based checking rules and all the associated forecast-based checking rules is, perhaps, the limit of what can be achieved by pure empiricism.[1]

1.1 Implications of the Main Result

1.1.1 Testing Knowledge

Suppose that we must distinguish between forecasters who have some prior knowledge of the stochastic process from those who know naught but the data itself. If the forecasters can only make deterministic predictions, the hypothesis that they know the stochastic process can be rejected if a realized state contradicts the forecast.

The task of designing a test that differentiates the knowledge of forecasters is more difficult if probabilistic forecasts are permitted. If we assume that the states are always generated under identical conditions and the forecasters must repeat "once and for all" forecasts, we can still reject the hypothesis that a forecaster knows the relevant probabilities if the empirical frequencies contradict them. If we cannot assume that the states are always generated under identical conditions then we might consider comparing empirical frequencies with the forecasts in the periods that the forecasts were similar (we can reasonably assume that, according to the forecaster, in these periods the conditions are roughly identical). We can perform a more elaborate test; compare the time average of the forecasts of a and the empirical frequencies of a in the periods where a was forecasted with probability p, in the even periods where a was forecasted with probability p, when the last four outcomes were identical and a was forecasted with probability p, etc.

Our main result implies that we will not be able to differentiate forecasters who know the stochastic process from those whose knowledge was obtained by pure empiricism because they will both be able to drive the calibration scores to zero.

1.1.2 Setting Standards

In any good model of human behavior, it is posited that beliefs will be revised when the assumptions about the stochastic process are contradicted by the

[1] A generalization of our main result can be demonstrated. Given an arbitrary probability measure λ on the space of the history-based checking rules, there exists a forecasting scheme which simultaneously calibrates a full λ-measure set of history-based checking rules and all the forecast-based checking rules associated with them.

data. The difficulty is to define the appropriate sense in which the beliefs and the observed data should eventually comply. One definition, rational expectations, is dominant in economics. It assumes that agents' beliefs are identical to the truth. An alternative, due to Kurz (94), requires that expected and realized empirical frequencies are asymptotically equal. Another alternative is the temporary equilibrium literature (see Grandmont (82)). It assumes that agents' beliefs are such that there exists no arbitrage in the economy. The conclusions obtained in these different theories are also quite different. Therefore, the sense in which the beliefs relates to the data matters. The consistency between time average forecasts and empirical frequencies of countably many history-based checking rules and the associated forecast-based checking rules demonstrated in this paper can be interpreted as a minimal property that the agents' forecasts should satisfy in the long run because it can be obtained without prior knowledge of the stochastic process.

1.1.3 Calibration and Merging

In classical statistical inference, it is usual to demonstrate the existence of a forecasting scheme which, almost surely, merges with the true stochastic process - provided that the true stochastic process is in some class. The approach in this paper is different because we do not make assumptions about the stochastic process. It is, therefore, interesting to see the properties of our forecasting scheme under some of the usual structural assumptions imposed on the stochastic process. Assume, for example, that the true probability of a is 0.5 in all periods. An agent who adopts a forecasting scheme that calibrates the forecast-based checking rules associated with the always active checking rule (as in Foster and Vohra (98)) must eventually predict that the probability of a is close to 0.5, otherwise the calibration score of the forecast-based checking rules for predictions substantially different from 0.5 will not eventually be close to zero (since those predictions will not match the actual frequency of a). However, consider the forecasts in Lehrer's (98) forecasting scheme (which calibrates countably many checking rules, but not the associated forecast-based checking rules). There is no reason why the forecasts (as opposed to their average) should become close to the truth (0.5) and, in fact, this may never happen.

Assume that the data is a sequence such that a always follows b and, conversely, b always follows a. In this case, the forecasts in Lehrer's (98) scheme in the odd periods and in the even periods will eventually become close to the truth, otherwise the calibration score for the two checking rules active in those respective periods will not eventually be close to zero. However, the forecasts in Foster and Vohra's (98) scheme need not become close to the true probabilities.

Now suppose that the true stochastic process is a time-independent Markovian process of length one. Then, neither the forecasts in Foster and Vohra's (98) scheme nor the ones in Lehrer's (98) scheme need eventually be close to the true probabilities. Therefore, the forecasts generated by these two schemes may be systematically incorrect even if the underlying stochastic process is quite simple.

Now suppose that the forecasting scheme calibrates the forecast-based checking rules associated with the checking rule which is active when the current outcome is a and the checking rule which is active when the current outcome is b. Then, if the true stochastic process is a time-independent Markovian process of length one the agents' forecasts will eventually be close to the true probabilities (otherwise the predictions will not match the empirical frequencies after a or after b is realized).

If the forecasting scheme only calibrates finitely many history-based checking rules and the associated forecast-based checking rules, then the forecasts may not eventually be close to the true probabilities even if the stochastic process is relatively simple (such as a time-independent Markovian process of sufficiently large length or an eternal repetition of the same finite sequence). However, the forecasting scheme proposed in this paper calibrates countably many history-based checking rules and the associated forecast-based checking rules. Hence, in the case of a time-independent Markovian process of arbitrary (and unknown) finite length or in the case of an eternally repeating finite sequence, the forecasts will eventually become close to the truth.

1.1.4 Learning Correlated Equilibrium

We now turn to the question of learning in economics and game theory. The central problem in these learning models is that the relevant variables to be estimated are affected by the learning process itself. Agents are constantly revising their beliefs and, therefore, the relationship between exogenous states of nature and the payoff-relevant variables may not be stationary. Hence, some of the classical statistical inference techniques are not applicable. A possible solution to this problem is to appeal to the Blackwell-Dubins theorem (1962) which shows that, as long as the agent assigns positive probability to the true distribution, agents' posterior beliefs will eventually become identical to the truth - no assumptions on the nature of the true stochastic process are assumed. The problem with this approach is that it is not possible to assign positive probability to all distributions. Moreover, an agent who optimizes according to his beliefs may generate behavior that received zero probability by his own prior. Given a family Φ of stochastic processes, there may exist a belief such that eventually the posterior beliefs will be almost correct under any stochastic process in this family. However, this belief, and the behavior it generates, may not belong to Φ. So, for example, if Φ only contains Markovian distributions, this belief may not be Markovian. In fact, Nachbar (97) shows that, if Φ satisfies certain properties, then "prediction" and "optimization" cannot coexist because any belief will either generate behavior which does not belong to Φ or will not guarantee that the posterior beliefs become almost correct under any stochastic process in Φ.

We demonstrate that the conflict between prediction and optimization does not exist if by prediction we mean that the calibration scores of countably many history-based checking rules and the associated forecast-based checking rules become close to zero. All players can individually choose to have calibrated forecasts. Then, no matter what is the induced behavior, all players' forecasts will be calibrated. Moreover, Sandroni and Smorodinsky (98) show, based on a result by Foster and Vohra (97), that if all players' forecasts are calibrated then the time-average of play (in all subsequences in which the checking rules are active) must converge to the set of correlated equilibria. Therefore, an observer of the play (who uses the checking rules of the players) will not be able to reject the hypothesis that the players are playing a correlated equilibrium. This result need not hold if the players calibrate the history-based checking rules and not the associated forecast-based checking rules.

1.2 Related Literature

The idea that calibration is a desirable property of probabilistic forecasts is due to Dawid (82). He shows that the posterior beliefs of a coherent Bayesian will become calibrated with probability one under the Bayesian forecasters own prior. Subsequently Oakes (85) showed that no deterministic forecasting scheme could be guaranteed to calibrate the forecast-based checking rule associated with the always active checking rule. The existence of a randomized forecasting scheme that calibrates the forecast-based checking rule associated with the always active checking rule was established by Foster and Vohra (98). This result was first generalized by Fudenberg and Levine (97) who show the existence of forecasting schemes that calibrates certain classes of checking rules. Foster and Vohra's (98) result also inspired alternative proofs of the same result by Fudenberg and Levine (95) and Hart and Mas-Collel (96) which use the minimax theorem and the Blackwell approachability theorem. These proofs are conceptually simpler and inspired the proof in this paper.

Variations and strengthenings of Dawid's (82) original notion of calibration were introduced and explored by Kalai, Lehrer and Smorodinsky (95) who introduced the language of checking rules and demonstrated a connection between calibration and merging.

The existence of a forecasting scheme that simultaneously calibrates countable many history-based checking rules was first established by Lehrer (98). His result is based on a generalization of the Blackwell approachability theorem. The main result of this paper uses his result as well as the minimax theorem.

2 Definitions and Result

Let N be the set of natural numbers. Let N_+ be $N \cup \{0\}$. Denote by $S \equiv \{1, ..., n\}$ the state space and S^0 be the set containing only the null history. Let S^t, $t \in N \cup \{\infty\}$, be the $t-$Cartesian product of S, and let $\bar{S} \equiv \cup_{t=0}^{\infty} S^t$ be the set of all finite histories. Given any infinite history $s \in S^\infty$, we denote by s_t the t^{th} coordinate of s and by $s^t = (s_1, s_2, \ldots, s_t) \in S^t$ the prefix of length t of s. Call any infinite history $s \in S^\infty$ a path. In our model, at the end of each stage $t \in N$ a forecaster makes his prediction regarding the probability distribution for the stage and and a realization of S is observed.

Definition 1 Let $\Delta(S)$ be the space of probability distributions over S. A *pure forecasting scheme* is a function $f : \bar{S} \rightarrow \Delta(S)$.

Given any finite history $s^t \in S^t$, a pure forecasting scheme gives the probabilities $f(s^t) = (f_1(s^t), ..., f_n(s^t))$ of the states next period, where $f_i(s^t)$, $i \in S$, is the forecast that state i will occur next period. By Kolmogorov's Extension Theorem (see Shriyaev (1984)) any pure forecasting scheme f determines a unique probability measure f^* on the space of infinite histories S^∞. Denote by Ω the space of all pure forecasting schemes.

Let $(\Delta(S) \times S)^0$ be S^0. Given $t \in N$, let $(\Delta(S) \times S)^t$ be the set of t-histories of forecasts and states. So, $x^t \in (\Delta(S) \times S)^t$, $x^t = (x_1, ..., x_t)$, is a $t-$history where each element x_j is the state realized at period j and the forecast made at period $j-1$. Let $H \equiv \cup_{t=0}^{\infty} (\Delta(S) \times S)^t$ be the set of finite histories of forecasts and states. Let $\Delta(\Delta(S))$ be the set of probability distributions over $\Delta(S)$.

Definition 2 A *mixed forecasting scheme* is a function $\zeta : H \to \Delta(\Delta(S))$.

Given any finite history of states and forecasts, a mixed forecasting scheme assigns probabilities to probability distributions over the state to be realized in the next period. At history $x^t \in (\Delta(S) \times S)^t$, the forecaster randomizes according to $\zeta(x^t)$ and announces the realized forecast at period t. By Kolmogorov's Extension Theorem, any mixed forecasting scheme ζ determines a unique probability measure ζ^* on the space of all pure forecasting schemes Ω.

Definition 3 A *history-based checking rule* is a function $C : \bar{S} \to \{0, 1\}$.

Given $s \in S^\infty$, if $C(s^t) = 1$ then we say that C is *active* at period t. Analogously, if $C(s^t) = 0$ then we say that C is *inactive* at period t. For example, a history-based checking rule could be always active, active in odd periods, active in period t whenever term t of the binary expansion of π is 1. The frequencies of the states will be compared with the predictions in which the history-based checking rule was active.

Definition 4 Given a subset $D \subset \Delta(S)$ and a pure forecasting scheme f, let $\chi^D : \bar{S} \to \{0, 1\}$ be a function defined by $\chi^D(s^t) = 1$ if $f(\mathrm{s}^t) \in \mathrm{D}$ and $\chi_t^D(s) = 0$ otherwise. Given a checking rule C, a subset $D \subset \Delta(S)$ and a pure forecasting scheme f, the *forecast-based checking rule*, C^D, is the function $C^D \equiv C\chi^D$.

So, a forecast-based checking rule is a function $C\chi^D$ which is active (i.e., equal to one) if and only if the history-based checking rule is active and the forecast belongs to D. For example, a forecast-based checking rule could be active when the forecast for state 1 is greater than 0.5, active in the odd periods in which the forecast for state n is between 0.1 and 0.4, active when the last four states were identical and the forecast for any state is smaller than $2/n$, etc. Clearly, the history-based checking rule C is identical to the forecast-based checking rule $C^{\Delta(S)}$.

Definition 5 Given a checking rule C^D, and a pure forecasting scheme f, the *calibration score on* $s \in S^\infty$, at time T, is the $n-$dimensional vector

$$\rho_T(C^D, f, s) \equiv \frac{\sum_{t=0}^{T-1} C^D(s^t)\,(I(s^{t+1}) - f(s^t))}{\sum_{t=0}^{T-1} C^D(s^t)},$$

where $I(s^{t+1})$ is the unit vector in $\Re^n$ with 1 in the s_{t+1}^{th}-coordinate and zero elsewhere.

So, ρ_T is the difference between the vector of empirical frequencies and the average forecasts, in the sequence of checking times when the forecast was in D.

Definition 6 A pure forecasting scheme, f*, calibrates a forecast-based checking rule* C^D *on* $s \in S^\infty$ if $\sum_{t=0}^{\infty} C_0^D(s^t)$ is finite or if

$$\lim_{T\to\infty} \rho_T(C^D, f, s) = \bar{0},$$

where $\bar{0}$ is the n-dimensional vector $(0, \ldots, 0)$, and the equality is coordinate-wise.

So, a pure forecasting scheme calibrates a forecast-based checking rule on a path if the calibration scores converge to zero on this path (whenever the forecast-based checking rule is active infinitely often on this path).

Definition 7 A pure forecasting scheme, f*, calibrates a forecast-based checking rule* C^D if f calibrates C^D on every path $s \in S^\infty$.

So, a pure forecasting scheme calibrates a forecast based checking rule if it brings the calibration score to zero no matter what infinite sequence of states has been realized.

Example 1 Let $S = \{a, b\}$. There is no pure forecasting scheme such that the time average of its forecasts will match the empirical frequencies in all subsequences that a was forecasted with probability $p \in [0, 1]$ (see Oakes (85)).

Consider an arbitrary pure forecasting scheme. Assume that state a occurs if and only if the forecast of a was less than 0.5. Clearly, in the periods that a was forecasted with probability less than 0.5, the empirical frequency of a is greater than 0.5.

Definition 8 A mixed forecasting scheme, ζ, *calibrates a forecast-based checking rule* C^D on $s \in S^\infty$ if ζ^*-almost every pure forecasting scheme f calibrates C^D on s.

That is, a mixed forecasting scheme calibrates a forecast-based checking rule on a path if, the calibration scores of almost every pure forecasting scheme converge to zero on this path (whenever the forecast-based checking rule is active infinitely often on this path). The almost everywhere statement in the definition refers to the fact that the forecasts are chosen according to a probability distribution determined by the mixed forecasting scheme. Our main result implies that a mixed forecasting scheme is immune to the problems presented by example 1 above.

Definition 9 A mixed forecasting scheme, ζ, *calibrates a forecast-based checking rule* C^D if ζ calibrates C^D on every path $s \in S^\infty$.

So, a mixed forecasting scheme calibrates a forecast based checking rule if the calibration scores of almost every pure forecasting scheme converge to zero for *all* sequences of states. We now show the main result of this paper.

Proposition 1 Let $\mathcal{D} = \{D^1, D^2, \ldots\}$ be an arbitrary countable collection of subsets of $\Delta(S)$. Let $\mathcal{C} = \{C_1, C_2, \ldots\}$ be an arbitrary countable collection of history-based checking rules.[2] Then, there exists a mixed forecasting scheme that simultaneously calibrates all forecast-based checking rules $C_j^{D^k}, C_j \in \mathcal{C}$, $D^k \in \mathcal{D}$.

Proof - See section 3.

It follows immediately from proposition 1 that there exists a mixed forecasting scheme that simultaneously calibrates all history-based checking rules in $\mathcal{C}$. Simply consider the trivial partition $\mathcal{D} = \{\Delta(S)\}$. Lehrer (97) shows that there exists a *pure* forecasting scheme that calibrates countable history-based checking rules. This does not follow immediately from proposition 1 (although a separate proof of Lehrer's (97) result, based on proposition 1, can be obtained). Example 1 shows that the general result in proposition 1 cannot be obtained by pure a forecasting scheme.

[2] Clearly, the collections of subsets of $\Delta(S)$ and/or the collection of history-based checking rules may also be finite.

Foster and Vohra's (98) main result is the special case of proposition 1 where $\mathcal{C}$ contains only the always active checking rule and $\mathcal{D}$ is a finite partition of $\Delta(S)$. Foster and Vohra (98) are interested in checking rules which are activated on sequences of positive density, whereas we look at checking rules which are activated infinitely often, even if this occurs with density zero. Moreover, Foster and Vohra (98) show that the calibration score is eventually small - provided that the area covered by each element of the partition of $\Delta(S)$ is also small. We show that the calibration scores converge to zero.

Example 2 There is no mixed forecasting scheme with the property that, given any deterministic process, almost every pure forecasting schemes will merge with the process.

Assume that there are two states 1 and 2. Consider an arbitrary mixed forecasting scheme ζ. Consider the infinite history that is 1 if and only if the joint probability of all forecasts assigning probability greater than 0.5 to 1 is smaller than 0.5. Clearly, on this particular infinite history, the set of pure forecasting schemes which eventually forecasts correctly the next state has zero ζ^*-measure.

The same example demonstrates there is no mixed forecasting scheme that simultaneously calibrates all checking rules. Consider all history-based checking rules. One of them will be active in the periods in which the joint probability of all forecasts assigning probability greater than 0.5 to 1 is smaller than 0.5 (and 1 occurred in the next period). On this path, the set of pure forecasting schemes which calibrates this checking rule has zero ζ^*-measure. However, proposition 1 can be slightly generalized. Let λ be an arbitrary probability measure on the space $\mathcal{M}$ of all history-based checking rules. Let $\mathcal{D} = \{D^1, D^2, \ldots\}$ be an arbitrary countable collection of subsets of $\Delta(S)$. Then, we can show that there exists a full λ-measure $\Theta \subset \mathcal{M}$ of checking rules ($\lambda(\Theta) = 1$) and a mixed forecasting scheme that calibrates all checking rules C^{D^k}, $C \in \Theta$, $D^k \in \mathcal{D}$.

3 Proof of Main Result

To prove our main result we use a minor generalization of proposition 1 of Lehrer (97), which itself generalizes Blackwell's (56) celebrated Approachability Theorem.

Let $\langle , \rangle$ be the inner product in $\Re^n$, $\langle a,b\rangle = \sum_{i=1}^{n} a_i b_i$, $a \in \Re^n$, $b \in \Re^n$. Let $\|\cdot\|$ be the norm in $\Re^n$, $\|a\| = \sqrt{\langle a,a\rangle}$, $a \in \Re^n$.

Lemma A1 Let $(X, \mathcal{B}, \nu)$ be a measure space. Let $\{T_t\}_{t=0}^{\infty}$ be a sequence of random variables taking values in N_+ and let $\{g_t\}_{t=0}^{\infty}$ and $\{\rho_t\}_{t=0}^{\infty}$ be sequences of random variables taking values in $\Re^n$. Let $\{\varepsilon_t\}_{t=1}^{\infty}$ be a sequence of non-negative numbers such that $\sum_{t=1}^{\infty} \varepsilon_t < \infty$. Assume:

1. T_t is non-decreasing, $t \in N$; $T_t - T_{t-1} \leqslant 1$; and $T_0 = 0$;
2. $g_t \in [-1,1]^n$;
3. $T_t - T_{t-1} = 0$ implies $g_t = \bar{0}$;
4. $\rho_t = \frac{T_{t-1}\rho_{t-1} + g_t}{T_t}$; where the ratio is taken for each coordinate of $\Re^n$.
5. $E_\nu \left(\left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle \right) \leqslant\leq \varepsilon_t$, where E_ν is the expectation operator associated with ν, and $\{\varepsilon_t\}_{t=1}^{\infty}$ is a sequence of positive numbers such that

Then, $\nu - a.e.$, $T_t \to \infty$ implies that $\rho_t \to \bar{0}$.

Proof: The case $n = 1$ follows from Proposition 1 of Lehrer (97). We show that the case $n \geq 2$ can be reduced to the case $n = 1$. Let $X \times \{1, \ldots, n\}$ be endowed with the probability $\tilde{\nu}(\cdot, j) = \frac{\nu(\cdot)}{n}$. For any random variable Y on X taking values in $\Re^n$, let $\tilde{Y}$ be a real valued random variable on $X \times \{1, \ldots, n\}$ defined by $\tilde{Y}(x,j) \equiv (Y(x))_j$, where $x \in X$ and $(Y(x))_j$ is the j^{th} coordinate of $Y(x)$.

By definition,

$$E_{\tilde{\nu}}(\tilde{g}_t \frac{\tilde{\rho}_{t-1}}{T_t}) = \frac{1}{n} E_\nu \left(\left\langle g_t, \frac{\rho_{t-1}}{T_t} \right\rangle \right) \Longrightarrow E_{\tilde{\nu}}(\tilde{g}_t \frac{\tilde{\rho}_{t-1}}{T_t}) \leq \frac{\varepsilon_t}{n}.$$

Therefore, we are back to the case $n = 1$ replacing $\tilde{g}_t$ with g_t and $\tilde{\rho}_t$ with ρ_t. Hence,

$$\tilde{\rho}_t \to 0 \ \tilde{\nu} - a.e. \Longrightarrow \rho_t \to \bar{0} \ \nu - a.e.$$

Q.E.D.

Let $\mathcal{R} \equiv \mathcal{C}\text{x}\mathcal{D}$ be a countable collection of checking rules. Let X be $\mathcal{R} \times \Omega$. An element of X is a checking rule c and a pure forecasting scheme f. Fix any path $s \in S^{\infty}$. We now define the following functions:

1. Let $T_t(c, f) \in N$ be the number of times that the checking rule $c \in \mathcal{R}$ was active until period $t-1$, i.e. $T_t = T_t(C^D, f) = \sum_{j=1}^{t} C^D(f, s^j)$.

2. Let $g_t(c, f) = \bar{0}$ if the checking rule c is inactive at period $t-1$ and $g_t(c, f) = s_t - f(s^{t-1})$ if the checking rule c is active at period $t-1$, where $f(s^{t-1}) \in \Delta(S)$ is the forecast announced at period $t-1$ and $s_t \in S$ is the state observed at period t.

3. Let $\rho_t(c, f) \in \Re^n$ be the calibration score of checking rule c at period t. That is $\rho_t = \rho_t(C^D, f) = \frac{\sum_{j=0}^{t-1} C^D(s^j)\left(I(s^{j+1}) - f(s^j)\right)}{\sum_{j=0}^{t-1} C^D(s^j)} = \frac{T_{t-1}\rho_{t-1} + g_t}{T_t}$.

Notice that these definitions match the first four assumptions of Lemma A1.

Let $\{\epsilon_t\}_{t=1}^{\infty}$ be a sequence of positive numbers satisfying (6) of Lemma A.1. That is,

$$\sum_{t=1}^{\infty} \epsilon_t < \infty.$$

Let λ be a probability distribution over $\mathcal{R}$ such that $\lambda(c) > 0$ for all $c \in \mathcal{R}$. By definition, $\sum\limits_{c \in \mathcal{R}} \lambda(c) = 1$.

To prove Proposition 1 it suffices to demonstrate the existence of a mixed forecasting scheme ζ such that $\nu = \lambda \times \zeta^*$ satisfies the fifth assumption of Lemma A1, i.e.,

$$E_{\nu}\left(\left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle\right) \leqslant\leq \epsilon_t.$$

Therefore, $\nu - a.e.$, $T_t(c, f) \to \infty$ implies that $\rho_t(c, f) \to \bar{0}$. By definition, all probabilities $\lambda(c)$ are strictly positive. Hence, $\zeta^* - a.e.$, for all $c \in \mathcal{R}$, $T_t(c, f) \to \infty$ implies that $\rho_t(c, f) \to \bar{0}$.

Fix $\varepsilon > 0$ and let be $Q(\varepsilon)$, be a finite subset of $\Delta(S)$ such that for every $q \in \Delta(S)$ exists $\hat{q} \in Q(\varepsilon)$ so that $\|q - \hat{q}\| < \frac{\varepsilon}{\sqrt{n}}$. The existence of such a subset follows from the compactness of $\Delta(S)$. Consider the following auxiliary zero-sum game:

1. The set of pure strategies for player 1 is $Q(\varepsilon)$.
2. The set of pure strategies for player 2 is S.
3. For any pair $(q,s) \in (Q(\varepsilon) \times S)$ the payoff from player 1 to player 2 is given by
$$G(q,s) \equiv \sum_{C^D \in \mathcal{R}} \lambda(C^D) \left\langle z_{C^D}(q),\ I(s) - q \right\rangle,$$
where $z_{C^D} : \Delta(S) \longrightarrow \Re^n$ satisfies $\|z_{C^D}(q)\|^2 \leq n$, for all $C^D \in \mathcal{R}$.

Lemma A2 There exists a mixed strategy for player 1 which limits the payoff to player 2 to no more than ε.

Proof: Given a mixed strategy for player 2, $p \in \Delta(S)$, $p = (p_1, ..., p_n)$, choose $\hat{q} \in Q(\varepsilon)$ for player 1 so that $\|p - \hat{q}\| \leqslant \frac{\varepsilon}{\sqrt{n}}$. Then,

$$\sum_{s \in S} p_s G(\hat{q}, s) = \sum_{s \in S} p_i \left(\sum_{c \in \mathcal{R}} \lambda(c) \left\langle\ z(c,\hat{q}),\ I(s) - \hat{q} \right\rangle \right) =$$

$$\sum_{c \in \mathcal{R}} \lambda(c) \left(\sum_{s \in S} p_s\ \left\langle z(c,\hat{q}),\ I(s) - \hat{q} \right\rangle \right) = \sum_{c \in \mathcal{R}} \lambda(c) \left(\left\langle z(c,\hat{q}),\ \sum_{s \in S} p_s\ (I(s) - \hat{q}) \right\rangle \right) =$$

$$\sum_{c \in \mathcal{R}} \lambda(c) \left(\ \langle z(c,\hat{q}),\ p - \hat{q} \rangle \right) \leqslant\leq \sum_{c \in \mathcal{R}} \lambda(c) \left(\|z(c,\hat{q})\|\ \|p - \hat{q}\| \right) \leqslant\leq \sum_{c \in \mathcal{R}} \lambda(c) \varepsilon = \varepsilon.$$

Hence, for any mixed strategy for player 2 there exists a pure strategy for player 1 that gives player 2 a payoff smaller than ε. By the Minimax Theorem, there exists a mixed strategy for player 1 which guarantees that the payoff of player 2 will not exceed ε, independently of player 2's action.

Q.E.D

An immediate consequence of Lemma A2 is that for any $\varepsilon > 0$ and any collection of functions $z_{C^D} : \Delta(S) \longrightarrow \Re^n$ such that $\|z_{C^D}(q)\|^2 \leq n$, for all $C^D \in \mathcal{R}$, there exists a measure $\mu \in \Delta(\Delta(S))$ such that for every $i \in S$

$$E_{\lambda \times \mu} \langle\ z_{C^D}(q),\ I(i) - q \rangle \leqslant\leq \varepsilon.$$

where $E_{\lambda \times \mu}$ is the expectation operator associated with $\lambda \times \mu$.

We now describe the mixed forecasting scheme ζ that will satisfy condition 5 of Lemma A1. Assume that ζ has been defined for all elements, x^{t-1} of $(\Delta(S) \times S)^{t-1}$. We construct $\zeta(x^t)$, for an arbitrary $x^t \in (\Delta(S) \times S)^t$.

Assume the calibration score at stage $t-1$, for each $C^D \in \mathcal{R}$, is ρ_{t-1}. Define the collection of functions $z_{C^D} : \Delta(S) \longrightarrow \Re^n$ as follows:

$$z_{C^D}(q) = \begin{cases} \frac{\rho_{t-1}(C^D,f)}{T_t(C^D,f)} & \text{if } C^D(s^{t-1}, q) = 1 \\ \bar{0} & \text{otherwise} \end{cases}$$

By Lemma A2 there exists a measure $\mu \in \Delta(\Delta(S))$, with a finite support Q_t, such that for any realization of $s_{t+1} \in S$

$$E_{\lambda \times} \mu \langle\ z_{C^D}(q),\ I(s_{t+1}) - q \rangle \leqslant\leq \varepsilon_{t+1}.$$

Let $\zeta(x^t) = \mu$. Observe that:

$$E_\nu \left(\left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle \right) = E_{\lambda \times \mu} \left(\left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle \right)$$

$$= \sum_{\{(C^D,q) \in \mathcal{R} \times Q_t\}} \lambda(C^D)\mu(q) \left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle$$

$$= \sum_{\{(C^D,q) : C^D(s^{t-1},q)=1\}} \lambda(C^D)\mu(q) \left\langle \frac{\rho_{t-1}}{T_t}, g_t \right\rangle$$

$$= \sum_{\{(C^D,q) : C^D(s^{t-1},q)=1\}} \lambda(C^D)\mu(q) \left\langle z_{C^D}(q), I(s_{t+1}) - q \right\rangle$$

$$= \sum_{\{(C^D,q) \in \mathcal{R} \times Q_t\}} \lambda(C^D)\mu(q) \left\langle z_{C^D}(q), I(s_{t+1}) - q \right\rangle$$

$$= E_{\lambda \text{x} \mu} \langle\ z_{C^D}(q),\ I(s_{t+1}) - q \rangle \leqslant\leq \varepsilon_{t+1}.$$

And so assumption (5) of lemma $A1$ is satisfied.

The proof of the case where λ is an arbitrary probability distribution over $\mathcal{R}$ is identical to the proof presented here. It suffices to replace the summations by appropriately defined integrals. Therefore, we omit this proof.

References

[1] Blackwell, D. (1956) "A Vector Valued Analog of the Mini-Max Theorem," *Pacific Journal of Mathematics* **6**, 1–8.

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

[3] Foster, D.P. and R.V. Vohra (1997) "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, 21, 40-55.

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

[5] Fudenberg, D. and D. Levine (1995) "Consistency and Cautious Fictitious Play," *Journal of Economics Dynamics and Control,* 19, 1065-1089

[6] Fudenberg, D. and D. Levine (1995) "An Easier Way to Calibrate," mimeo, UCLA.

[7] Fudenberg, D. and D. Levine (1996) "Conditional Universal Consistency," mimeo, UCLA.

[8] Grandmont, J.M. (1982) "Temporal General Equilibrium Theory." In: Arrow, K.J., Intriligator, M.D. (eds.) Handbook of Mathematical Economics, 2, 879-922. Amsterdam: North-Holland.

[9] Hart, S. and A. Mas-Colell (1996) "A Simple Adaptive Procedure Leading to Correlated Equilibrium," mimeo, Hebrew University.

[10] Kalai, E., E. Lehrer and R. Smorodinsky (1995) "Calibrated Forecasting and Merging," to appear in *Games and Economic Behavior*.

[11] Kurz, M. (1994) "On Rational Belief Equilibria," *Economic Theory* 4, 859-876.

[12] Lehrer, E. (1997) "Any Inspection Rule is Manipulable," Manuscript, Tel-Aviv University.

[13] Oakes, D. (1985) "Self Calibrated Priors Do Not Exist," *Journal of the American Statistical Association*, **80**, 339.

[14] Sandroni, A. and R. Smorodinsky (1998) "Dynamic Foundations of Equilibrium in Games - A Second Look," mimeo, Technion.

[15] Shiryaev, A.N. (1984), *Probability*, Springer-Verlag, New-York.