Calibration with Many Checking Rules

Alvaro Sandroni †  Rann Smorodinsky ‡
Rakesh V. Vohra §

May 13, 2002

Abstract

Each period an outcome (out of finitely many possibilities) is observed. For simplicity assume two possible outcomes, a and b. Each period, a forecaster announces the probability of a occurring next period based on the past.

Consider an arbitrary subsequence of periods (e.g., odd periods, even periods, all periods in which b is observed, etc). Given an integer n, divide any such subsequence into associated sub-subsequences in which the forecast for a is between \([i/n, i+1/n]\), \(i \in \{0, 1, ..., n\}\).

We compare the forecasts and the outcomes (realized next period) separately in each of these sub-subsequences. Given any countable partition of \([0, 1]\) and any countable collection of subsequences, we construct a forecasting scheme such that for all infinite strings of data, the long run average forecast for a matches the long run frequency of realized a’s.

*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 Economics, University of Rochester, Rochester NY 14627 (alsn@troi.cc.rochester.edu) and 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. (ramn@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 litteram,
Book II.

1 Introduction

In any good model of human behavior, it is posited that beliefs will be
revised when they are contradicted by the data. The difficulty is to define
the appropriate sense in which the beliefs and the data should eventually
comply. Currently, the dominant assumption in economics is that agents’
beliefs coincide with the true probabilities. However, if agents’ beliefs do not
coincide with the truth will agents be able to recognize it? Assume just two
possible outcomes, a and b. If the probability of a is fixed throughout time
then the empirical frequency of a will reveal this probability. On the other
hand, assume that the assumption of a fixed probability is incorrect. Would
agents be able to recognize it and drop the assumption of a fixed probability?

This question motivates us to look for a general definition of when the
data contradicts the forecasts and when it does not (by general we mean a
definition that applies for any sequence of data and probabilistic forecasts).
For simplicity, consider the sequence of forecasts “a will occur with proba-
bility p in all periods.” These forecasts would be contradicted by the data if
the frequency of a is not close to p. However, assume the frequency of a is
p. The table below shows that the data could still contradict the forecast.

<table>
<thead>
<tr>
<th>outcome</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>forecasts</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Assuming the pattern repeats indefinitely, the long run frequency of a is
0.5. The forecast of a is also 0.5. However, on the odd numbered periods,
the frequency of a is 1 but the forecast is 0.5. Thus, the forecast tracks the
long run frequency of a’s, but misses the alternating pattern of a’s and b’s.

We demand more by breaking the sequence up into two subsequences; one
corresponding to even periods and the other to odd periods, and require the
forecast to match the frequency on each subsequence. The next table shows
that this hurdle can be satisfied and yet the data contradicts the forecast.

\[
\begin{array}{cccccccccc}
\text{outcome} & a & a & b & b & a & a & b & b & a & a \\
\text{forecasts} & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
\end{array}
\]

Assuming the pattern of \(aa\) followed by \(bb\) repeats indefinitely, the long run frequency of \(a\) is 0.5 as anticipated by the forecast. In the odd periods, the long run frequency of \(a\) is also, as anticipated, 0.5. In the even periods, it is also 0.5. However, if the probability of \(a\) in every period were, in fact, 0.5 then we would expect that the frequency of \(a\), after \(aa\) was observed, to be 0.5. In the data, this frequency is zero. Analogously, the frequency of \(a\) in the period \(4n+1\), \(n\) natural number, should be 0.5 when it is one.

The table above suggests that dividing the sequence up into just two subsequences is not enough. So, how many subsequences would suffice? And which ones should they be? To answer these questions, we formalize the notion of dividing up the entire sequence of observations into subsequences. Imagine a rule that, at the end of each period, decides whether or not to mark the period (as a function of the past). The marked periods define a subsequence on which the forecasts (made in those periods) could be compared to the outcomes realized next period. One rule might be to mark every even numbered period. The forecasts made in the even periods will be compared with the outcomes realized next period. Another rule would be to mark the period if the current outcome is \(a\). The forecasts made in the periods that \(a\) occurred will be compared with the outcomes realized next period.

A rule that decides which periods to mark (as a function of past and current outcomes) is called an \textit{outcome-based checking rule}. Formally, an outcome-based checking rule is a function from finite sequences of outcomes to \(\{0,1\}\). We say that the rule is active when it assumes the value 1 for that period. The marked periods are those in which the rule is active. An outcome-based checking rule could be active when the last three observation were \(aba\), when the period is a prime-number, etc.

Outcome-based checking rules mark a period based on past and current outcomes only. However, if forecasts change then we may want a checking rule that marks a period as a function of the forecasts as well. Fix an outcome-based checking rule and an interval \(D\) of possible forecasts. An associated forecast-based checking rule will mark those periods marked by the outcome-based checking rule when the forecast lies in \(D\). That is, a forecast-based checking rule is active when the outcome-based checking rule is active \textit{and} the forecast is within some interval (these intervals form a partition of \([0,1]\)).
For example, consider the outcome-based checking rule that is active in the even periods. Consider the partition \([0, 0.5)\) and \([0.5, 1]\). A forecast-based checking rule (associated with this outcome-based checking rule) is active in the even periods when the forecast for \(a\) is less than 0.5. Another forecast-based checking rule is active in the even periods when the forecast for \(a\) is greater than 0.5. For each forecast-based checking rule, there is an associated subsequence of active periods. The forecasts will be compared to the data separately in each of these subsequences.

Given a collection \(\mathcal{C}\) of outcome-based checking rules and a partition of \([0, 1]\), we say that a sequence of forecasts is calibrated with respect to the observed data if the average forecasts match empirical frequencies, in the subsequence specified by the forecast-based checking rule associated with the outcome-based checking rules in \(\mathcal{C}\). Informally, a sequence of forecasts is calibrated if, in the subsequences specified by \(\mathcal{C}\), the frequency of \(a\) is \(p\) in the sub-subsequences in which the forecast is \(p\).

The examples above show that forecasts matching empirical frequencies for finitely many checking rules may fail to capture relatively simple patterns. However, consider a countable collection of outcome-based checking rules that include all functions (mapping finite sequences of outcomes to \(\{0, 1\}\)) implementable by a recursive algorithmic. Consider a countable partition of \([0, 1]\). This collection of forecast-based checking rules is also countable. If the forecasts match the empirical frequencies for all these forecast-based checking rules then no comparison between frequencies and the forecasts, that is implementable by a Turing-Machine, would reject the hypothesis that the forecasts are correct.

The main result in this paper shows that, given any countable collection of outcome-based checking rules and countable partition of the entire interval, there is a forecasting scheme that generates sequences of calibrated forecasts on every possible infinite string of data. So, if a forecaster uses this forecasting scheme then after some point in the future, when he looks backwards, he will always see that the time average of the forecasts are close to the empirical frequencies. In this sense, he will not see a contradiction between the forecasts and the data.

### 1.1 Related Literature

The idea that calibration is a desirable property of probabilistic forecasts is due to Dawid (1982). He shows that the posterior beliefs of a coherent
Bayesian will become calibrated with probability one under the Bayesian forecaster’s own prior. Subsequently Oakes (1985) showed that no deterministic forecasting scheme could be guaranteed to calibrate the forecast-based checking rule associated with the always active outcome-based checking rule. The existence of a randomized forecasting scheme that calibrates the forecast-based checking rule associated with the always active outcome-based checking rule was established by Foster and Vohra (1998). This result was first generalized by Fudenberg and Levine (1999a) who show the existence of forecasting schemes that calibrate certain classes of checking rules. Foster and Vohra’s (1998) result also inspired alternative proofs of the same result by Fudenberg and Levine (1999b) and Hart and Mas-Collel (2001) which use the minimax theorem and the Blackwell approachability theorem, respectively. These proofs are conceptually simpler and inspired the proof in this paper. Foster and Vohra (1999) also show a potential use of Blackwell’s approachability theorem in a variety of settings.

Variations and strengthenings of Dawid’s (1982) original notion of calibration were introduced in Dawid (1985) which discusses the relationship between calibration and merging. This topic is also central in Kalai, Lehrer and Smorodinsky (1999). Dawid and Vovk (1999) show a connection between calibration and gambling strategies (see also Shafer and Vovk (2001)).

The existence of a forecasting scheme that simultaneously calibrates countably many outcome-based checking rules (but not the associated forecast-based checking rules) was first established by Lehrer (2001). The main result of this paper generalizes the result of Lehrer (2001) to forecast-based checking rules. Thus, the results of Lehrer (2001) and Foster and Vohra (1998) are special cases of our result. Further, we strengthen the result of Foster and Vohra (1998) to show that the difference between the forecasts and the long-run frequencies converge to zero rather than simply being sufficiently small.

2 Definitions and Result

Let \( N \) be the set of natural numbers. Let \( N_+ \) be \( N \cup \{0\} \). We denote by \( S \equiv \{1, \ldots, n\} \) the state space and by \( S^0 \) the empty set. We call an element of \( S \) an outcome. Let \( S^t, t \in N \cup \{\infty\} \), be the \( t \)-Cartesian product of \( S \), and let \( \bar{S} \equiv \bigcup_{t=0}^{\infty} S^t \) be the set of all finite sequences of outcomes. Given an infinite sequence of outcomes \( s \in S^\infty \), 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 \(s\). We call an infinite sequence of outcomes \(s \in S^\infty\) a path.

Let \(\Delta(S)\) be the set of probability measures over \(S\) and let \(\Delta(S)^0\) be the empty set. We call an element of \(\Delta(S)\) a forecast. We denote by \(\Delta(S)^t\), \(t \in N \cup \{\infty\}\), the \(t\)-Cartesian product of \(\Delta(S)\). Given an infinite sequence of forecasts \(f \in \Delta(S)^\infty\), \(f = (f_0, f_1, \ldots, f_t, \ldots)\), we denote by \(f_t \in \Delta(S)\) the \((t+1)\)th element of \(f\), and by \(f^{t-1} = (f_0, f_1, \ldots, f_{t-1}) \in \Delta(S)^t\) the prefix of length \(t\) of \(f\).

A forecast determines the probability that an outcome will be realized next period. For example, at period zero a forecast \(f_0\) is made. This forecast determine probabilities for the outcomes that will only be observed at period 1. A \(t\)-sequence of forecasts \(f^{t-1} \in \Delta(S)^t\), \(f^{t-1} = (f_0, f_1, \ldots, f_{t-1})\), are the forecasts \(f_j\) made at periods \(j \leq t - 1\), starting from the forecast made at period zero. A \(t\)-sequence of outcomes \(s^t \in S^t\), \(s^t = (s_1, s_2, \ldots, s_t)\) are the outcomes \(s_j\) realized at periods \(j \leq t\), starting from the outcome realized at period one. We call the pair \(x^t = (s^t, f^{t-1}) \in S^t \times \Delta(S)^t\), a \(t\)-history. An infinite history is called a history.

Let \(X^t = S^t \times \Delta(S)^t\) be the set of \(t\)-histories. Let \(H \equiv \bigcup_{t=0}^\infty X^t\) be the set of finite histories. Let \(\Delta(\Delta(S))\) be the set of probability distributions over \(\Delta(S)\).

**Definition 1** A forecasting scheme is a function \(\zeta : H \rightarrow \Delta(\Delta(S))\).

At the end of each stage \(t \in N\), a \(t\)-history is observed. Based on this observation, the forecaster must decide which forecast \(f_t \in \Delta(S)\) to make at period \(t\). We allow the forecaster to randomize. So, \(f_t \in \Delta(S)\) can be selected (possibly) at random, using a probability distribution \(\mu_t \in \Delta(\Delta(S))\). The realization of \(\mu_t\) is, of course, observed at period \(t\).

If the forecaster chooses the forecasts deterministically, i.e., if for any \(t\)-history \(x^t \in H\), \(\zeta(x^t)\) assigns probability one to a probability measure on \(\Delta(S)\), then we say that \(\zeta\) is a pure forecasting scheme. We now define how the forecasts will be compared to the data.

**Definition 2** An outcome-based checking rule is a function \(C : \hat{S} \rightarrow \{0, 1\}\).

An outcome-based checking rule is an arbitrary function mapping finite sequences of outcomes to \(\{0, 1\}\). If \(C(s^t) = 1\) then we say that \(C\) is active at \(s^t\). Analogously, if \(C(s^t) = 0\) then we say that \(C\) is inactive at \(s^t\). An outcome-based checking rule could be always active, active in odd periods, active in period \(t\) whenever term \(t\) of the binary expansion of \(\pi\) is 1, etc.
Definition 3 Given a set $D \in \Delta(S)$ and an outcome-based checking rule $C$, a forecast-based checking rule is a function $C^D : H \rightarrow \{0, 1\}$ such that
\[
C^D(s^t, f^{t-1}) = 1 \quad \text{if } C(s^t) = 1 \text{ and } f_{t-1} \in D; \\
C^D(s^t, f^{t-1}) = 0 \quad \text{otherwise};
\]
where $f^{t-1} = (f_0, f_1, \ldots, f_{t-1})$.

If $C^D(x^t) = 1$ then we say that $C^D$ is active at $x^t = (s^t, f^{t-1})$. Analogously, if $C^D(x^t) = 0$ then we say that $C^D$ is inactive at $x^t$. So, given a subset $D \subseteq \Delta(S)$, and an outcome-based checking rule, a forecast-based checking rule is a function $C^D$ which is active (i.e., equal to one) if and only if the outcome-based checking rule $C$ is active and the forecast made last period belongs to $D$. For example, a forecast-based checking rule could be active when the forecast for outcome 1 is greater than 0.5, active in the odd periods in which the forecast for outcome $n$ is between 0.1 and 0.4, active when the last four outcomes were identical and the forecast for any outcome is smaller than $2/n$, etc. Clearly, the outcome-based checking rule $C$ is identical to the forecast-based checking rule $C^\Delta(S)$.

Definition 4 Given a forecast-based checking rule $C^D$ and a $T$-history $x^T = (s^T, f^{T-1})$, the calibration score at time $T$ is the $n$-dimensional vector
\[
\rho_T(C^D, x^T) = \frac{\sum_{t=1}^{T} C^D(x^t)(I(s^t) - f_{t-1})}{\sum_{t=1}^{T} C^D(x^t)},
\]
where $f^{T-1} = (f_1, \ldots, f_{T-1})$, $s^T = (s_1, \ldots, s_T)$ and $I(s^t)$ is the $n$-vector with 1 in the $s^t_{i^*}$-coordinate and zero elsewhere.

So, the calibration score is the difference between empirical frequencies and average forecasts, in the periods that the forecast-based checking rule was active.

Definition 5 An infinite sequence of forecasts $f \in \Delta(S)^\infty$ calibrate a forecast-based checking rule $C^D$ on a path $s \in S^\infty$ if $\sum_{t=1}^{\infty} C^D(x^t)$ is finite or if
\[
\lim_{T \to \infty} \rho_T(C^D, x^T) = 0,
\]
where $x^T = (s^T, f^{T-1})$, 0 is the $n$-vector $(0, \ldots, 0)$, and the equality is coordinate-wise.
That is, the realized forecasts calibrate a forecast-based checking rule on a path if the calibration score converges to zero on this sequence (whenever the forecast-based checking rule is active infinitely often on this sequence).

Given a path \( s \in S^\infty \) and a forecasting scheme \( \zeta \), let the function

\[
\tilde{\zeta}_s : \bigcup_{t=0}^{\infty} \Delta(S)^t \rightarrow \Delta(\Delta(S))
\]

be given by

\[
\tilde{\zeta}_s(f^{t-1}) = \zeta(s^t, f^t).
\]

By the Kolmogorov’s Extension Theorem, a function \( \tilde{\zeta}_s \) determine a unique probability measure \( \zeta_s^* \) on \( \Delta(S)^\infty \), the space of infinite sequences of forecasts. Since \( \tilde{\zeta}_s \) is determined by \( s \) and \( \zeta \), it follows that a path \( s \in S^\infty \) and a forecasting scheme \( \zeta \) determine a unique probability measure \( \zeta_s^* \) on \( \Delta(S)^\infty \).

**Definition 6** A forecasting scheme, \( \zeta \), calibrates a forecast-based checking rule \( C_D \) on \( s \in S^\infty \) if \( \zeta_s^* \)-almost every infinite sequence of forecasts \( f \) calibrates \( C_D \) on \( s \).

That is, a forecasting scheme calibrates a forecast-based checking rule on a path if, the calibration scores of almost every infinite sequence of forecasts converge to zero on this path (whenever the forecast-based checking rule is active infinitely often on this path).

**Definition 7** A forecasting scheme, \( \zeta \), calibrates a forecast-based checking rule \( C_D \) if \( \zeta \) calibrates \( C_D \) on every path \( s \in S^\infty \).

So, a forecasting scheme calibrates a forecast-based checking rule if on any given path, the calibration scores of almost every infinite sequence of forecasts converge to zero. We now state the main result of this paper.

**Proposition 1** Let \( D = \{D^1, D^2, \ldots \} \) be an arbitrary countable collection of subsets of \( \Delta(S) \). Let \( C = \{C_1, C_2, \ldots \} \) be an arbitrary countable collection of outcome-based checking rules. Then, there exists a forecasting scheme that simultaneously calibrates all forecast-based checking rules \( C^{D^k}_j \), \( C_j \in C \), \( D^k \in D \).
Proof - See section 4.

Proposition 1 shows that there exists a forecasting scheme such that given any path, almost every infinite sequence of realized forecasts (under the distribution induced by the forecasting scheme) will match the empirical frequencies observed in the data.

Foster and Vohra’s (1998) main result is the special case of proposition 1 where \( C \) contains only the always active outcome-based checking rule and \( D \) is a finite partition of \( \Delta(S) \). Foster and Vohra (1998) are interested in checking rules which are active 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 (1998) 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.

It follows immediately from proposition 1 that there exists a forecasting scheme that simultaneously calibrates all outcome-based checking rules in \( C \). Simply consider the trivial partition \( D = \{\Delta(S)\} \). Lehrer (2001) shows that there exists a pure forecasting scheme that calibrates countable outcome-based checking rules. This does not follow immediately from proposition 1 (although a separate proof of Lehrer’s (2001) result, based on proposition 1, can be derived). Example 1 shows that the general result in proposition 1 cannot be obtained by a pure forecasting scheme.

Example 1 Let \( S = \{a, b\} \). Let \( C \) be the always active outcome-based checking rule. Let \( D_1 = [0, 0.5) \) and \( D_2 = [0.5, 1] \) be a partition of \( \Delta(S) \). There is no pure forecasting scheme that calibrates \( C^{D_1} \) and \( C^{D_2} \) on all paths (see Oakes (1985)).

Consider an arbitrary pure forecasting scheme. Consider the path in which outcome \( a \) occurs if and only if the forecast of \( a \) was strictly less than 0.5. In the periods that \( a \) was forecast with probability strictly less than 0.5, the empirical frequency of \( a \) is 1. In the periods that \( a \) was forecast with probability greater or equal than 0.5, the empirical frequency of \( a \) is 0.

There is no contradiction between example 1 and proposition 1. To see this, note that the forecasting rule in proposition 1 is not necessarily pure. The forecaster may choose the forecasts at random. Moreover, the realized outcome is not a function of the realized forecast (i.e. nature is oblivious to
the realized forecasts). Given an arbitrary path, the realized forecasts will match the frequencies with probability one under the distribution induced by the forecasting scheme, but given any infinite sequence of forecasts there are paths in which this match does not occur.

3 Discussion

In this section, we argue that, under some distributional assumptions, calibrated forecasts must eventually approach the given probabilities of outcomes. We also comment on calibration as a test of knowledge and on the connection between calibration and the foundations of Nash-equilibrium.

3.1 Calibration and Statistical Inference

It is usual to assume that the outcomes being forecasted are generated by an underlying stochastic process. The approach in this paper avoids such assumption. It is nevertheless interesting to examine the properties of calibrated forecasts under some of the usual distributional assumptions on the outcomes. A full analysis of this topic is beyond the scope of this paper. We confine ourselves to presenting some plausible conjectures. For a formal analysis of related issues, we refer the reader to Kalai, Lehrer and Smorodinsky (1999).

Assume, for example, that there are two possible outcomes, $a$ and $b$. 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 outcome-based checking rule (as in Foster and Vohra (1998)) must eventually predict that the probability of $a$ is close to 0.5. Otherwise the calibration score of the forecast-based checking rules which are active 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 (2001) forecasting scheme (which calibrate countably many outcome-based 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 (2001)
scheme in the odd periods and in the even periods will eventually become close to the truth, otherwise the calibration score for the two outcome-based checking rules active in those respective periods will not eventually be close to zero. However, the forecasts in Foster and Vohra’s (1998) scheme need not become close to the true probabilities.

Suppose the true stochastic process is a time-independent Markovian process of one step. Then, neither the forecasts in Foster and Vohra’s (1998) scheme nor the ones in Lehrer’s (2001) 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 the forecasting scheme calibrates the forecast-based checking rules associated with two outcome-based checking rule. One of them is active when the current outcome is $a$ and the other 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 either $a$ or $b$ is realized).

If the forecasting scheme only calibrates finitely many outcome-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 outcome-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.

### 3.2 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 outcome contradicts the forecast.

The task of designing a test that differentiates the knowledge of forecast-
ers is more difficult if probabilistic forecasts are permitted. If we assume that the outcomes 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 outcomes 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). A more elaborate test compares the time average of the forecasts of $a$ and the empirical frequencies of $a$ in the periods where $a$ was forecast with probability $p$, in the even periods where $a$ was forecast with probability $p$, when the last four outcomes were identical and $a$ was forecast with probability $p$, etc.

Our main result implies that both forecasters who know the stochastic process and those whose knowledge was obtained from the data will be able to drive the calibration scores to zero.

3.3 Foundations of Nash equilibrium

In a Nash-equilibrium players’ beliefs are correct, i.e. players behave as if they knew the (perhaps mixed) strategy the opponents are using. However, the play can be sufficiently non-stationary so that even if players have observed the play for a long period of time they may not be able to deduce (from the past play) the relevant future probabilities. On the other hand, the results in this paper shows that after many interactions, players’ beliefs can become calibrated no matter how the play turns out to be. Foster and Vohra (1997) and Sandroni and Smorodinsky (2001) show if players optimize myopically and their beliefs are calibrated then the empirical frequencies of the play path (if they exist) will coincide with the empirical frequencies of a Correlated-equilibrium play. In addition, if the stage game has a unique correlated equilibrium then empirical frequencies of the play path do exist and coincide with the empirical frequencies of a Nash equilibrium play.
4 Proof of Main Result

4.1 Preliminary results

Lemma A.0 Let \( z_i \in \{-1, 1\} \) be a sequence of random variables and \( X_n \equiv \frac{\sum_{i=1}^{n} z_i}{n} \). If \( \sum_{j=1}^{n} \frac{E(X_j^2)}{j} \) converges than \( X_n \to 0 \) almost surely as \( n \to \infty \).

Proof: This argument is taken from Lehrer (2002). Fix \( \varepsilon > 0 \). Let \( A_j \) denote the event that \((X_j z_{j+1})^2 > \varepsilon^2 \) and \( \chi(A_j) \) the characteristic function of this event. By Chebyshev’s inequality it follows that \( P(A_j) \leq \frac{E(X_j^2)}{\varepsilon^2} \).

Therefore, \( \sum_{j=1}^{n} \frac{P(A_j)}{j} \) converges.

For \( m < k \), let \( S(m, k) \equiv \sum_{j=m+1}^{k} \chi(A_j) \) and \( S(k) \equiv S(0, k) \). The proof of lemma A.0 follows from the Borel-Cantelli Lemma and facts 1 and 2 below.

Fact 1. If \( X_j^2 > 9 \varepsilon \) infinitely often then \( S((1+\varepsilon)^j, (1+\varepsilon)^{j+1}) > \varepsilon \) infinitely often.

Fact 2.

\[
\sum_{k} P(S((1+\varepsilon)^k, (1+\varepsilon)^{k+1}) > \varepsilon) < \infty.
\]

To demonstrate fact 1, first assume that \( X_j^2 > 9 \varepsilon \). Observe that

\[
X_j^2 = (\frac{j-1}{j} X_{j-1})^2 + 2(\frac{j-1}{j})X_{j-1}(\frac{z_j}{j}) + (\frac{z_j}{j})^2.
\]

Repeated application of this identity yields:

\[
X_j^2 = \frac{X_1^2}{j^2} + (2/j) \sum_{i=2}^{j} (\frac{i-1}{j})X_{i-1}z_i + \sum_{i=2}^{j} (\frac{z_i}{j})^2 \leq (2/j) + (2/j) \sum_{i=2}^{j} (\frac{i-1}{j})X_{i-1}z_i.
\]

So, for \( j \) sufficiently large, \((2/j) \sum_{i=2}^{j} (\frac{i-1}{j})X_{i-1}z_i \) is greater than \( 8 \varepsilon \). Since

\[
(2/j) \sum_{i=2}^{j} (\frac{i-1}{j})X_{i-1}z_i = j^{-1} \sum_{i=2}^{j} (2/j) \sum_{k=i}^{j} X_{k-1}z_k
\]

it follows that there exists some \( i \leq j \) such that \((2/j) \sum_{k=i}^{j} X_{k-1}z_k \geq 8 \varepsilon \). So,

\[
S(j) + \varepsilon \geq (1/j) \sum_{k=1}^{i-1} \chi(A_k) + (1/j) \sum_{k=i}^{j} (\varepsilon + \chi(A_k)) \geq
\]

13
\[
(1/j) \sum_{k=1}^{i-1} \chi(A_k) + (1/j) \sum_{k=i}^{j} X_{k-1} z_k \geq 4\epsilon.
\]

So, if \(X_j^2 > 9\epsilon \) infinitely often then \(S(j) > 3\epsilon \) infinitely often. This implies that \(S((1 + \epsilon)^j, (1 + \epsilon)^{j+1}) > \epsilon \) infinitely often because \(S(j)\) is an average of terms of the form \(S((1 + \epsilon)^k, (1 + \epsilon)^{k+1})\) for \(k \leq j\) as well as \(S((1 + \epsilon)^k, j)\).

We now show fact 2. By Markov’s inequality we have

\[
P(S(m, k) > \epsilon) \leq \frac{1}{\epsilon} \frac{\sum_{i=m+1}^{k} P(A_i)}{k - m}.
\]

Thus,

\[
\sum_{k} P(S((1 + \epsilon)^k, (1 + \epsilon)^{k+1}) > \epsilon) \leq \sum_{k} \frac{1}{\epsilon(1 + \epsilon)^{k+1}} \leq \frac{1 + \epsilon}{\epsilon^2} \sum_{j} \frac{P(A_j)}{j} < \infty.
\]

The next result is a minor generalization of Proposition 1 of Lehrer (2002) which generalizes Blackwell’s (1956) celebrated Approachability Theorem (see also Lehrer (2001)). Let \(\langle \cdot, \cdot \rangle\) be the inner product in \(\mathbb{R}^n\), \(\langle a, b \rangle = \sum_{i=1}^{n} a_i b_i\), \(a \in \mathbb{R}^n, b \in \mathbb{R}^n\). Let \(\| \cdot \|\) be the norm in \(\mathbb{R}^n\), \(\|a\| = \sqrt{\langle a, a \rangle}, a \in \mathbb{R}^n\).

**Lemma A1** Let \((X, B, \nu)\) be a probability measure space. Let \(\{\bar{T}_t\}_{t=0}^{\infty}\) be a sequence of random variables taking values in \(\mathbb{N}^+\), \(\langle \bar{g}_t, a \rangle \leq \sum_{i=1}^{n} a_i b_i\), \(a \in \mathbb{R}^n, b \in \mathbb{R}^n\). Let \(\bar{\rho}_t\) be a sequence of non-negative numbers such that \(\sum_{t=1}^{\infty} \bar{\epsilon}_t < \infty\). Assume:

1. \(\bar{T}_t\) is non-decreasing, \(t \in \mathbb{N}; \bar{T}_t - \bar{T}_{t-1} \leq 1;\) and \(\bar{T}_0 = 0;\)
2. \(\bar{g}_t \in [-1, 1]^n;\)
3. \(\bar{T}_t - \bar{T}_{t-1} = 0\) implies \(\bar{g}_t = 0;\)
4. \(\bar{\rho}_t = \frac{\bar{T}_{t-1+\bar{g}_{t-1}}}{\bar{T}_t},\) where the ratio is taken for each coordinate of \(\mathbb{R}^n;\)
5. $E_{\nu} \left( \left( \frac{\bar{\rho}_t}{T_{t+1}}, \bar{g}_{t+1} \right) \right) \leq \varepsilon_{t+1}$, where $E_{\nu}$ is the expectation operator associated with $\nu$.

Then, $\nu - a.e., \bar{T}_t \to \infty$ implies that $\bar{\rho}_t \to 0$.

**Proof:** The case $n = 1$ was demonstrated by Lehrer (2002). We reproduce his proof here. We prove that $\sum_{i=1}^{\infty} E_{\nu}(\bar{\rho}_{t}^{2})$ is convergent.

Denote by $A^i$ the event that $\bar{T}_i - \bar{T}_{i-1} = 1$ and $\chi_i$ the characteristic function of this event. Set

$$x_t = \chi_t \frac{\bar{g}_t - \bar{\rho}_{t-1}}{T_t}.$$ 

Then

$$\bar{\rho}_t = \bar{\rho}_{t-1} + x_t.$$ 

Hence

$$\bar{\rho}_t^2 = \bar{\rho}_{t-1}^2 + 2\bar{\rho}_{t-1} \times \chi_t \frac{\bar{g}_t}{T_t} - 2\bar{\rho}_{t-1} \times \chi_i \frac{\bar{\rho}_{t-1}}{T_t} + x_t^2.$$ 

Hypothesis (3) of the lemma implies that $\chi_t \frac{\bar{g}_t}{T_t} = \bar{\rho}_t$. Hypothesis (5) of the lemma implies that

$$E_{\nu}(\bar{\rho}_t^2) \leq E_{\nu}(\bar{\rho}_{t-1}^2) + 2\varepsilon_t - 2E_{\nu} \left( \bar{\rho}_{t-1} \times \chi_i \frac{\bar{\rho}_{t-1}}{T_t} \right) + E_{\nu}(x_t^2).$$ 

Applying this inequality recursively yields:

$$E_{\nu}(\bar{\rho}_t^2) \leq E_{\nu}(\bar{\rho}_1^2) + 2 \sum_{i=2}^{t} \varepsilon_i - \sum_{i=2}^{t} 2E_{\nu} \left( \bar{\rho}_{i-1} \times \chi_i \frac{\bar{\rho}_{i-1}}{T_i} \right) + \sum_{i=2}^{t} E_{\nu}(x_i^2).$$ 

Let $r_i = \sum_{k=1}^{i} \chi_k$. Then each $|x_i|$ is at most $2/r_i$. By adding the non-zero terms of the series $\{x_i^2\}$ we obtain

$$\sum_{i=2}^{\infty} x_i^2 \leq \sum_{j=2}^{\infty} 2/j^2 \implies \sum_{i=2}^{\infty} E_{\nu}(x_i^2) < \sum_{j=2}^{\infty} 2/j^2 < \infty.$$ 

By assumption $\sum_{i=2}^{\infty} \varepsilon_i < \infty$. It follows that

$$0 \leq \sum_{i=2}^{t} 2E_{\nu}(\bar{\rho}_{i-1} \times \chi_i \frac{\bar{\rho}_{i-1}}{T_i}) \leq -E_{\nu}(\bar{\rho}_1^2) + E_{\nu}(\bar{\rho}_1^2) + 2 \sum_{i=2}^{t} \varepsilon_i + \sum_{i=2}^{t} E_{\nu}(x_i^2) < \infty.$$ 

15
Let $f(t)$ be $\min\{k : T_k \geq t\}$, $z_t \equiv g_{f(t)}$ and $X_t \equiv \sum_{i=1}^{f(t)} z_i$. Hypothesis (3) and (4) implies that $\{X_t\}$ is a subsequence of the $\{\bar{p}_j\}$'s. More precisely, for each $X_i$ there is a unique $j \geq i$ such that $\chi_j = 1$, $\bar{T}_j = i \Rightarrow i + 1 \geq \bar{T}_{j+1}$ and $X_i = \bar{p}_j$. Thus,

$$\sum_{i=1}^{\infty} \frac{X_i^2}{i+1} \leq \sum_{j=1}^{\infty} \bar{p}_j \times \chi_j \frac{\bar{p}_j}{\bar{T}_{j+1}} \Rightarrow \sum_{i=1}^{\infty} \frac{E_\nu(X_i^2)}{i+1} \leq \sum_{j=1}^{\infty} E_\nu(\bar{p}_j \times \chi_j \frac{\bar{p}_j}{\bar{T}_{j+1}}).$$

Hence the series $\sum_{t=1}^{\infty} E_\nu(X_t^2)$ is convergent. Therefore, by Lemma A.0, $X_n \to 0 \nu$-almost surely. It is straightforward to show that if $\bar{T}_t \to \infty$ then $X_n \to 0 \Rightarrow \bar{p}_n \to 0$. Hence, $\nu$-a.e., $\bar{T}_t \to \infty$ implies that $\bar{p}_t \to 0$.

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 $\bar{\nu}(\cdot, j) = \frac{\nu(j)}{n}$. For any random variable $Y$ on $X$ taking values in $\mathbb{R}^n$, let $\tilde{Y}$ be a real valued random variable on $X \times \{1, \ldots, n\}$ defined by $\tilde{Y}(x, j) = (Y(x))_j$, where $x \in X$ and $(Y(x))_j$ is the $j$th coordinate of $Y(x)$.

By definition,

$$E_\nu(\bar{g}_t \frac{\bar{p}_{t-1}}{\bar{T}_t}) = \frac{1}{n} E_\nu \left( \left( \bar{g}_t, \frac{\bar{p}_{t-1}}{\bar{T}_t} \right) \right) \Rightarrow E_\nu(\bar{g}_t \frac{\bar{p}_{t-1}}{\bar{T}_t}) \leq \frac{\varepsilon_t}{n}.$$

Therefore, we are back to the case $n = 1$ replacing $\bar{g}_t$ with $\bar{g}_t$ and $\bar{p}_t$ with $\bar{p}_t$. Hence,

$$\bar{p}_t \to 0 \tilde{\nu} - \text{a.e.} \Rightarrow \bar{p}_t \to 0 \nu - \text{a.e.} \quad \blacksquare$$

Let $\mathcal{C}$ be a countable collection of outcome-based checking rules. Let $\mathcal{D}$ be a countable collection of subsets of $\Delta(S)$. Let $\mathcal{R}$ be $\mathcal{C} \times \mathcal{D}$. An element $(C, D)$ of $\mathcal{R}$ can be identified with the forecast-based checking rule $C^D$. Let $\lambda$ be a probability measure over $\mathcal{R}$ such that

$$\lambda(C^D) > 0$$

for every $(C, D) \in \mathcal{R}$. The existence of $\lambda$ follows from the fact that $\mathcal{R}$ is countable.

Given $\varepsilon > 0$, let $Q(\varepsilon)$ be a finite subset of $\Delta(S)$ such that for every $q \in \Delta(S)$ exists $\hat{q} \in Q(\varepsilon)$ with the property that $\|q - \hat{q}\| < \frac{\varepsilon}{\sqrt{n}}$. The existence of such subset follows from the compactness of $\Delta(S)$.
Let \( z \) be a sequence of functions \( z_{CD} : \Delta(S) \rightarrow \mathbb{R}^n, C^D \in \mathcal{R} \), such that 
\[
\|z_{CD}(q)\|^2 \leq n, \text{ for all } C^D \in \mathcal{R}.
\]

Given \( \varepsilon > 0 \), 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 
   \[
   G(q, i) \equiv \sum_{C^D \in \mathcal{R}} \lambda(C^D) \langle z_{CD}(q), I(i) - q \rangle ,
   \]
   where \( I(i) \) is the \( n \)-vector with 1 in the \( i \)-th coordinate and zero elsewhere.

**Lemma A2** There exists a mixed strategy for player 1 which limits the payoff to player 2 to no more than \( \varepsilon \).

**Proof:** Given a mixed strategy for player 2, \( p \in \Delta(S), p = (p_1, \ldots, p_n) \), choose \( \hat{q} \in Q(\varepsilon) \) for player 1 so that \( \|p - \hat{q}\| \leq \frac{\varepsilon}{\sqrt{n}} \). Then,
\[
\sum_{i \in S} p_i G(\hat{q}, i) = \sum_{i \in S} p_i \left( \sum_{C^D \in \mathcal{R}} \lambda(C^D) \langle z_{CD}(\hat{q}), I(i) - \hat{q} \rangle \right) = \]
\[
\sum_{C^D \in \mathcal{R}} \lambda(C^D) \left( \sum_{i \in S} p_i \langle z_{CD}(\hat{q}), I(i) - \hat{q} \rangle \right) = \]
\[
\sum_{C^D \in \mathcal{R}} \lambda(C^D) \left( \sum_{i \in S} p_i (I(i) - \hat{q}) \right) = \]
\[
\sum_{C^D \in \mathcal{R}} \lambda(C^D) (\langle z_{CD}(\hat{q}), p - \hat{q} \rangle) \leq \sum_{C^D \in \mathcal{R}} \lambda(C^D) (\|z_{CD}(\hat{q})\| \|p - \hat{q}\|) \leq \]
\[
\sum_{C^D \in \mathcal{R}} \lambda(C^D) \varepsilon = \varepsilon.
\]
Hence, for any mixed strategy for player 2 there exists a pure strategy for player 1 that gives player 2 a payoff smaller than $\varepsilon$. By the Minimax Theorem, there exists a mixed strategy for player 1 which guarantees that the payoff of player 2 will not exceed $\varepsilon$, independently of player 2’s action.

**Corollary 1.** Given any $\varepsilon > 0$ and any collection of functions $z = \{z_{CD}\}$, $C^D \in \mathcal{R}$, $z_{CD} : \Delta(S) \rightarrow \mathbb{R}^n$ such that $\|z_{CD}(q)\|^2 \leq n$, there exists a probability measure $\mu \in \Delta(\Delta(S))$ such that for every $i \in S$

$$E_{\lambda \times \mu}(z_{CD}(q), I(i) - q) \leq \varepsilon,$$

where $E_{\lambda \times \mu}$ is the expectation operator associated with $\lambda \times \mu$. By definition $\mu \in \Delta(\Delta(S))$ depends only on $\varepsilon$ and $z$.

### 4.2 Description of the forecasting scheme

We now describe the forecasting scheme $\zeta$. Let $\varepsilon_t > 0$ be a sequence of real numbers such that:

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

For example, let $\varepsilon_t$ be $\frac{1}{t}$. We now define the following functions: Given a forecast-based checking rule $C^D \in \mathcal{R}$, and a $t$–history $x^t \in X^t$, $x^t = (s^t, f^{t-1})$,

1. Let $T_t(C^D, x^t) \in \mathbb{N}$ be the number of times that $C^D \in \mathcal{R}$ was active until period $t$, i.e.

$$T_t = T_t(C^D, x^t) = \sum_{j=1}^{t} C^D(x^j),$$

where $x^j$ is the prefix of length $j$ of $x^t$.

2. Let $g_t(C^D, x^t) \in \mathbb{R}^n$ be $0$ if $C^D$ is inactive at $x^t$ and

$$g_t(C^D, x^t) = I(s_t) - f_{t-1}$$

if $C^D$ is active at $x^t$, where $f_{t-1} \in \Delta(S)$ is the forecast announced at period $t - 1$ and $s_t \in S$ is the outcome observed at period $t$. 

18
3. Let \( \rho_t(C^D, x^t) \in \mathbb{R}^n \) be the calibration score at period \( t \). That is,

\[
\rho_t = \rho_t(C^D, x^t) = \frac{\sum_{j=1}^{t} C^D(x^j) (I(s_j) - f_{j-1})}{\sum_{j=1}^{t} C^D(x^j)} = \frac{T_t - T_{t-1} + g_t}{T_t}.
\]

Given \( x^t \in X^t \), define the collection of functions \( z_{C^D}: \Delta(S) \rightarrow \mathbb{R}^n \) as follows:

\[
z_{C^D}(q) = \begin{cases} 
\frac{\rho_t(C^D, x^t)}{T_t(C^D, x^t) + 1} & \text{if } q \in D; \\
0 & \text{if } q \notin D.
\end{cases}
\]

By corollary 1, given \( x^t \in X^t \), there exists a probability measure \( \mu \in \Delta(\Delta(S)) \), with a finite support \( Q \), such that for any realization of \( s_{t+1} \in S \)

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

We define

\[
\zeta(x^t) = \mu.
\]

By construction, \( \zeta \) is well-defined because \( \mu \) depends only on \( x^t \) (and \( \varepsilon_{t+1} \) which is a function of \( t \)).

### 4.3 Proof of proposition 1

Fix a path \( s \in S^\infty \). Let \( X = \mathcal{R} \times \Delta(S)^\infty \) and let \( \nu \) be \( \lambda \times \zeta^* \). Let \((X, \mathcal{B}, \nu)\) be a probability measure space. Let \( \{T_t\}_{t=0}^{\infty} \) be a sequence of random variables taking values in \( N_+ \) and let \( \{\bar{g}_t\}_{t=0}^{\infty} \) and \( \{\bar{\rho}_t\}_{t=0}^{\infty} \) be sequences of random variables taking values in \( \mathbb{R}^n \) such that for any \((C^D, f) \in \mathcal{R} \times \Delta(S)^\infty, \)

\[
\bar{T}_t(C^D, f) = T_t(C^D, x^t);
\]

\[
\bar{g}_t(C^D, f) = g_t(C^D, x^t);
\]

\[
\bar{\rho}_t(C^D, f) = \rho_t(C^D, x^t);
\]

where \( x^t = (s^t, f^{t-1}) \) is the prefix of length \( t \) of \((s, f)\) and \( T_t, g_t, \) and \( \rho_t \) are defined in section 4.2.

Assumptions 1 – 4 of Lemma A1 are satisfied by \( \bar{T}_t, \bar{g}_t, \bar{\rho}_t \). To see this simply note that \( T_t \) is non-decreasing since it is a sum of non-negative numbers; \( T_t - T_{t-1} \leq 1 \) because \( T_t - T_{t-1} \) is either zero or one; \( g_t \in [-1, 1]^n \) follows directly from its definition and \( T_t - T_{t-1} = 0 \) implies \( g_t = 0 \) follows from the
fact that \( g_t \) is \( \bar{0} \) when the forecast-based checking rule is inactive. By point (3) in the description of forecasting scheme (see section 4.2),

\[
\rho_t = \frac{T_{t-1} \rho_{t-1} + g_t}{T_t}.
\]

So, the same properties hold for \( \bar{T}_t, \bar{g}_t, \bar{\rho}_t \). We now demonstrate that assumption 5 of Lemma A1 is also satisfied. That is,

\[
E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \right) \leq \varepsilon_{t+1}.
\]

Given a \( t \)-sequence of forecasts \( f^{t-1} \in \Delta(S)^t \), let \( x^t \) be the \( t \)-history \( (s^t, f^{t-1}) \), where \( s^t \) is the prefix of length \( t \) of the path \( s \). Note that \( x^t \) is well-defined because the path \( s \in S \) is fixed (independent of \( f^{t-1} \)). Let

\[
E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \mid x^t \right)
\]

be the expectation of \( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \) conditional on the \( t \)-sequence of forecasts being \( f^{t-1} \in \Delta(S)^t \). Given a forecast \( q \in \Delta(S) \), let \( x^{t+1}(q) = (s^{t+1}, f^t) \), where \( f^t = (f^{t-1}, q) \). Let \( Q, z_{CD}, \mu \) be as defined in section 4.2. Then,

\[
E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \mid x^t \right) =
\sum_{\{(C^D, q) \in R \times Q\}} \lambda(C^D) \mu(q) \left\langle \frac{\rho_t(C^D, x^t)}{T_{t+1}(C^D, x^t)}, g_{t+1}(C^D, x^{t+1}(q)) \right\rangle =
\sum_{\{(C^D, q) \in R \times Q\}} \lambda(C^D) \mu(q) \langle z_{CD}(q), I(s_{t+1}) - q \rangle
\]

\[
= E_{\lambda \times \mu} ( z_{CD}(q), I(s_{t+1}) - q ) \leq \varepsilon_{t+1}.
\]

So,

\[
E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \mid x^t \right) \leq \varepsilon_{t+1}.
\]

Therefore,

\[
E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \right) = E_\nu \left( E_\nu \left( \left\langle \frac{\bar{\rho}_t}{\bar{T}_{t+1}}, \bar{g}_{t+1} \right\rangle \mid x^t \right) \right) \leq \varepsilon_{t+1}.
\]
So, assumption 5 of Lemma A.1 is satisfied. By Lemma A.1, \( \nu - a.e. \), \( \bar{T}_t(C^D, f) \to \infty \) implies that \( \bar{\rho}_t(C^D, f) \to 0 \). By definition, all probabilities \( \lambda(C^D) \) are strictly positive. Hence, \( \bar{\zeta}_s - a.e. \), for all \( C^D \in \mathcal{R} \), \( \bar{T}_t(C^D, f) \to \infty \) implies that \( \bar{\rho}_t(C^D, f) \to 0 \). Equivalently, for all \( C^D \in \mathcal{R} \), \( \bar{\zeta}_s - a.e. \), \( T_t(C^D, x^t) \to \infty \) implies that \( p_t(C^D, x^t) \to 0 \).

References


21


