Good randomized sequential probability forecasting is always possible

Vladimir Vovk
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England
vovk@cs.rhul.ac.uk

Glenn Shafer
Rutgers Business School - Newark and New Brunswick
180 University Avenue, Newark, New Jersey 07102 USA
and
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England
gshafer@rutgers.edu

July 25, 2005

Abstract

Building on the game-theoretic framework for probability, we show that it is possible, using randomization, to make sequential probability forecasts that will pass any given battery of statistical tests. This result, an easy consequence of von Neumann's minimax theorem, simplifies and generalizes work by earlier authors.

Keywords: probability forecasting; calibration; game-theoretic probability

1 Introduction

In a recent book (Shafer and Vovk, 2001), we introduced a purely game-theoretic framework for probability theory. In this article, we build on that framework to demonstrate the possibility of good probability forecasting.

In Section 2, we review the prototypical game studied in our book. One of the players, a sceptic, bets repeatedly at odds given by a probability forecaster.

The sceptic can become infinitely rich unless reality respects the forecaster's odds over the long run. In Section 3, we formulate a game that better represents the challenge faced by the forecaster, as opposed to the sceptic. In this new game, which we did not consider in our book, the forecaster faces a sceptic whose strategy is revealed in advance, and he is allowed to use a degree of randomization to conceal each of his probability forecasts until the corresponding outcome has been announced. Our main result, stated and proven in Section 4, says that the forecaster can keep the sceptic from becoming infinitely rich, no matter how reality chooses the outcomes. This means that relative to the sceptic's strategy, the outcomes will look like random events with the forecasted probabilities. This result is an easy consequence of von Neumann's minimax theorem, but it is somewhat surprising. As we explain in Section 5, it suggests that we can make an arbitrary sequential process in the real world look stochastic in any specified respect.

In the usual measure-theoretic framework for probability, an asymptotic statistical test can be defined by specifying an event of probability zero; the test is passed if the event does not happen (Martin-Löf 1966). In our framework, a statistical test is a betting strategy for the sceptic; the test is passed if the sceptic does not become infinitely rich. Because we consider any betting strategy for sceptic (any statistical test), our result strengthens earlier results in Foster and Vohra (1998), Fudenberg and Levine (1999), Lehrer (2001), and Sandroni et al. (2003). These articles made weaker demands on the forecaster; instead of requiring that the probabilities he gives as forecasts pass any statistical test, Foster and Vohra asked only that the entire sequence of probabilities be properly calibrated, and the other authors added only the demand that certain subsequences also be properly calibrated. A sequence of probabilities is properly calibrated when the difference between the average of the probabilities and the observed relative frequency of the events being forecast converges to zero; for a precise statement see equations (4) and (6) below. When calibration fails, whether on the entire sequence or on a subsequence, a statistical test has been failed. But as Ville (1939) demonstrated, there are other statistical tests that go beyond calibration (see also Li and Vitányi 1997, p. 313). We can check, for example, whether the convergence required by calibration is at the rate required the law of the iterated logarithm

In the measure-theoretic framework, violation of the law of the iterated logarithm is an event of probability zero. In our framework, there is a betting strategy for sceptic that makes him infinitely rich when it happens (Shafer and Vovk 2001, Chapter 5).

So far as mathematical technique is concerned, this article holds little novelty, for our argument from the minimax theorem was already used by some of the authors concerned with proper calibration (Foster and Vohra 1998, Fudenberg 1999, Sandroni et al. 2003). Our contribution is to put the argument in our game-theoretic framework and to show that it can lead to forecasts with stochastic properties going beyond calibration. The earlier articles we have cited did not quite exhaust the argument's potential even for calibration. As we show in Appendix A (see the comments preceding Theorem 5), our result implies a general statement about tests of calibration that is stronger and simpler than the strongest previous statement, the one given by Sandroni et al. (2003).

In another recent article, Sandroni (2003) has given a measure-theoretic version of our result. As we show in Appendix B, Sandroni's result can be derived quite easily, modulo technicalities, from our Theorem 4. It is less general than our Theorem 4 in several respects, most notably in that it makes the restrictive assumption that outcomes are generated by a probability measure.

Our result also has philosophical significance beyond that of the earlier work, because it goes beyond calibration to questions about the meaning of probability. Within the game-theoretic framework, the requirement that probabilities resist any betting strategy defines their very meaning. Some readers might consider calibration equally fundamental, arguing that the project of properly calibrating subsequences is in the spirit of the frequentist foundation of probability advanced by von Mises (1919), and it is true that von Mises's approach is still sometimes presented as a legitimate competitor to interpretations of probability based on betting. But as Ville (1939) pointed out, it is deficient because it does not require as much irregularity as classical probability theory does. A sequence satisfying von Mises's conditions satisfies the law of large numbers, but it need not satisfy other predictions of probability theory, such as the law of the iterated logarithm. Our game-theoretic framework provides one way, in our judgement the simplest way formulated to date, of correcting this deficiency. From our point of view, the work on properly calibrated forecasting that we are extending stays too close to von Mises, and this makes it unnecessarily complex. As we see it, the story is simpler in the game-theoretic framework. For more on the historical background of the game-theoretic framework, see Shafer and Vovk (2004), Vovk and Shafer (2004) and Chapter 2 of Shafer and Vovk (2001).

2 The game-theoretic framework for probability

In this section, we review the elements of our game-theoretic framework, with an emphasis on the idea of probability forecasting. For a review of earlier work on probability forecasting, see Dawid (1986).

Probability forecasting can be thought of as a game with two players, Forecaster and Reality. On each round, Forecaster gives probabilities for what Reality will do. Assuming, for simplicity, that Reality makes a binary choice on each round, we might begin our description of the game with this protocol:

FOR $n = 1, 2, \dots$:
  Forecaster announces $p_n \in [0, 1]$.
  Reality announces $x_n \in \{0, 1\}$.

This is a perfect-information protocol; the players move in the order indicated, and each player sees the other player's moves as they are made. The players may also receive other information as play proceeds; we make no assumption about what other information each player does or does not receive.

Forecaster's goal, broadly conceived, is to state probabilities that pass all possible statistical tests in light of Reality's subsequent moves. To formalize this goal, we add a third player, Sceptic, who seeks to refute Forecaster's probabilities. Sceptic is allowed to bet at the odds defined by Forecaster's probabilities, and he refutes the probabilities if he gets infinitely rich. This produces a fully specified perfect-information game:

BINARY FORECASTING GAME I
Players: Forecaster, Sceptic, Reality
Protocol:

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $p_n \in [0, 1]$.
Sceptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n)$.

Restriction on Sceptic: Sceptic must choose M_n so that his capital remains nonnegative ($\mathcal{K}_n \geq 0$) no matter what value Reality announces for x_n.
Winner: Sceptic wins if $\mathcal{K}_n$ tends to infinity. Otherwise Forecaster wins.

This protocol specifies both an initial value for Sceptic's capital ($\mathcal{K}_0 = 1$) and a lower bound on its subsequent values ($\mathcal{K}_n \geq 0$). The asymptotic conclusions we draw about the game will not change if these numbers are changed, but some lower bound is needed in order to prevent Sceptic from recouping losses by borrowing ever more money to make ever larger bets.

An *internal strategy* for one of the players in this game is a rule that tells the player how to move on each round based on the previous moves by the other players. The word "internal" here refers to the fact that the strategy uses only information internal to the game, ignoring other information the player might receive. We call an internal strategy for Sceptic *legal* if it respects the condition that Sceptic move so that his capital always remains nonnegative, no matter how the other players move.

In the game as we have defined it, Forecaster and Sceptic have opposite goals. One of them wins, and the other loses. We have not assigned a goal to Reality, but she is in a position to determine which of the other two players wins. The exact sense in which this is true is spelled out in the next two theorems.

Making Forecaster win is easy; Reality can do this even without Forecaster's cooperation:

Theorem 1 *Reality has an internal strategy that assures that Forecaster wins.*

Proof Consider the strategy for Reality that always sets

$$x_n := \begin{cases} 1 & \text{if } M_n \leq 0 \\ 0 & \text{if } M_n > 0. \end{cases}$$

When Reality follows this strategy, Sceptic's capital increment $M_n(x_n - p_n)$ is never positive and so his capital cannot tend to infinity. ∎

Making Sceptic win is harder, because Sceptic can keep himself from winning by never betting (always setting $M_n := 0$). But if Sceptic makes large enough bets, Reality can assure that he wins. If Sceptic and Reality cooperate closely, they can assure that Sceptic wins spectacularly:

Theorem 2 *Sceptic and Reality can jointly assure that Sceptic wins. More precisely, there is a legal internal strategy for Sceptic and an internal strategy for Reality such that Sceptic wins when the two players follow these strategies.*

Proof Consider the strategies that call for Sceptic to announce

$$M_n := \begin{cases} \mathcal{K}_{n-1}/(1-p) & \text{if } p_n < 0.5 \\ -\mathcal{K}_{n-1}/p & \text{otherwise,} \end{cases}$$

and for Reality to announce

$$x_n := \begin{cases} 1 & \text{if } p_n < 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

The strategy for Sceptic is legal, and when Sceptic and Reality follow these strategies, Sceptic's capital doubles on each round. ∎

The simple argument in this proof has goes back at least to Putnam (1963); see Dawid's comments in Oakes (1985).

Because Reality can largely determine the winner, the following hypothesis is a nonvacuous prediction about Reality's behaviour:

> **Hypothesis of the Excluded Gambling System:** No matter how Sceptic plays, Reality will play so that Sceptic does not win the game.

This hypothesis is not a mathematical assumption. Rather, it is a way of connecting our mathematical formalism, a formal game, with the real world, where the x_n appear. It provides an interpretation in the real world of the probabilities p_n.

When we adopt the hypothesis of the excluded gambling system in a particular real-world forecasting setup, we are expressing confidence in the theory or the person supplying the probabilities. Of course, we never adopt it more than provisionally. When we do adopt it, we say that a property E of the sequence $p_1x_2p_2x_2\ldots$ happens *almost surely* in the game-theoretic sense if Sceptic has an internal strategy that wins the game whenever the actual sequence $p_1x_2p_2x_2\ldots$ fails to satisfy E.

In Shafer and Vovk (2001), we justify the hypothesis of the excluded gambling strategy by showing that it gives meaning to the classical predictions of probability theory. Consider, for simplicity, a probability measure P on $\{0,1\}^\infty$ that assigns non-zero probability to every finite sequence $x_1\ldots x_n$. Suppose Forecaster uses P's conditional probabilities as his moves,

$$p_n := P(x_n = 1 \mid x_1, \ldots, x_{n-1}), \qquad (1)$$

and suppose E is a measurable subset of $\{0,1\}^\infty$. Then, as we show in Section 8.1 of (Shafer and Vovk 2001):

1. If $P(E) = 0$, then Sceptic has a legal internal strategy that wins the game if E happens.

2. If $P(E) > 0$, then Sceptic does not have such a strategy.

In other words, an event happens almost surely in the game-theoretic sense if and only if it has probability one.

These results for the binary case go back to Ville (1939), but we show that they generalize to more general forecasting games. Instead of having Reality choose from $\{0,1\}$, we can have her choose from some other measurable space Ω. Forecaster's moves will then come from $\mathcal{P}(\Omega)$, the set of all probability measures on Ω, and Sceptic will gamble on each round by choosing a payoff that has zero or perhaps negative expected value with respect to the measure chosen by Forecaster on that round. (See, for example, the Randomization Subgame described in Section 3.2.) In this case as in the binary case, Reality must avoid any given set of probability zero in order to keep Sceptic from becoming infinitely rich.

Historically, the principle that a given set of small or zero probability will not happen has been considered fundamental to the interpretation of probability by many authors, including Kolmogorov (1933). It is sometimes called Cournot's principle (Shafer and Vovk 2004). Within our game-theoretic framework, we use *Cournot's principle* as another name for our hypothesis of the excluded gambling system.

We conclude this brief review of the game-theoretic framework with these clarifications:

1. Because they allow Reality to play strategically and even collaborate with other players, our games diverge from the usual picture of stochastic processes, in which the outcome is not thought to be affected by how anyone is betting. But the principal mathematical results in Shafer and Vovk (2001) assert that Sceptic has strategies that achieve certain goals. If Sceptic can achieve a goal even when Reality and Forecaster do their worst against him as a team, he can also achieve it when Reality is indifferent to the game and Forecaster has no advance knowledge of how Reality will behave. Allowing Forecaster and Reality to play as a team makes our results worst-case results.

2. The framework does not require that Forecaster's move on the nth round be derived from a probability measure for $x_1 x_2 \dots$ specified in advance of the game. On the contrary, when deciding how to move on the nth round, Forecaster may use any new information and any new ideas that come his way by that time.

3. Instead of giving Sceptic the goal of making his capital tend to infinity (so that Forecaster's goal is to keep it from tending to infinity), we sometimes

give Sceptic the goal of making his capital unbounded (so that Forecaster's goal is to keep it bounded). The two formulations are equivalent for our purposes, because if Sceptic has a strategy that guarantees his capital will be unbounded, then he has a strategy that guarantees it will tend to infinity. (See the last paragraph of the proof of Theorem 3.) We will not hesitate to take advantage of this equivalence. (See, for example, the first two bullets in Section 5.1.)

4. Many of the classical results of probability theory hold in our framework even when Forecaster offers bets that fall short of defining probability distributions for Reality's move. This feature of our framework is important for applications to finance, because only a limited number of instruments can be priced by a securities market. But we are not concerned with this feature in the present article.

5. Finally, infinities are not essential to our story. Although we have been talking, for brevity, about a probability forecaster's performance being tested by an adversary who seeks to become infinitely rich over an infinite sequence of bets, we can also develop a more useful but more complicated finitary picture, where the adversary seeks only to become very rich by means of finitely many bets. See Chapters 6 and 7 of Shafer and Vovk (2001) and Theorem 4 in Section 4 of this article.

3 The challenge to Forecaster

The work reviewed in the preceding section emphasizes what Sceptic can achieve—how Sceptic can become infinitely rich if Reality violates various predictions. In the remainder of this article, we look at the challenge faced by Forecaster. We know (Theorem 2) that Forecaster cannot win if Sceptic and Reality are both working against him. But what can he achieve if Reality is somehow neutral, paying no attention to the forecasts?

Previous authors, most recently Sandroni et al. (2003), have noted that Forecaster's position can be strengthened if he is allowed to randomize his forecasts. When he does this, he still cannot be absolutely sure of winning, but as we will show, he can win almost surely with respect to the probabilities involved in the randomization. We can think of the randomization as a way of hiding full knowledge of the p_n from Reality and thus assuring a measure of neutrality on her part.

To make this story fully game-theoretic, we need a game that represents Sceptic's strength—the power of the internal strategies available to him—while also representing Forecaster's randomization. In this section, we construct such a game. In this game, a number of strategies Sceptic might want to use are combined into a single strategy, which makes Sceptic infinitely rich if any of the individual strategies do. Forecaster's randomization, on the other hand, is represented by a subgame between him and a random number generator.

In Subsection 3.1 we explain how Sceptic combines strategies. In Subsection 3.2, we explain the randomization subgame. In Subsection 3.3, we combine these two ideas into a formal game. We will show that Forecaster has a winning strategy in this game in Section 4.

3.1 Sceptic's strength

Consider first the problem of representing Sceptic's strength.

The most important point here is the following feature of our game:

Proposition 1 *Suppose $\mathcal{R}_1$ and $\mathcal{R}_2$ are legal internal strategies for Sceptic, and set*

$$\mathcal{R} = \alpha_1 \mathcal{R}_1 + \alpha_2 \mathcal{R}_2,$$

where α_1 and α_2 are nonnegative real numbers adding to one. Then $\mathcal{R}$ is also a legal internal strategy for Sceptic, and $\mathcal{R}$ wins whenever $\mathcal{R}_1$ wins or $\mathcal{R}_2$ wins.

Proof An internal strategy $\mathcal{Q}$ for Sceptic is a function that assigns a real number to every finite sequence of moves by Forecaster and Reality of the form $p_1 x_1 \ldots p_n$. Such a function recursively determines a capital process $\mathcal{L}$ for Sceptic: $\mathcal{L}(\Box) = 1$, where $\Box$ is the empty sequence, and

$$\mathcal{L}(p_1 x_1 \ldots p_n x_n) = \mathcal{L}(p_1 x_1 \ldots p_{n-1} x_{n-1}) + \mathcal{R}(p_1 x_1 \ldots p_n)(x_n - p_n). \qquad (2)$$

The internal strategy $\mathcal{Q}$ is legal if and only if $\mathcal{L}$ is everywhere nonnegative.

Let $\mathcal{K}_1$, $\mathcal{K}_2$, and $\mathcal{K}$ be the capital processes for $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}$, respectively. By (2), $\mathcal{K} = \alpha_1 \mathcal{K}_1 + \alpha_2 \mathcal{K}_2$. It follows that (i) $\mathcal{K}$ is everywhere nonnegative whenever $\mathcal{K}_1$ and $\mathcal{K}_2$ are, and thus $\mathcal{R}$ is legal whenever $\mathcal{R}_1$ and $\mathcal{R}_2$ are, and (ii) on any path $p_1 x_1 p_2 x_2 \ldots$ where $\mathcal{K}_1$ or $\mathcal{K}_2$ tends to infinity, $\mathcal{K}$ also tends to infinity. ∎

If Forecaster is considering two different legal internal strategies for Sceptic, $\mathcal{R}_1$ and $\mathcal{R}_2$, and he wants to find a strategy of his own that beats both of them, then according to this proposition, it is enough for him to find a strategy that beats $\alpha_1 \mathcal{R}_1 + \alpha_2 \mathcal{R}_2$. This conclusion generalizes from any pair to any finite set of legal internal strategies and even to any countably infinite set of legal internal strategies. It also generalizes from internal strategies to strategies that use any other information that we can assume in advance will be available to Sceptic.

There are no more than a countable number of statistical tests we might ask Forecaster's probabilities to pass, and hence no more than a countable number of legal strategies for Sceptic that Forecaster needs to counter. Indeed, as Wald (1937) explained, there are only a countable number of sentences in any formal language that we might use to formulate tests. But as a practical matter, we cannot specify all the tests that interest us, and Proposition 1 depends on an asymptotic notion of winning that loses contact with practicality when we try to average too many strategies. The average will tend to infinity when any of its components does, but not as fast. So we do not want to exaggerate the significance of the possibility of averaging strategies for Sceptic. We assert only that in some circumstances it can allow Sceptic to capture the aspects of randomness (including calibration) that interest us. For a closer look, see Vovk (2005).

Because we are asking Forecaster to defeat only a single strategy for Sceptic, we can clarify Forecaster's task by requiring Sceptic to announce this strategy before Forecaster moves. If the players did not receive information from outside the game in the course of play, then we might simply require Sceptic to announce an internal strategy for the whole game at the outset. But because our framework does allow both Forecaster and Sceptic to receive information from outside of game, and because some of this information might be unanticipated or informal, we instead require only that Sceptic announce a strategy for each round before Forecaster moves on that round. For simplicity, we assume that this strategy is internal at least in the sense that at the point where it is announced, the only not-yet-received information it uses is Forecaster's not-yet-announced move.

The following game, which can be thought of as a subgame of the game we will formulate in Subsection 3.3, expresses these ideas formally. It differs from Binary Forecasting Game I in only one way: Sceptic now moves first on each round, announcing a strategy that is bounded as a function of Forecaster's forthcoming move.

FORECASTING SUBGAME
Players: Sceptic, Forecaster, Reality
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \dots$:
Sceptic announces a bounded function $S_n : [0, 1] \to \mathbb{R}$.
Forecaster announces $p_n \in [0, 1]$.
Reality announces $x_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$.
Restriction on Sceptic: Sceptic must choose S_n so that his capital remains nonnegative ($\mathcal{K}_n \geq 0$) no matter what values Forecaster and Reality announce for p_n and x_n.
Winner: Sceptic wins if $\mathcal{K}_n$ tends to infinity.

The strategy S_n takes only the subsequent move by Forecaster, p_n, into account. But in choosing S_n, Sceptic may take into account both previous moves in the game and any other information received before the round begins. When we say that S_n is bounded, we mean that $\sup_p |S_n(p)| < \infty$. We do not require that a uniform bound exist for all the S_n together, and we require no other regularity of the S_n; we do not even require that they be measurable.

3.2 Allowing Forecaster to randomize

In our purely game-theoretic approach, the notion that Forecaster moves randomly must also be represented game-theoretically. We can do this by splitting

Forecaster into two players. The first player decides on the probabilities and sets up the randomizing device (this is the role of a person who constructs and spins a roulette wheel); the second player then decides on the outcome of the randomization (this is the role of the roulette wheel). Calling the first player Forecaster and the second Random Number Generator, we can describe their interaction in terms of a game analogous to our Binary Forecasting Game I:

RANDOMIZATION SUBGAME
Players: Forecaster, Random Number Generator
Protocol:

$\mathcal{F}_0 := 1$.
FOR $n = 1, 2, \dots$:
Forecaster announces $P_n \in \mathcal{P}[0,1]$.
Forecaster announces $f_n : [0,1] \to \mathbb{R}$ such that $\int f_n dP_n \leq 0$.
Random Number Generator announces $p_n \in [0,1]$.
$\mathcal{F}_n := \mathcal{F}_{n-1} + f_n(p_n)$.

Restriction on Forecaster: Forecaster must choose P_n and f_n so that his capital remains nonnegative ($\mathcal{F}_n \geq 0$) no matter what value Random Number Generator announces for p_n.
Winner: Forecaster wins if his capital $\mathcal{F}_n$ tends to infinity.

The f_n must be measurable; this is needed in order for the integral to be defined. But as we will see when we prove Theorem 3 in Section 4, Forecaster can achieve what we want him to achieve even if we put much stronger restrictions on the f_n; e.g., we can require that they be continuous and piece-wise linear.

On the first of his two moves on the nth round, Forecaster announces a probability distribution P_n for p_n. On the second, he sets up the randomizing device, by making bets that force Random Number Generator to make p_n look random with respect to P_n. He bets by choosing a gamble f_n that is either fair ($\int f_n dP_n = 0$) or unfavourable to him ($\int f_n dP_n < 0$). If Forecaster gets infinitely rich with such bets, then we will think that Random Number Generator has done a bad job—i.e., has not made his $p_1, p_2, \dots$ look like draws from the sequence $P_1, P_2, \dots$ of probability measures.

We will adopt Cournot's principle for the Randomization Subgame. We will assume, that is to say, that Random Number Generator will do a good job, playing so that $\mathcal{F}_n$ necessarily stays bounded.

3.3 The game to challenge Forecaster

Combining our two ideas—announcing Sceptic's strategy at the outset of each round and randomizing the probability forecasts—we obtain a perfect-information game involving four players:

BINARY FORECASTING GAME II
Players: Sceptic, Forecaster, Reality, Random Number Generator
Protocol:

$\mathcal{K}_0 := 1$.
$\mathcal{F}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Sceptic announces a bounded function $S_n : [0,1] \to \mathbb{R}$.
Forecaster announces $P_n \in \mathcal{P}[0,1]$.
Reality announces $x_n \in \{0,1\}$.
Forecaster announces $f_n : [0,1] \to \mathbb{R}$ such that $\int f_n dP_n \leq 0$.
Random Number Generator announces $p_n \in [0,1]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$.
$\mathcal{F}_n := \mathcal{F}_{n-1} + f_n(p_n)$.

Restriction on Sceptic: Sceptic must choose S_n so that his capital remains nonnegative ($\mathcal{K}_n \geq 0$) no matter how the other players move.
Restriction on Forecaster: Forecaster must choose P_n and f_n so that his capital remains nonnegative ($\mathcal{F}_n \geq 0$) no matter how the other players move.
Winner: Forecaster wins if either (i) his capital $\mathcal{F}_n$ tends to infinity or (ii) Sceptic's capital $\mathcal{K}_n$ stays bounded.

As we explained when we described the subgames in the preceding subsections, the game puts no restrictions on S_n or f_n aside from the requirements that each S_n be bounded and each f_n be measurable.

If Random Number Generator does a good job in the game ($\mathcal{F}_n$ does not tend to infinity), then Forecaster wins the game if and only if Sceptic does not detect any disagreement between Forecaster and Reality ($\mathcal{K}_n$ stays bounded). In the next section, we prove that Forecaster has a winning strategy. If Forecaster uses this strategy, then Random Number Generator can guarantee that the p_n are good probability forecasts for the x_n just by making sure that they look like random draws from the P_n.

In the protocol as we have set it up, Random Number Generator actually announces p_n after Reality announces x_n, and this makes it awkward to think of p_n as a probability forecast of x_n. But our result (Forecaster has a winning strategy) is not affected if we make an exception to our presumption of perfect information by supposing that Random Number Generator sees the x_n later or perhaps never at all, and this should not hamper his ability to make the p_n look like random draws from the P_n. We will return to this point in Section 5.1, where we recast the protocol so that x_n is announced after p_n.

One might also worry that the protocol grants too much to Sceptic and Reality. We could tie Sceptic down more by requiring him to announce an internal strategy for the entire game at the outset. We could make the neutrality of Reality more explicit by requiring her to choose her entire sequence of moves $x_1 x_2 \ldots$ before play begins, even though she announces these moves to the other players according to the indicated schedule. But because Forecaster has a winning strategy in the game as laid out, there is no point in changing the game to strengthen Forecaster's hand. Forecaster's winning strategy will remain a winning strategy when the other players are weakened.

4 Good probability forecasts

Theorem 3 *Forecaster has a winning internal strategy in Binary Forecasting Game II.*

Proof Imagine for a moment that Forecaster and Reality play the following zero-sum game on round n after Sceptic announces the bounded function S_n:

GAME ON ROUND n
Players: Forecaster, Reality
Protocol:
  Simultaneously:
    Forecaster announces $p_n \in [0,1]$.
    Reality announces $x_n \in \{0,1\}$.
Payoffs: Forecaster loses (and Reality gains) $S_n(p_n)(x_n - p_n)$.

The value of this game is at most zero, because for any mixed strategy Q for Reality (any probability measure Q on $\{0,1\}$), Forecaster can limit Reality's expected gain to zero by choosing $p_n := Q\{1\}$. In order to apply von Neumann's minimax theorem, which requires that the move spaces be finite, we replace Forecaster's move space $[0,1]$ with a finite subset A_n of $[0,1]$. Fixing $\epsilon > 0$ and using the boundedness of S_n, we choose A_n dense enough in $[0,1]$ that the value of the game is smaller than $\epsilon 2^{-n}$. The minimax theorem then tells us that Forecaster has a mixed strategy P_n (a probability measure on $[0,1]$ concentrated on A_n) such that

$$\int S_n(p)(x-p)P_n(dp) \le \epsilon 2^{-n} \tag{3}$$

for both $x = 0$ and $x = 1$.

Returning now to Binary Forecasting Game II, consider the strategy for Forecaster that tells him, on round n, to use the P_n just identified and to use as his second move the function f_n given by

$$f_n(p) := \frac{1}{1+\epsilon}\left(S_n(p)(x_n - p) - \epsilon 2^{-n}\right)$$

for $p \in A_n$ and defined arbitrarily for $p \notin A_n$. (This allows Forecaster to make f_n continuous and piece-wise linear if he wishes.) The condition $\int f_n dP_n \le 0$ is then guaranteed by (3). Comparing the sums

$$\mathcal{K}_n = 1 + \sum_{i=1}^{n} S_i(p_i)(x_i - p_i)$$

and

$$\mathcal{F}_n = 1 + \frac{1}{1+\epsilon}\sum_{i=1}^{n}\left(S_i(p_i)(x_i - p_i) - \epsilon 2^{-i}\right),$$

we see that

$$(1+\epsilon)\mathcal{F}_n = \mathcal{K}_n + \epsilon 2^{-n},$$

so that $\mathcal{K}_n \leq (1+\epsilon)\mathcal{F}_n$. This establishes that $\mathcal{F}_n$ is never negative and that either $\mathcal{K}_n$ will stay bounded or $\mathcal{F}_n$ will be unbounded.

To complete the proof, it suffices to show that for every legal strategy $\mathcal{T}$ for Forecaster, we can construct another legal strategy $\mathcal{T}^*$ such that whenever $\mathcal{T}$'s capital is unbounded, $\mathcal{T}^*$'s tends to infinity. This is easy to do. We choose some number larger than 1, say 2. Starting, as the game requires, with initial capital 1 for Forecaster, we have him play $\mathcal{T}$ until its capital exceeds 2. Then he sets aside 1 of this capital and continues with a rescaled version of $\mathcal{T}$, scaled down to the reduced capital. (This means he multiplies $\mathcal{T}$'s moves on succeeding rounds by the same factor as he has multiplied the capital at this point, thus assuring that the capital on succeeding rounds is also multiplied by this factor.) When the capital again exceeds 2, he again sets aside 1, and so forth. The money set aside, which is part of the capital earned by this strategy, grows without bound. For another way of constructing $\mathcal{T}^*$ from $\mathcal{T}$, see Shafer and Vovk (2001), p. 68.
∎

The essential idea of this proof—the application of von Neumann's minimax theorem—was used by several of the authors who worked on properly calibrated randomized forecasting, including Hart (Foster and Vohra 1998, pp. 383–384), Fudenberg and Levine (1999), and Sandroni et al. (2003). In Appendix A, we show that Theorem 1 implies the result obtained by Sandroni et al. (2003), the strongest result on properly calibrated randomized forecasting of which we are aware.

Like the results of previous authors, Theorem 3 generalizes beyond the case where Reality's moves are binary. Our proof generalizes directly to the case where Reality's move space Ω is a finite set, and the argument can probably also be extended to yet other games considered by Shafer and Vovk (2001).

Whereas previous work on properly calibrated forecasting seems to be essentially asymptotic (and has been criticized on this account; see Schervish's comment in Oakes (1985)), our game-theoretic result is not. We stated Theorem 3 in asymptotic form, but in the course of proving it, we also established a finitary result:

Theorem 4 *For any $\epsilon > 0$, Forecaster has a strategy in Binary Forecasting Game II that guarantees $\mathcal{K}_n \leq (1+\epsilon)\mathcal{F}_n$ for each n.*

Forecaster can guarantee $\mathcal{F}_n \geq \mathcal{K}_n$ to any approximation required, which means that every dollar gained by Sceptic can be attributed to the poor performance of Random Number Generator.

5 Discussion

Theorems 3 and 4 say that if you have a good random number generator, you can do a good job forecasting probabilistically how reality will behave. In this concluding section, we elaborate this message and its implications for how we think about stochasticity.

5.1 Variations on the game

We can vary Binary Forecasting Game II in several ways without losing its intuitive message. Here we look at a couple of variations that may be helpful to some readers.

Two games at once

The rule for winning in Binary Forecasting Game II treats the game as a single game involving four players. We could just as well, however, return to the picture developed in Section 3, in which Random Number Generator is simultaneously participating in two different games by announcing the p_n. All the players are in the same protocol—the protocol in Binary Forecasting Game II—but there are two games because there are two rules for winning:

- Against Forecaster, Random Number Generator is playing the Randomization Subgame we described in Section 3.2. Random Number Generator wins this game if and only if $\mathcal{F}_n$ stays bounded (so that the p_n look random with respect to the P_n). (Recall point 3 at the end of Section 2 concerning the equivalence of requiring that a player's capital not tend to infinity and requiring that it be bounded.)
- Against Sceptic, Random Number Generator is playing the Forecasting Subgame (in the role we gave to Forecaster when we first described that game on p. 9), which he wins if and only if $\mathcal{K}_n$ stays bounded (so that the x_n look random with respect to the p_n).

Because we have not changed the protocol for Binary Forecasting Game II, it remains true that Forecaster has a strategy that guarantees $\mathcal{K}_n \leq (1+\epsilon)\mathcal{F}_n$ for all n (Theorem 4). By playing this strategy, Forecaster guarantees that Random Number Generator wins the Forecasting Subgame whenever Random Number Generator wins the Randomization Subgame.

Putting x_n after p_n

As we have already mentioned, one counter-intuitive feature of Binary Forecasting Game II is that p_n is announced after the outcome x_n it is supposed to predict. Here is a way of changing the protocol so that x_n comes last, where it seems to belong.

FOR $n = 1, 2, \dots$:
  Sceptic announces a bounded function $S_n : [0,1] \to \mathbb{R}$.
  Forecaster announces $P_n \in \mathcal{P}[0,1]$.
  Forecaster announces, for $x = 0$ and $x = 1$,
    $f_n^x : [0,1] \to \mathbb{R}$ such that $\int f_n^x dP_n \leq 0$.
  Random Number Generator announces $p_n \in [0,1]$.
  Reality announces $x_n \in \{0,1\}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$.
$\mathcal{F}_n := \mathcal{F}_{n-1} + f_n^{x_n}(p_n)$.

Changing the protocol in this way does not invalidate the conclusion that Forecaster has a winning strategy. In order to see this, we need to think separately about two changes and why they do not weaken Forecaster:

- Forecaster now makes his second move before Reality announces x_n. Instead of waiting to see x_n and then announcing f_n, Forecaster announces a strategy for how f_n will depend on x_n.
- Random Number Generator announces p_n before Reality announces x_n.

A winning strategy remains a winning strategy when it is announced, in whole or in part, in advance. So it does not weaken Forecaster to announce how f_n will depend on x_n. Nor can the order in which Forecaster's opponents make their moves after he is finished diminish what he can achieve.

This version of the protocol helps us see that the randomization is a way of requiring Reality's neutrality. Intuitively, Reality's moves $x_1 x_2 \ldots$ have nothing to do with how well Random Number Generator can simulate random draws from announced probability measures on $[0,1]$. But if Forecaster makes f_n depend on x_n, Reality may be able to choose the x_n so that Random Number Generator fails Forecaster's test. When we adopt Cournot's principle for the Randomization Subgame, we are assuming Reality will not behave in this malicious way.

5.2 Is randomization really needed?

Theorem 2 seems to establish that Forecaster cannot win against Sceptic without the randomization we have studied in this article. It may be unreasonable, however, to ask Forecaster to defeat the extremely precise collaboration between Sceptic and Reality in the example we used to prove Theorem 2. Is it not possible that Forecaster might succeed without randomization if we ask him only to defeat more reasonable strategies by Sceptic?

One reasonable thing for Sceptic to do is to test for calibration. The traditional way of checking calibration is to divide the range of the forecasts, $[0,1]$, into a large number of small intervals and to check whether the average of the x_n for the p_n in each interval itself falls in or near that interval. We can easily translate this into a strategy for Sceptic in Binary Forecasting Game I. Say we use 4 intervals: $[0,0.25)$, $[0.25,0.5)$, $[0.5,0.75)$, $[0.75,1]$. Let $\mathcal{R}$ be the strategy constructed for Sceptic in Proposition 3.3 of Shafer and Vovk (2001), which keeps his capital nonnegative and makes him infinitely rich unless

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n}(x_i - p_i)}{n} = 0. \qquad (4)$$

Suppose Sceptic runs 4 copies of $\mathcal{R}/4$; copy 1 on the rounds where $p_n \in [0,0.25)$ (pretending the other rounds do not happen), copy 2 on the rounds where

$p_n \in [0.25, 0.5)$ (again pretending the other rounds to not happen), etc. Then Reality can defeat Forecaster by always choosing

$$x_n := \begin{cases} 1 & \text{if } p_n < 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

This is in the spirit of examples discussed by Oakes, Dawid and Schervish (1985), who conclude that many sequences of outcomes cannot be predicted probabilistically by a computable theory.

Even this example can be questioned, however. A strategy for Sceptic that requires him to distinguish whether $p_n < 0.25$ or $p_n \geq 0.25$ may be reasonable from a mathematical point of view, but it is not reasonable from a computational point of view. It is not continuous in p_n, and so it cannot really be implemented. It is not computable.

The fact that only continuous functions are computable, together with the fact that the strategies for Sceptic that we studied in Shafer and Vovk (2001) are continuous, suggests that we study a version of Binary Forecasting Game II in which Sceptic's move S_n is required to be continuous. Recent work by Vovk, Takemura, and Shafer (2005) shows that Forecaster can win such a game without randomization.

5.3 The meaning of stochasticity

We have shown that good randomized probability forecasting for a sequence $x_1x_2\ldots$ is always possible. If Forecaster is allowed to announce his probability p_n for each x_n after observing $x_1x_2\ldots x_{n-1}$, and he is allowed to randomize when choosing these probabilities, then he can make sure that they pass a given battery of statistical tests. This shows that probability theory applies broadly to the real world. But this breadth of applicability undermines some conceptions of stochasticity. If everything is stochastic to a good approximation, then the bare concept of stochasticity has limited content.

There is content in the assertion that a sequence obeys a probability distribution P that we specify fully before any observation. The assertion can be refuted when the sequence is observed, and if it is not refuted then Forecaster can avoid refutation himself only by agreeing with P's predictions in the limit (Vovk and Shafer 2004; Dawid 1984, p. 281). But there seems to be very little content in the assertion that a sequence is governed by a completely unknown probability distribution.

When he follows the strategy suggested by the proof of Theorem 3, is Forecaster using experience of the past to predict the future? He is certainly taking the past into consideration. Sceptic's moves signal emerging discrepancies that he would like to take advantage of, and Forecaster chooses his ps to avoid extending these discrepancies. But because he succeeds regardless of the xs, it is awkward to call Forecaster's ps predictions. Perhaps we should call them descriptions of the past rather than predictions of the future.

Kolmogorov once expressed puzzlement about the appearance in the real world of the kind of irregularity described by probability (Kolmogorov 1983, p. 1):

> In everyday language we call random those phenomena where we cannot find a regularity allowing us to predict precisely their results. Generally speaking there is no ground to believe that random phenomena should possess any definite probability. Therefore, we should have distinguished between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of the probability theory).
>
> There emerges a problem of finding the reasons for applicability of the mathematical theory of probability to the phenomena of the real world.

But when probability is used in a way that succeeds regardless of how events turn out, we do not need to look farther to find reasons for its success.

Acknowledgments

We are grateful to Phil Dawid, who inspired this article by calling our attention to Sandroni et al. (2003), and to Akimichi Takemura, who pointed out some awkward aspects of an earlier version. The referees were also very helpful. This work was partially supported by EPSRC through grants GR/R46670 and GR/M16856.

References

[1] Dawid, A. P. (1984) Statistical theory: The prequential approach (with discussion). *J. R. Statist. Soc. A*, **147**, 278–292.

[2] Dawid, A. P. (1986) Probability forecasting. *Encyclopaedia of Statistical Sciences* (eds S. Kotz, N. L. Johnson, and C. B. Read), **7**, 210–218. New York: Wiley.

[3] Foster, D. P. and Vohra, R. V. (1998) Asymptotic calibration. *Biometrika*, **85**, 379–390.

[4] Fudenberg, D. and Levine, D. K. (1999) An easier way to calibrate. *Games and Economic Behavior*, **29**, 131–137.

[5] Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.

[6] Kolmogorov, A. N. (1983) On logical foundations of probability theory. *Probability Theory and Mathematical Statistics; Proc. 4th USSR-Japan Symp., 1982* (eds K. Itô and J. V. Prokhorov), *Lecture Notes in Mathematics*, **1021**, 1–6. Berlin: Springer.

[7] Lehrer, E. (2001) Any inspection is manipulable. *Econometrica*, **69**, 1333-1347.

[8] Li, M. and Vitányi, P. (1997) *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd edn. New York: Springer.

[9] Martin-Löf, P. (1966) The definition of random sequences. *Information and Control*, **9**, 602–619.

[10] Oakes, D. (1985) Self-calibrating priors do not exist (with discussion). *J. Amer. Statist. Assoc.*, **80**, 339–342.

[11] Putnam, H. (1963) “Degree of Confirmation” and inductive logic. *The philosophy of Rudolf Carnap*, (ed P. A. Schilpp), pp. 761–783. La Salle, IL: Open Court.

[12] Sandroni, A. (2003) The reproducible properties of correct forecasts. *Int. J. Game Theory*, **32**, 151–159.

[13] Sandroni, A. Smorodinsky, R. and Vohra. R. V. (2003) Calibration with many checking rules. *Math. Oper. Res.*, **28**, 141–153.

[14] Shafer, G. and Vovk, V. (2001) *Probability and Finance: It's only a Game!* New York: Wiley.

[15] Shafer, G. and Vovk, V. (2004) The origins and legacy of Kolmogorov's *Grundbegriffe*. Working Paper #4, Game-Theoretic Probability and Finance Project, www.probabilityandfinance.com.

[16] Shiryaev, A. N. (1996) *Probability*. 2nd edn. New York: Springer.

[17] Ville, J. (1939) *Étude critique de la notion de collectif*. Paris: Gauthier-Villars.

[18] von Mises, R. (1919) Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Zeitschrift*, **5**, 52–99.

[19] Vovk, V. (2005) Non-asymptotic calibration and resolution. Working Paper #13, www.probabilityandfinance.com.

[20] Vovk, V. and Shafer, G. (2004) Good sequential probability forecasting is always possible. Working Paper #7, www.probabilityandfinance.com.

[21] Vovk, V., Takemura, A. and Shafer, G. (2005) Defensive forecasting. *AISTATS 2005: Proc. 10th Int. Workshop AI and Stat.*

[22] Wald, A. (1937) Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, **8**, 38–72.

A Properly calibrated randomized forecasting

As we explained in the introduction to this article, the work on randomized forecasting by Foster and Vohra (1998), Fudenberg and Levine (1999), Lehrer (2001), and Sandroni et al. (2003) demonstrated only the existence of randomized forecasts with certain calibration properties. Foster and Vohra showed that the whole sequence of forecasts can be made properly calibrated, and the other authors showed that subsequences of forecasts selected by certain rules can also be made properly calibrated. In this appendix, we show that our Theorem 3, together with a game-theoretic strong law of large numbers that we proved in Shafer and Vovk (2001), implies the existence of randomized forecasts that are properly calibrated even with respect to the widest possible class of rules for selecting subsequences. For brevity, we continue to consider only the binary case.

Selecting subsequences can be more complicated in probability forecasting than in von Mises's theory, because we can use $p_1x_1 \dots p_{n-1}x_{n-1}p_n$, not merely $x_1 \dots x_{n-1}$, when deciding whether to include the nth trial in the subsequence. Among our predecessors, however, only Sandroni et al. went beyond $x_1 \dots x_{n-1}$, and they used only $x_1 \dots x_{n-1}$ and p_n, still ignoring the prior forecasts $p_1, \dots, p_{n-1}$. We will be as broad as possible, allowing rules that use all the internal information $p_1x_1 \dots p_{n-1}x_{n-1}p_n$.

Let us write $\mathcal{U}$ for the set of all sequences of the form $p_1x_1 \dots p_{n-1}x_{n-1}p_n$. We call any measurable function $F : \mathcal{U} \to \{0, 1\}$ a *selection rule*; we interpret it by including the nth round in the subsequence when $F(p_1x_1 \dots p_{n-1}x_{n-1}p_n) = 1$. We say that an infinite sequence of forecasts $p_1p_2 \dots$ is *properly calibrated* with respect to a selection rule F on a path $x_1x_2 \dots$ if

$$\sum_{n=1}^{\infty} F(p_1x_1 \dots p_{n-1}x_{n-1}p_n) < \infty \tag{5}$$

or

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} F(p_1x_1 \dots p_{i-1}x_{i-1}p_i)(x_i - p_i)}{\sum_{i=1}^{n} F(p_1x_1 \dots p_{i-1}x_{i-1}p_i)} = 0. \tag{6}$$

Let us write $\mathcal{V}$ for the set of all sequences of the form $p_1x_1 \dots p_nx_n$. A *forecasting system* is a measurable function $\zeta : \mathcal{V} \to \mathcal{P}[0, 1]$. Given a path $s = x_1x_2 \dots$, we define $\overline{\zeta}_s : [0, 1]^* \to \mathcal{P}[0, 1]$ by

$$\overline{\zeta}_s(p_1, \dots, p_{n-1}, p_n) := \zeta(p_1x_1 \dots p_nx_n).$$

By Ionescu-Tulcea's extension theorem (Shiryaev 1996, Section II.9), there exists a unique probability measure ζ_s^* on $[0, 1]^\infty$ having $\overline{\zeta}_s(p_1, \dots, p_n)$ as a conditional distribution for p_{n+1} given $p_1, \dots, p_n$ for $n = 0, 1, \dots$. We say that ζ is *properly calibrated* with respect to a selection rule F on a path s if ζ_s^*-almost every forecast sequence $p_1p_2 \dots$ is properly calibrated with respect to F on s; we say that ζ is *properly calibrated* with respect to F if it is properly calibrated with respect to F on every path s.

The following theorem, which we prove using Theorem 3, can also be proven using the methods of Sandroni et al. Technically, it is stronger than Proposition 1 of Sandroni et al., because they use, unnecessarily, a notion of selection rule that is less general than ours.

Theorem 5 *Given an arbitrary countable collection of selection rules, there exists a forecasting scheme that is properly calibrated with respect to all the rules in the collection.*

Proof Let $\mathcal{C}$ be a countable collection of selection rules. For each selection rule F in $\mathcal{C}$, fix a strategy $\mathcal{R}_F$ for Sceptic in Binary Forecasting Game I that is legal and makes him infinitely rich if neither (5) nor (6) holds. Such a strategy $\mathcal{R}_F$ can be constructed in an almost trivial way from the winning strategy $\mathcal{R}_{\text{SLLN}}$ for Sceptic we constructed in Section 3.3 of Shafer and Vovk (2001) to prove the game-theoretic strong law of large numbers in the bounded forecasting game studied there; we simply ignore any round n in Binary Forecasting Game I for which $F(p_1x_1 \dots p_{n-1}x_{n-1}p_n) = 0$. More precisely, $\mathcal{R}_F$ tells Sceptic to set M_n equal to zero whenever $F(p_1x_1 \dots p_{n-1}x_{n-1}p_n) = 0$ and to use $\mathcal{R}_{\text{SLLN}}$'s recommendation for round k of the bounded forecasting game on the round of Binary Forecasting Game I for which $F(p_1x_1 \dots p_{n-1}x_{n-1}p_n) \neq 0$ for the kth time.

By Lemmas 3.1 and 3.2 of Shafer and Vovk (2001), Sceptic can average the $\mathcal{R}_F$ for $F \in \mathcal{C}$ to obtain a legal strategy $\mathcal{R}$ in Binary Forecasting Game I that makes him infinitely rich whenever any of the $\mathcal{R}_F$ does so. This strategy $\mathcal{R}$ makes him infinitely rich if there is any F in $\mathcal{C}$ for which neither (5) nor (6) holds. It can be chosen measurable. Since it is known in advance, it can be translated into Sceptic's strategy in Binary Forecasting Game II. This makes Forecaster's winning strategy in Binary Forecasting Game II (which exists by Theorem 1) a function of only Reality's and Random Number Generator's moves—i.e., a forecasting system. Call it ζ.

To complete the proof, we need to check that ζ is properly calibrated with respect to all $F \in \mathcal{C}$ on all $s = x_1x_2 \dots$. In other words, for each s and each F, we need to show that ζ_s^*-almost all forecast sequences $p_1, p_2, \dots$ are properly calibrated with respect to F on s. But lack of proper calibration leads to Sceptic becoming infinitely rich, and by Theorem 3, this leads to Forecaster also becoming infinitely rich. And since Forecaster's capital is a supermartingale with respect to ζ_s^*, Forecaster becomes infinitely rich with probability zero. ∎

B Forecasting under stochasticity

Suppose we fix a horizon N and a function $T(p_1, x_1, \dots, p_N, x_N)$ for testing the probability forecasts $p_1, \dots, p_N$. The function T takes only two values: "accept" and "reject". Suppose further that under any probability distribution P for $x_1, \dots, x_N$, the probability that T accepts when P's conditional probabilities are used for the p_n is at least $1 - \epsilon$. Sandroni (2003) shows that there is a randomized strategy for giving the p_n that makes the probability that T accepts at least $1 - \epsilon$ no matter how $x_1, \ldots, x_N$ come out. This is a measure-theoretic version of our game-theoretic Theorem 4. It is weaker than Theorem 4 in several respects:

- It assumes a fixed horizon N. (This is a minor point, because the assumption that N is finite can easily be relaxed in Sandroni's approach.)
- It assumes that Reality chooses the $x_1, \ldots, x_N$ randomly rather than playing strategically.
- It fixes at the outset a test T based on all the forecasts and outcomes, whereas our Sceptic can also use other information that arrives as the game proceeds or merely change his strategy on a whim.

But it expresses in measure-theoretic terms the proposition that randomized probability forecasts can perform as well true probabilities.

Typically, measure-theoretic counterparts can be derived from game-theoretic results in probability (Shafer and Vovk 2001, Chapter 8). We now demonstrate that this rule holds in the present case by deriving a simplified version of Sandroni's result from our Theorem 4.

Following Sandroni, we assume that the game is played for only a finite number of rounds, say N. To avoid technicalities, we make two simplifying assumptions:

1. We assume, as we have done throughout this article, that the outcomes x_n are binary. (Sandroni allows any finite outcome space.)
2. We assume that all probabilities are chosen from a fixed finite subset $\mathbf{P}$ of $[0, 1]$. The forecaster is required to choose his p_n from $\mathbf{P}$, and the unknown probability distribution P has all its conditional probabilities (its probabilities for $x_n = 1$ given $x_1, \ldots, x_{n-1}$) in $\mathbf{P}$.

Under these assumptions, our Theorem 4 holds with $\epsilon = 0$.

A *test* T is a function that maps each sequence $(p_1, x_1, \ldots, p_N, x_N)$ with $p_n \in \mathbf{P}$ and $x_n \in \{0, 1\}$ to 0 or 1. We interpret $T = 1$ to mean that the test rejects the p_n (this reverses Sandroni's convention). The test *does not reject the truth with probability* $1 - \epsilon$ if, for any probability distribution P on $\{0, 1\}^N$ with conditional probabilities in $\mathbf{P}$, the P-probability that $T(p_1, x_1, \ldots, p_N, x_N) = 1$, where p_n is the conditional P-probability that $x_n = 1$ given $x_1, \ldots, x_{n-1}$, does not exceed ϵ. The test *can be passed with probability* $1 - \epsilon$ if there exists a forecasting system (in the sense of the preceding appendix but with the p_n restricted to $\mathbf{P}$) ζ such that, for any path $s = x_1, \ldots, x_N$,

$$\zeta_s^* \left\{ (p_1, \ldots, p_N) \in [0, 1]^N \mid T(p_1, x_1, \ldots, p_N, x_N) = 1 \right\} \le \epsilon,$$

where ζ_s^* is defined as in the preceding appendix but with n restricted not to exceed N.

Corollary 1 *If a test does not reject the truth with probability* $1 - \epsilon$ *then it can be passed with probability* $1 - \epsilon$.

Proof Let T be a test that does not reject the truth with probability $1-\epsilon$. We use it to define a martingale $\mathcal{K}$ as follows: define a function $\mathcal{K}'(p_1, x_1, \ldots, p_n, x_n)$, where $n = 0, 1, \ldots, N$, $p_i \in [0,1]$, and $x_i \in \{0,1\}$, by the requirements

$$\mathcal{K}'(p_1, x_1, \ldots, p_N, x_N) := T(p_1, x_1, \ldots, p_N, x_N)$$

and

$$\mathcal{K}'(p_1, x_1, \ldots, p_n, x_n) := \max_{p \in \mathbf{P}} \left(p\mathcal{K}'(p_1, x_1, \ldots, p_n, x_n, p, 1) + (1-p)\mathcal{K}'(p_1, x_1, \ldots, p_n, x_n, p, 0)\right)$$

for $n = N-1, N-2, \ldots, 1, 0$. Since T does not reject the truth with probability $1-\epsilon$, $\mathcal{K}'(\Box) \le \epsilon$. It is easy to see that $\mathcal{K} := \mathcal{K}'/\mathcal{K}'(\Box)$ is a capital process of some strategy in Binary Forecasting Game I in which Sceptic is allowed to throw part of his money away at each trial. Consider the corresponding strategy (i.e., the strategy with the same M_n) in which Sceptic keeps all his money at each trial; this determines Sceptic's strategy in Binary Forecasting Game II ($S_n(p)$ is Sceptic's move in Binary Forecasting Game I made in response to $p_1, x_1, \ldots, p_{n-1}, x_{n-1}, p$). Consider also the randomized strategy for Random Number Generator that draws $p_n \in \mathbf{P}$ randomly from P_n at each trial and the deterministic strategy for Reality that outputs a fixed sequence $s = (x_1, \ldots, x_N)$ of moves. Let ζ be the part of Forecaster's winning strategy that produces P_n (in the game with the goal $\mathcal{F}_n \ge \mathcal{K}_n$ for all n; cf. Theorem 4) as played against these strategies for Sceptic, Random Number Generator, and Reality. Because $\mathcal{F}_n \ge \mathcal{K}_n$, $\mathcal{F}_n$ is a nonnegative supermartingale with respect to ζ_s^*, and it starts at 1. So Doob's inequality implies that ζ_s^* gives the event $\mathcal{F}_N \ge 1/\epsilon$ probability at most ϵ. It follows that ϵ is also an upper bound for the probability that $\mathcal{K}_N \ge 1/\epsilon$ and hence for the probability that $\mathcal{K}'_N \ge 1$ and hence for the probability that $T(p_1, x_1, \ldots, p_N, x_N) = 1$. ∎