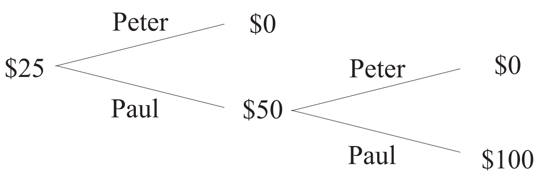
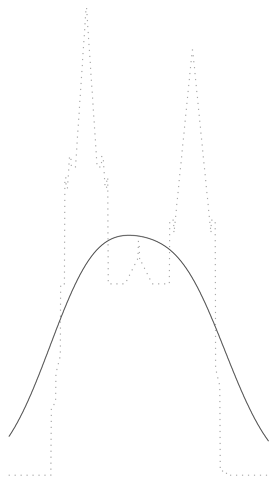
Good sequential probability forecasting is always possible

Vladimir Vovk Glenn Shafer





The Game-Theoretic Probability and Finance Project

Working Paper #7

First posted June 11, 2003. Last revised September 12, 2007.

Project web site:
http://www.probabilityandfinance.com

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.

Contents

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 (5) and (7) below. When calibration fails, whether on the entire sequence or on a subsequence, a statistical test has been failed. But there are other statistical tests that go beyond calibration. We can check, for example, whether the convergence required by calibration is at the rate required the law of the iterated logarithm (Ville 1939).[1] In the measure-theoretic framework, violation of the law of the iterated logarithm is an event of

[1] A referee has asked whether a violation of the law of the iterated logarithm can be detected in finite time. It can be, in the same sense in which a failure of calibration can be detected in finite time (Kolmogorov 1929; Shafer and Vovk 2001, p. 120). But longer sequences of probability forecasts are needed to detect violations of the law of the iterated logarithm than to detect violations of calibration.

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, 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 alone 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, 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 its relative complexity should be seen as another deficiency of von Mises's picture. As we show here, the story is simpler in our game-theoretic framework. For more on the historical background of the game-theoretic framework, see Shafer and Vovk (2004), Appendix C, and Chapter 2 of Shafer and Vovk (2001).

An abridged version of this working paper, omitting in particular all the appendices, was published in 2005 as "Good randomized sequential probability forecasting is always possible," *Journal of the Royal Statistical Society, Series B*, volume 67, pp. 747–764. This working paper remains useful, however, for its more thorough exposition. In addition to the appendices, it includes an informative additional result, Theorem 5, added in 2007.

In 2004, before the published version appeared, we learned that randomization is not needed in the quite common case where the properties we want for the forecasts can be enforced by a strategy for the sceptic that is continuous in the forecast. The proof for the continuous case is remarkably simple [41]. Moreover, Vovk has recently shown that Theorem 3 of this paper can be derived naturally from this continuous case [40]. We now call the whole approach, whether or not randomization is used, *defensive forecasting*. For more information, including extensions beyond the binary case, see the whole series of working papers on defensive forecasting at www.probabilityandfinance.com.

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 $n = 1, 2, \dots$:
    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,

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

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 3.1.) 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.

Theorem 2 says that Sceptic and Reality can defeat Forecaster in Binary Forecasting Game I, where they have full knowledge of Forecaster's moves and no constraints on their own moves. But the strategies for Sceptic and Reality used to prove Theorem 2 are rather delicate: as functions of Forecaster's move p_n, they change discontinuously at $p_n = 1/2$. It turns out that Forecaster is in a much stronger position if the protocol is changed even slightly to prevent this delicate and very precise collaboration between Sceptic and Reality. There are at least two different ways to do this:

- We can constrain Sceptic to use a strategy continuous in p_n. As we mentioned in Section 1, we have shown in work subsequent to the original posting of this working paper that this constraint suffices to give Forecaster a winning strategy. This is explained in [41] and at more length in [31].

- We can restrict Reality's knowledge of p_n slightly, so that she may be unable to tell, for example, whether or not it exceeds $1/2$. We can do this by having Forecaster announce not p_n itself but a probability distribution P_n, possibly very concentrated about a single point, from which p_n will subsequently be drawn. This idea was used by several authors in the 1990s and more recently; see Sandroni et al. (2003).

The purpose of this section is to formulate a game, which we call Binary Forecasting Game II, in which Forecaster gives only a probability distribution P_n for p_n. We give the protocol for this game in Subsection 3.3. As we see there, Forecaster's winning means his winning almost surely with respect to the probabilities involved in the randomization. We show in Section 4 that Forecaster does have a winning strategy.

We begin this section, in Subsection 3.1, by explaining abstractly how the idea of drawing p_n from a probability distribution can be expressed game-theoretically. The game we use for the explanation can be thought of as a subgame of Binary Forecasting Game II.

Binary Forecasting Game II also incorporates the idea that Sceptic can combine a number of strategies he might want to use into a single strategy, which makes him infinitely rich if any of the individual strategies do. We explain this in Subsection 3.2. This explanation also involves a game that can be thought of as a subgame of Binary Forecasting Game II.

Because we will not use the subgames developed in Subsections 3.1 and 3.2 in isolation, some readers may prefer to skip these subsections, at least initially, turning directly to the protocol for Binary Forecasting Game II in Subsection 3.3.

3.1 Allowing Forecaster to randomize

In our purely game-theoretic approach, the notion that Forecaster's move is selected at random from a probability distribution he announces 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, and writing $\mathcal{P}[0,1]$ for the set of all probability measures on $[0,1]$, 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 $\int f_n dP_n \le 0$.
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 ($\mathcal{F}_n \ge 0$) no matter what value Random Number Generator announces for p_n.
Winner: Forecaster wins if his capital $\mathcal{F}_n$ tends to infinity.

In this game, Forecaster 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.

Random Number Generator can easily defeat Forecaster. We will assume that he does so, playing so that $\mathcal{F}_n$ stays bounded. In other words, we will adopt Cournot's principle for this Randomization Subgame. This assumption will be implicit in our formulation of Binary Forecasting Game II (Subsection 3.3), inasmuch as that game requires the forecasts p_n to defeat Sceptic and Reality only when $\mathcal{F}_n$ is bounded.

The protocol of the Randomization Subgame implicitly assumes that f_n is 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 piecewise linear.

3.2 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 (2004).

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, except that here we have Forecaster announcing p_n and there Random Number Generator will do so, 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 a function of Forecaster's forthcoming move.

Forecasting Subgame
Players: Sceptic, Forecaster, Reality
Protocol:

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Sceptic announces a 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. The game does not require S_n to be continuous or even measurable. The requirement that $\mathcal{K}_n$ remain nonnegative no matter how Forecaster and Reality move does, however, mean that S_n must be a bounded function of p.

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 $n = 1, 2, \dots$:
Sceptic announces a 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 rules of this game impose only regularity conditions on S_n or f_n. The requirement that $\mathcal{K}_n \geq 0$ no matter how the other players move does imply a bound on each S_n, and the requirement $\int f_n dP_n \leq 0$ does imply that f_n must be integrable and therefore measurable. But this is about all that can be said.

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 also ensure the neutrality of Reality by requiring her to choose her entire sequence of moves

$x_1x_2\ldots$ 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

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

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

$$f_n(p) := \frac{1}{1+\epsilon}\left(S_n(p)(x_n - p) - \epsilon 2^{-n}\right) \tag{4}$$

for $p \in A_n$ and defined arbitrarily for $p \notin A_n$. (This allows Forecaster to make f_n continuous and piecewise linear if he wishes.) The condition $\int f_n dP_n \leq 0$ 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.

A more careful look at the proof of Theorem 3 allows one to strengthen Theorem 4 as follows.

Theorem 5 *For any $\epsilon > 0$ and any sequence $\epsilon_1, \epsilon_2, \dots$ of positive reals, Forecaster has a strategy in Binary Forecasting Game II that guarantees:*

- $\mathcal{K}_n \le (1+\epsilon)\mathcal{F}_n$ *for each n;*
- *each P_n is concentrated on a set $\{p'_n, p''_n\} \subseteq [0,1]$ containing at most two elements (we allow $p'_n = p''_n$) such that $|p'_n - p''_n| < \epsilon_n$.*

The condition $|p'_n - p''_n| < \epsilon_n$ shows that a tiny amount of randomization is sufficient; a similar observation was made by Kakade and Foster (2004).

Proof Fix n. We are required to show that Forecaster can achieve his goal (3) using P_n concentrated on $\{p', p''\} \subseteq [0,1]$ with $|p' - p''| < \epsilon_n$. Consider the curve C consisting of the points

$$C_p := (S_n(p)(-p), S_n(p)(1-p)), \quad p \in [0,1],$$

in the (x,y)-plane $\mathbb{R}^2$. Inequality (3) requires that a convex combination of points on this curve be southwest of the point $(\epsilon 2^{-n}, \epsilon 2^{-n})$ (where (x_1, y_1) being southwest of (x_2, y_2) means that $x_1 \le x_2$ and $y_1 \le y_2$). We will see that a convex mixture of $C_{p'}$ and $C_{p''}$ with arbitrarily close p' and p'' can be made arbitrarily close to the origin $(0,0)$ (except in trivial cases); this will establish the theorem.

We may assume that C does not pass through the origin. (If it does, we can set $p' = p'' := p$ for $C_p = (0,0)$.)

Clearly, C lies in the second ($x \le 0, y \ge 0$) and fourth ($x \ge 0, y \le 0$) quadrants of the plane. For each $p \in [0,1]$ let L_p be the straight line passing through the origin and the point $(-p, 1-p)$ (so that C_p is the only point of the intersection of C and L_p). Let A be the set of p for which C_p is in the second quadrant, and let B be the set of p for which C_p is in the fourth quadrant. Without loss of generality we assume that the sets A and B are non-empty (if A is empty, we can set $p' = p'' := 0$, and if B is empty, we can set $p' = p'' := 1$). Since $A \cup B = [0,1]$ and the set $[0,1]$ is connected, the closures of A and B cannot be disjoint.

Let $p' \in [0,1]$ be any point in the intersection of the closures of A and B. The line $L_{p'}$ contains a point of C either in the second or in the fourth quadrant; suppose, for concreteness, that $C_{p'}$ is in the second quadrant. There are C_ps in the fourth quadrant with p arbitrarily close to p'. We can take one of them as p''.

Forecaster's strategy must also include a rule for selecting f_n. We retain the rule indicated in the proof of Theorem 3: the support of the probability measure P_n now consists of the two points $p = p'$ and $p = p''$, so we define $f_n(p)$ by (4) and extend it to $[0,1]$ by making it piecewise linear and therefore integrable. ∎

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[2]:

- Against Forecaster, Random Number Generator is playing the Randomization Subgame we described in Section 3.1. 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).
- 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. 10), 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).

Theorem 4 says that Forecaster has a strategy that guarantees $\mathcal{K}_n \leq (1+\epsilon)\mathcal{F}_n$ for all n. 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 $x_n \in \{0,1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$.
  $\mathcal{F}_n := \mathcal{F}_{n-1} + f_n^{x_n}(p_n)$.

[2]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.

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 $x_1 x_2 \dots$ 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

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} (x_i - p_i)}{n} = 0. \qquad (5)$$

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 (Appendix E; 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 of the journal version (*JRSSB* 2005) were also very helpful. Sasha Shen noticed in August 2007 that Forecaster's moves P_n in Theorems 3 and 4 can be taken concentrated on a two-element set, which has led to Theorem 5. This work was partially supported by EPSRC through grants GR/R46670, GR/M16856, and EP/F002998/1.

References

[1] David Blackwell and Lester E. Dubins. Merging of opinions with increasing information. *Annals of Mathematical Statistics*, 33:882–886, 1962.

[2] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

[3] Alonzo Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46:130–135, 1940.

[4] J. H. Curtiss. An elementary mathematical model for the interpretation of precipitation probability forecasts. *Journal of Applied Meteorology*, 7:3–17, 1968.

[5] A. Philip Dawid. Statistical theory: The prequential approach (with discussion). *Journal of the Royal Statistical Society. Series A*, 147:278–292, 1984.

[6] A. Philip Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13:1251–1285, 1985.

[7] A. Philip Dawid. Probability forecasting. In S. Kotz, N. L. Johnson, and C. B. Read, editors, *Encyclopedia of Statistical Sciences*, volume 7, pp. 210–218. Wiley, New York, 1986.

[8] Bruno de Finetti. *Compte rendu critique du colloque de Genève sur la théorie des probabilités*. Number 766 in *Actualités Scientifiques et Industrielles*. Hermann, Paris, 1939. This is the eighth fascicle of [43]. It summarizes a colloquium held in Geneva in 1937, and it includes additional comments that de Finetti subsequently solicited from some of the participants.

[9] Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953. Classical monograph that popularized martingales.

[10] Dean P. Foster and Rakesh V. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[11] Drew Fudenberg and David K. Levine. An easier way to calibrate. *Games and Economic Behavior*, 29:131–137, 1999.

[12] Yuri M. Kabanov, Robert Sh. Liptser, and Albert N. Shiryaev. On the question of absolute continuity and singularity of probability measures. *Mathematics of the USSR Sbornik*, 33:203-221, 1977.

[13] Sham M. Kakade and Dean P. Foster. Deterministic calibration and Nash equilibrium. In John Shawe-Taylor and Yoram Singer, editors, *Proceedings of the Seventeenth Annual Conference on Learning Theory*, volume 3120 of *Lecture Notes in Computer Science*, pages 33–48, Heidelberg, 2004. Springer.

[14] Andrei N. Kolmogorov. Über das Gesetz der iterierten Logarithmus. *Math. Annalen*, 101:126–135, 1929.

[15] Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965.

[16] Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions of Information Theory*, IT-14:662–664, 1968.

[17] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.

[18] Andrei N. Kolmogorov. On logical foundations of probability theory. *Probability Theory and Mathematical Statistics; Fourth USSR-Japan Symposium Proceedings, 1982*, K. Itô and Yu. V. Prokhorov, Eds., vol. 1021 of *Lecture Notes in Mathematics*, pp. 1–6. Springer, Berlin. 1983. A footnote states, "The report is recorded by Novikov A. A., Zvonkin A. K., Shen' A."

[19] Ehud Lehrer. Any inspection is manipulable. *Econometrica*, 69:1333-1347, 2001.

[20] Leonid A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14:1413, 1973.

[21] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York, 2nd edition, 1997.

[22] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[23] Per Martin-Löf. The literature on von Mises' Kollectivs revisited. *Theoria*, 35:12–37, 1969.

[24] Per Martin-Löf. Personal communication with Vladimir Vovk, 2003.

[25] David Oakes. Self-calibrating priors do not exist (with discussion). *Journal of the American Statistical Association*, 80:339–342, 1985.

[26] Hilary Putnam. "Degree of Confirmation" and inductive logic. *The philosophy of Rudolf Carnap*, P. A. Schilpp, Ed., vol. 11 of *The Library of Living Philosophers*, Chapter 24, pp. 761–783. Open Court, La Salle, IL. 1963.

[27] Alvaro Sandroni. The reproducible properties of correct forecasts. *Int. J. Game Theory*, 32:151–159, 2003.

[28] Alvaro Sandroni, Rann Smorodinsky, and Rakesh V. Vohra. Calibration with many checking rules. *Mathematics of Operations Research*, 28:141–153, 2003.

[29] Claus Peter Schnorr. Über die Definition von effektiven Zufallstests. Parts I and II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 15:297–312 and 313–328, 1970.

[30] Claus Peter Schnorr. *Zufälligkeit und Wahrscheinlichkeit: Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*, volume 218 of *Lecture Notes in Mathematics*. Springer, Berlin, 1971.

[31] Glenn Shafer. Game-theoretic probability and its uses, especially defensive forecasting. Working Paper #21, Game-Theoretic Probability and Finance Project, www.probabilityandfinance.com, 2007.

[32] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's only a Game!* Wiley, New York, 2001. See www.probabilityandfinance.com for reviews and sample chapters.

[33] Glenn Shafer and Vladimir Vovk. The origins and legacy of Kolmogorov's *Grundbegriffe*. Working Paper #4, Game-Theoretic Probability and Finance Project, www.probabilityandfinance.com, 2004.

[34] Albert N. Shiryaev. *Probability.* Springer, New York, second edition, 1996.

[35] Jean Ville. *Étude critique de la notion de collectif.* Gauthier-Villars, Paris, 1939.

[36] Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

[37] Vladimir Vovk. On a randomness criterion. *Soviet Mathematics Doklady*, 35:656–660, 1987.

[38] Vladimir Vovk. A logic of probability, with applications to the foundations of statistics (with discussion). *Journal of the Royal Statistical Society. Series B*, 55:317–351, 1993.

[39] Vladimir Vovk. Non-asymptotic calibration and resolution. Working Paper #11, www.probabilityandfinance.com, 2004.

[40] Vladimir Vovk. Continuous and randomized defensive forecasting: a unified view. Working Paper #21, www.probabilityandfinance.com, 2007.

[41] Vladimir Vovk, Akimichi Takemura, and Glenn Shafer. Defensive forecasting. *AISTATS 2005: Proceedings of the Tenth International Workshop AI and Statistics*, 2005.

[42] Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937.

[43] Rolin Wavre. *Colloque consacré à la théorie des probabilités.* Hermann, Paris, 1938–1939. This celebrated colloquium was held in October 1937 at the University of Geneva, as part of a series (*Conférences internationales des sciences mathématiques*) that began in 1933. The proceedings of the colloquium were published by Hermann in eight fascicles of 50 to 100 pages each, in their series *Actualités Scientifiques et Industrielles.*

[44] Aleksandr K. Zvonkin and Leonid A. Levin. The complexity of finite objects and the algorithmic concepts of information and randomness. *Russian Mathematical Surveys*, 25:83–124, 1970.

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 $p_1 x_1 \dots p_{n-1} x_{n-1} p_n$, 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 $p_1 x_1 \dots p_{n-1} x_{n-1} p_n$.

Let us write $\mathcal{U}$ for the set of all sequences of the form $p_1 x_1 \dots p_{n-1} x_{n-1} p_n$. 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 $F(p_1 x_1 \dots p_{n-1} x_{n-1} p_n) = 1$. We say that an infinite sequence of forecasts $p_1 p_2 \dots$ is *properly calibrated* with respect to a selection rule F on a path $x_1 x_2 \dots$ if

$$\sum_{n=1}^{\infty} F(p_1 x_1 \dots p_{n-1} x_{n-1} p_n) < \infty \tag{6}$$

or

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} F(p_1 x_1 \dots p_{i-1} x_{i-1} p_i)(x_i - p_i)}{\sum_{i=1}^{n} F(p_1 x_1 \dots p_{i-1} x_{i-1} p_i)} = 0. \tag{7}$$

Let us write $\mathcal{V}$ for the set of all sequences of the form $p_1 x_1 \dots p_n x_n$. A *forecasting system* is a measurable function $\zeta : \mathcal{V} \to \mathcal{P}[0, 1]$. Given a path $s = x_1 x_2 \dots$, we define $\overline{\zeta}_s : [0, 1]^* \to \mathcal{P}[0, 1]$ by

$$\overline{\zeta}_s(p_1, \dots, p_{n-1}, p_n) := \zeta(p_1 x_1 \dots p_n x_n).$$

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 $p_1 p_2 \dots$ 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.

Because our notion of a selection rule is more general than the notion used by Sandroni et al., the following theorem is stronger than Proposition 1 of Sandroni et al.:

Theorem 6 *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 (6) nor (7) 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 $F(p_1 x_1 \dots p_{n-1} x_{n-1} p_n) = 0$. More precisely, $\mathcal{R}_F$ tells Sceptic to set M_n equal to zero whenever $F(p_1 x_1 \dots p_{n-1} x_{n-1} p_n) = 0$ 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 $F(p_1 x_1 \dots p_{n-1} x_{n-1} p_n) \neq 0$ 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 (6) nor (7) 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 $s = x_1 x_2 \dots$. 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 $x_1, \dots, x_N$ 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.
- It assumes that Reality chooses the $x_1, \dots, x_N$ randomly rather than playing strategically.
- It fixes a single test T at the outset, whereas our Sceptic can change his strategy as the game proceeds.

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 $x_1, \dots, x_{n-1}$) in $\mathbf{P}$.

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

A *test* T is a function that maps each sequence $(p_1, x_1, \dots, p_N, x_N)$ 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 $T(p_1, x_1, \dots, p_N, x_N) = 1$, where p_n is the conditional P-probability that $x_n = 1$ given $x_1, \dots, x_{n-1}$, 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 $s = x_1, \dots, x_N$,

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

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 $\mathcal{K}'(p_1, x_1, \dots, p_n, x_n)$, where $n = 0, 1, \dots, N$, $p_i \in [0,1]$, and $x_i \in \{0,1\}$, by the requirements

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

and

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

for $n = N-1, N-2, \dots, 1, 0$. 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 $p_1, x_1, \dots, p_{n-1}, x_{n-1}, p$). 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 $s = (x_1, \dots, x_N)$ 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 $T(p_1, x_1, \dots, p_N, x_N) = 1$. ∎

C Roots of the game-theoretic framework

In this appendix, we review some of the historical roots of the game-theoretic framework. For a broader account, see [32], Chapter 2, and [33].

C.1 Von Mises's collectives

Richard von Mises was the first to give the hypothesis of the excluded gambling system a role in the foundation of probability. Von Mises thought of a probability as a limiting relative frequency in a sequence of trials, but he recognized that the existence of a limiting frequency does not imply the irregularity needed to exclude a gambling system. In 1919 [36], von Mises suggested that this regularity could be guaranteed by requiring that the same limiting frequency be obtained when one selects a subsequence of trials step by step, deciding just before each trial whether it will be included in the subsequence. Von Mises called a sequence of outcomes satisfying this requirement a *Kollectif*.

Von Mises's collectives were widely studied and debated in mathematical circles in the 1930s. Many people pointed out that no sequence could satisfy his condition for every rule for selecting subsequences, while others noted that the condition can be satisfied for any countable collection of such rules. Abraham Wald argued that this is enough, because no formal language can express more than a countable number of rules [42], and Alonzo Church buttressed Wald's argument by pointing out that there are only countably many effectively computable selection rules [3].

C.2 Ville's martingales

A fundamental objection to von Mises's picture was advanced in the late 1930s by Jean Ville [35], who demonstrated that a sequence of outcomes may permit a gambler to get infinitely rich even though its limiting frequencies behave as von Mises required. Von Mises's conditions keep the gambler from getting rich simply by choosing when to bet, but the gambler may still be able to get rich by cleverly varying the amount he bets and which outcome he bets on.

At the time of Ville's work, measure theory in the modern abstract sense had not yet established itself as the standard foundation for mathematical probability, but coin tossing, always the fundamental example of probability, was already understood in terms of Lebesgue measure. For Ville and his contemporaries, the properties required of a binary sequence by "classical probability" were the subsets of $\{0, 1\}^\infty$ that have Lebesgue measure one. Ville was able to show that for every subset E of Lebesgue measure zero, there is a gambling system (a strategy for the gambler) that parlays initial finite wealth into infinite wealth if the sequence of outcomes $x_1 x_2 \dots$ falls in E. So by ruling out gambling systems we can guarantee any classical property, including, for example, the law of the iterated logarithm, which, as Ville showed, can be violated by a sequence that is a collective in von Mises's sense.

Ville called the capital process determined by a strategy for the gambler a *martingale*, and the requirement that the gambler use only his initial finite wealth, without borrowing, means that the martingale must be nonnegative. So we can restate Ville's result by saying that for every subset E of measure zero there is a nonnegative martingale that tends to infinity if $x_1 x_2 \dots$ falls in E. The requirement that nonnegative martingales not tend to infinity may be called the martingale definition of randomness.

Wald and Church's arguments apply to Ville's notion of a gambling system just as well as they apply to von Mises's notion of a subsequence selection rule (see Wald's comments on pp. 15–16 of [8]). Thus there exist sequences of outcomes that resist all the gambling systems that can be formulated in a given formal language. Ville could have reasonably argued that when the language is rich enough, these sequences are the truly random sequences, and he could have tried to make this martingale definition of randomness a foundation for probability theory as von Mises had done with his collectives.

In fact, neither Ville nor any of his contemporaries promoted the martingale definition of randomness, and the game-theoretic perspective that Ville had brought to the foundations of probability was neglected for many decades. There are a host of reasons for this:

- In spite of decades of effort, Von Mises never really produced a simple and attractive reconstruction of the mathematical theory of probability on his proposed new foundation, and his efforts in this direction did not provide much guidance for a similar effort based on Ville's sequences. By 1939, when Ville's book appeared, most students of mathematical probability, including Ville, had concluded that efforts to give a deep foundation to probability should be abandoned in favour of some more superficial axiom system, such as the one proposed by Kolmogorov [17].

- Ville himself misunderstood the implications of Church's argument (see [35], p. 134).
- The advent of Hitler and World War II disrupted the lives of Ville's European colleagues so severely that many who might have explored his ideas further never even became aware of them. Ville himself turned to other topics ([32], p. 198).
- Ville's notion of a martingale was co-opted by the American mathematician Joe Doob, who developed it formally within measure-theoretic probability [9].
- The frequentist interpretation of probability has roots in philosophy and popular culture that has sustained interest in von Mises's ideas. There has been no comparable impetus for the development of a game-theoretic interpretation.

Per Martin-Löf's article, "The literature on von Mises' Kollectivs revisited", published in 1969, thirty years after the appearance of Ville's *Étude critique de la notion de collectif*, was probably the first review of von Mises's work that gave Ville's contribution its due. Even today one sees discussions of von Mises's ideas that make no mention of Ville.

C.3 The revival of game-theoretic probability

In the 1960s, Kolmogorov brought a powerful new idea into the discussion of randomness: algorithmic complexity. Fixing a universal Turing machine, he defined the complexity of a finite sequence as the length of the shortest program that would generate it. A random finite sequence, he suggested, is a maximally complex one [15, 16]. This idea inspired work in many directions by many authors. Here we discuss only a development that can be seen as a step towards our game-theoretic framework: Claus Peter Schnorr's revival, in 1970, of Ville's martingale definition of randomness.

In order to explain something of what Schnorr did, we must mention the definition of randomness for infinite sequences given by Per Martin-Löf in 1966 [22]. Kolmogorov's definition of randomness for finite sequences cannot be applied directly to infinite sequences, but Martin-Löf reformulated it in terms of statistical tests. He showed that a universal object exists in the set of recursively enumerable statistical tests of randomness, and he demonstrated that the critical level of this universal test for a finite sequence x of length n is $n - C(x)$, where $C(x)$ is complexity in Kolmogorov's sense. This means that a finite sequence is random in Kolmogorov's sense to the extent that its randomness is not rejected by the universal test, and so it becomes natural to say that an infinite sequence is random if its randomness is never rejected by the universal test. The sequences that are not random in this sense form a constructively defined null set.

Schnorr's contribution was to show that Martin-Löf's definition of randomness is equivalent to a version of Ville's martingale definition of randomness. He showed that an infinite binary sequence is random in Martin-Löf's sense if and only if every lower semicomputable martingale (including those whose starting point is merely lower semicomputable, not computable) is bounded on the sequence ([30], p. 41). (Schnorr actually preferred a weaker condition, which he considered more constructive: he wanted to say that a sequence is random if no computable martingale grows fast enough on the sequence for its divergence to be detected in a certain constructive sense ([30], p. 77). This condition corresponds to a different notion of constructively defined null set, which originated with L. E. J. Brouwer.)

Because Ville himself did not espouse the martingale definition of probability, it can be argued that Schnorr was the first to do so. But Martin-Löf also played an essential role, not only by defining randomness in terms of statistical tests in 1966, but also by calling attention to Ville's work, of which he had been unaware in 1966 [24], in the historical article he published in 1969. Although it cannot be said that Martin-Löf advocated the martingale definition in this historical article, his comments about Ville were both positive and suggestive, and they may well have inspired Schnorr's attention to the martingale definition.

Leonid Levin independently covered the same ground as Schnorr in work that was not published until 1973 [20]. Levin expressed his results in a language that made their game-theoretic content less obvious (see Appendix D). But in at least one respect his results were more general than Schnorr's. Schnorr, like von Mises and Ville, seems to have been motivated by the problem of providing meaning to probability. A probability p is to be explained in terms of the randomness of a binary sequence of 0s and 1s. More generally, a probability measure P on a space Ω is to be explained in terms of the randomness of a sequence of elements of Ω. Levin, like Martin-Löf, went beyond this problem to consider the randomness of a sequence $x_1 x_2 \ldots$ with respect to a probability measure that is not necessarily a product measure. This is one way—though not the best way, as we will now argue—to approach the problem of evaluating sequential probability forecasts.

C.4 Dawid's prequential principle

The problem of evaluating a sequence of probability forecasts $p_1 p_2 \ldots$ seems to have first been raised in 1950, in weather forecasting [2], where it led to a literature on the topic of calibration (see [7] for a review). Probability forecasts, it was pointed out, should be properly calibrated—not too high or too low—relative to actual frequencies. On days when our probability forecast for rain is 0.2, it should rain approximately 20% of the time. And when we use a subsequence selection rule in the sense of von Mises, we want the same proper calibration on the subsequence: we should see rain approximately 20% of the time.

In the literature on calibration, as elsewhere, people were quicker to recognize the relevance of von Mises's subsequence selection than to recognize the relevance of Ville's gambling systems. Subsequence selection rules were discussed at least as early as 1968 [4]. Phil Dawid first called attention to gambling systems, somewhat in passing, in 1985 ([6], p. 1270).

Dawid also made a far more important contribution to our story. This is his prequential principle, which we see as an indispensable historical step in breaking the link, forged almost without reflection by mathematical probabilists, between probability forecasts and probability measures. Weather forecasters and others who actually make probability forecasts often do not see any such link; they may use all sorts of tricks to make up the probabilities $p_1p_2\dots$ as they go along, but they seldom obtain these numbers by calculating conditional probabilities from a probability measure. But for those schooled in mathematical probability, the idea that p_n should be the conditional probability for x_n given $x_1\dots x_{n-1}$ is almost inescapable.

In [6] and a series of other articles, Dawid noted that any forecasting system that does not use outside information—any rule for choosing p_n based on $x_1\dots x_{n-1}$—actually determines a probability measure. (This theorem is often attributed to Ionescu-Tulcea; see [34], §II.9.) But Dawid saw an analogy between probability forecasting and Bayesian statistics, where the *likelihood principle* says that hypotheses should be evaluated only in terms of the probabilities they assign to what actually happened. In the same spirit, Dawid insisted that one should evaluate not the forecasting system but only the forecasts $p_1, p_2, \dots$ actually made; this is his *prequential principle*.

Although the pictures developed by Ville, Schnorr, and Levin are largely game-theoretic, the game we can see in these pictures is only between Sceptic and Reality. There is no role for a player named Forecaster in the picture, because the forecasts are all specified in advance by a probability measure. By bringing the forecaster into the picture and talking about a probability measure on a sequence as a special kind of strategy for the forecaster, and then by focusing our attention on the forecaster's actual moves rather than on his strategies, Dawid more or less created our game-theoretic framework.

It was, in any case, by combining Schnorr's game-theoretic picture with Dawid's prequential principle that Vovk arrived at the early version of the game-theoretic framework he exposited in 1993 [38]. It is a relatively short step from this version to the general version exposited in our book [32], in which Forecaster can make more general moves.

Once we have the game-theoretic framework, however, we no longer need to talk about the prequential principle, for it holds automatically, at least to the extent we need it. The outcome of the game obviously depends only on the moves actually made by the players, not on strategies they may or may not have been following. This frees us from many complications and problems. We can get started, for example, without discussing whether strategies must be computable. Any strategy that we actually construct for Forecaster or Sceptic will be computable and more, but this is a fact of life, not a condition arbitrarily added to the setup.

D Sceptic's universal strategy

We have already mentioned Kolmogorov's introduction of the concept of universality into the foundations of probability and its further use by Martin-Löf. Kolmogorov defined randomness for a finite sequence in terms of the sequence's complexity relative to a universal Turing machine, and Martin-Löf proved the existence of a universal statistical test for the randomness of both finite and infinite sequences [22]. Schnorr gave a game-theoretic interpretation of Martin-Löf's universal test, and Levin, at about the same time, made clear the existence of an essentially maximal lower semicomputable supermartingale [44].

As we mentioned in §3, our Binary Forecasting Game I also has a universal object: a universal lower semicomputable game-theoretic supermartingale. Its existence follows readily from an argument given by Levin ([21], §4.4). We will not repeat this argument here, but in order to provide some perspective for the reader who is not familiar with Levin's work, we will explain what a universal lower semicomputable game-theoretic supermartingale is and say something about what its existence does and does not mean.

Recall that a game-theoretic martingale is a martingale in Ville's sense—the capital process resulting from a strategy for Sceptic. Given a strategy $\mathcal{P}$ for Sceptic, we obtain the resulting game-theoretic martingale $\mathcal{K}^{\mathcal{P}}$ by setting $\mathcal{K}^{\mathcal{P}}(\Box) := 1$, where $\Box$ is the empty sequence, and

$$\mathcal{K}^{\mathcal{P}}(p_1 x_1 \dots p_n x_n) := \mathcal{K}^{\mathcal{P}}(p_1 x_1 \dots p_{n-1} x_{n-1}) + \mathcal{P}(p_1 x_1 \dots p_{n-1} x_{n-1} p_n)(x_n - p_n). \quad (8)$$

(Cf. Shafer and Vovk, 2001, p. 82.) The concepts of strategy and game-theoretic martingale are largely interchangeable; we can recover the strategy for Sceptic from its game-theoretic martingale, and what we can say in terms of one we can say in terms of the other. One is computable if and only if the other is.

At first glance, we might hope to find a universal computable strategy for Sceptic, one that does as well, in some asymptotic sense, on every path $p_1 x_1 p_2 x_2 \dots$ as any other computable strategy for Sceptic. But it is fairly easy to see that there is no such thing. From any computable strategy $\mathcal{P}$ we can construct a path $p_1 x_1 p_2 x_2 \dots$ and another computable strategy that does infinitely better on that path. The key is to identify a path where $\mathcal{P}$ does poorly. We can do this by setting all the p_n equal to $\frac{1}{2}$ and then choosing the x_n step by step: always choose x_n so that $\mathcal{P}$'s gain, $\mathcal{P}(\frac{1}{2}, x_1, \dots, x_{n-1}, \frac{1}{2})(x_n - \frac{1}{2})$, is either negative or (if we find that $\mathcal{P}(\frac{1}{2}, x_1, \dots, x_{n-1}, \frac{1}{2})$ is close to zero before we find its sign) very small. Having identified the path, we then construct a strategy that does infinitely better than $\mathcal{P}$ on that path by betting all Sceptic's current capital on the winning side at each step.

In order to find a universal object, we must ask for something less than a computable game-theoretic martingale, and it turns out that the appropriate something less is a lower semicomputable game-theoretic supermartingale.

Notice that a process S (a real-valued function on sequences of the form $p_1 x_1 \dots p_n x_n$) qualifies as a game-theoretic martingale if and only if $S(\Box) = 1$

and for each $p_1x_1 \dots p_n$ there exists $M \in \mathbb{R}$ such that

$$S(p_1x_1 \dots p_nx_n) = S(p_1x_1 \dots p_{n-1}x_{n-1}) + M(x_n - p_n) \tag{9}$$

for all x_n. The notion of a *game-theoretic supermartingale* is defined analogously; it is a process S such that $S(\square) = 1$ and for each $p_1x_1 \dots p_n$ there exists $M \in \mathbb{R}$ such that

$$S(p_1x_1 \dots p_nx_n) \leq S(p_1x_1 \dots p_{n-1}x_{n-1}) + M(x_n - p_n) \tag{10}$$

for all x_n. The change from the equality in (9) to the inequality in (10) means, intuitively, that we now allow Sceptic to throw money away.

Recall that a process S is *lower semicomputable* if the relation

$$S(p_1x_1 \dots p_nx_n) > r,$$

where n ranges over natural numbers, $p_i \in [0,1]$, $x_i \in \{0,1\}$, and r ranges over rational (or real) numbers, is positively decidable. In the case of a game-theoretic martingale, being lower semicomputable is the same as being computable, but in the case of a game-theoretic supermartingale, being lower semicomputable is a weaker condition.

Here is the exact sense in which the lower semicomputable game-theoretic supermartingale given by Levin's argument is universal:

Theorem 7 *There exists a lower semicomputable game-theoretic supermartingale U such that for any lower semicomputable game-theoretic supermartingale S there exists a positive constant C such that, for every n and every situation $p_1x_1 \dots p_nx_n$,*

$$U(p_1x_1 \dots p_nx_n) \geq CS(p_1x_1 \dots p_nx_n).$$

The fact that U is a game-theoretic supermartingale means that for each $p_1x_1 \dots p_{n-1}x_{n-1}p_n$ there exists a number $\mathcal{P}(p_1x_1 \dots p_{n-1}x_{n-1}p_n)$ that satisfies

$$U(p_1x_1 \dots p_nx_n) \leq U(p_1x_1 \dots p_{n-1}x_{n-1}) + \mathcal{P}(p_1x_1 \dots p_{n-1}x_{n-1}p_n)(x_n - p_n)$$

for all x_n. If we allow definitions to depend on the axiom of choice, this defines a strategy $\mathcal{P}$ whose capital process is never less than U. But such a strategy $\mathcal{P}$ will be at best computable in the limit.

Although Theorem 7 follows from Levin's argument, it is stronger than the result stated by Levin, because it is prequential: it does not assume a strategy for Forecaster. We can reduce the picture to Levin's picture by fixing a strategy for Forecaster that depends only on Reality's moves. If this strategy, say $p(x_1 \dots x_n)$, is computable, then U reduces to a function of Reality's moves only:

$$V(x_1, \dots, x_n) := U(p(\square), x_1, \dots, p(x_1, \dots, x_{n-1}), x_n).$$

This function is a supermartingale with respect to the probability measure obtained from $p(x_1 \dots x_n)$ by Ionescu-Tulcea's extension theorem ([34], §II.9). The product Vp is Levin's *a priori* semimeasure.

E Jeffreys's law

In 1962 [1], David Blackwell and Lester E. Dubins showed that if two probability measures on for $x_1x_2\dots$ are equivalent (either is absolutely continuous with respect to the other), then they have conditional distributions for the future given the first n outcomes $x_1\dots x_n$ whose variation distance from each other converges to zero almost surely under both measures as n tends to infinity. This is an abstract measure-theoretic result, but it can be interpreted as giving conditions under which two probability measures or probability forecasters can be expected to agree more and more closely, or *merge*, as time goes on. Other theorems that can be interpreted similarly were subsequently proven by Kabanov et al. [12], Dawid [6], and Vovk [37]. Following Dawid ([5], p. 281), we call the general thesis that probability forecasts should merge *Jeffreys's law*.

The method we have studied in this article produces a sequence $P_1P_2\dots$ of probability measures on $[0,1]$ and draws p_n from P_n. We will now show that if there is another good forecaster who makes nonrandomized forecasts p'_n (forecasts p'_n that are known to everyone at the beginning of each round), then the P_n can succeed only if they concentrate around p'_n as we proceed.

To put the result in purely game-theoretic form, we use the following version of the forecasting game considered in §4.1:

Binary Forecasting Game III
Players: Theory, Sceptic, Forecaster, Random Number Generator, Reality
Protocol:

$\mathcal{K}_0 := 1$.
$\mathcal{K}'_0 := 1$.
$\mathcal{F}_0 := 1$.
FOR $n = 1, 2, \dots$:
  Theory announces $p'_n \in [0,1]$.
  Sceptic announces $S_n : [0,1] \to \mathbb{R}$.
  Forecaster announces $P_n \in \mathcal{P}[0,1]$.
  Forecaster announces, for $x = 0$ and $x = 1$,
    $f^x_n : [0,1] \to \mathbb{R}$ such that $\int f^x_n dP_n \le 0$.
  Sceptic announces $S'_n \in \mathbb{R}$.
  Random Number Generator announces $p_n \in [0,1]$.
  Reality announces $x_n \in \{0,1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + S_n(p_n)(x_n - p_n)$.
  $\mathcal{F}_n := \mathcal{F}_{n-1} + f_n(p_n)$.
  $\mathcal{K}'_n := \mathcal{K}'_{n-1} + S'_n(x_n - p'_n)$.

Restriction on Sceptic: Sceptic must choose the S_n and S'_n so that his capital on both accounts is always nonnegative ($\mathcal{K}_n \ge 0$ and $\mathcal{K}'_n \ge 0$ for all n) no matter how the other players move. Moreover, each S_n must be continuous.
Restriction on Forecaster: Forecaster must choose the P_n and f_n so that his capital is always nonnegative ($\mathcal{F}_n \ge 0$ for all n) no matter how the other players move.

Theorem 8 *Sceptic has a strategy in Binary Forecasting Game III such that if* $\mathcal{K}_n$ *and* $\mathcal{K}'_n$ *are bounded, then* $\lim_{n\to\infty}(p_n - p'_n) = 0$.

Proof Let us restate the protocol so that Sceptic's moves also look like probability forecasts:

> $\mathcal{K}_0 := 1$.
> $\mathcal{K}'_0 := 1$.
> $\mathcal{F}_0 := 1$.
> FOR $n = 1, 2, \ldots$:
>   Theory announces $p'_n \in [0,1]$.
>   Sceptic announces $q_n : [0,1] \to [0,1]$.
>   Forecaster announces $P_n \in \mathcal{P}[0,1]$.
>   Forecaster announces, for $x = 0$ and $x = 1$,
> $f^x_n : [0,1] \to \mathbb{R}$ such that $\int f^x_n dP_n \le 0$.
>   Sceptic announces $q'_n \in [0,1]$.
>   Random Number Generator announces $p_n \in [0,1]$.
>   Reality announces $x_n \in \{0,1\}$.
> $\mathcal{K}_n := \mathcal{K}_{n-1} q_n(p_n)/p_n$ if $x_n = 1$.
> $\mathcal{K}_n := \mathcal{K}_{n-1}(1 - q_n(p_n))/(1 - p_n)$ if $x_n = 0$.
> $\mathcal{F}_n := \mathcal{F}_{n-1} + f_n(p_n)$.
> $\mathcal{K}'_n := \mathcal{K}'_{n-1} q'_n/p'_n$ if $x_n = 1$.
> $\mathcal{K}'_n := \mathcal{K}'_{n-1}(1 - q'_n)/(1 - p'_n)$ if $x_n = 0$.

This change is motivated by the fact that a positive martingale with respect to a probability measure P can always be represented as Q/P, where Q is another probability measure. Formally, we can verify that the two protocols give the same capital processes for Sceptic using the relationships

$$q_n(p_n) = \frac{\mathcal{K}_{n-1} + S_n(1-p_n)}{\mathcal{K}_{n-1}} p_n \quad \text{and} \quad q'_n = \frac{\mathcal{K}'_{n-1} + S'_n(1-p'_n)}{\mathcal{K}'_{n-1}} p'_n.$$

In the modified protocol, consider the strategy for Sceptic that tells him to make the same move against both Theory and Forecaster:

$$q'_n = q_n(p_n) := \frac{\sqrt{p_n p'_n}}{\sqrt{p_n p'_n} + \sqrt{(1-p_n)(1-p'_n)}}.$$

This is the normalized geometric mean. By Cauchy's inequality, its denominator does not exceed 1. If both $\mathcal{K}_n$ and $\mathcal{K}'_n$ are bounded, than as $N \to \infty$,

$$\prod_{n=1}^{N} \frac{\sqrt{p_n p'_n}}{\sqrt{p_n p'_n} + \sqrt{(1-p_n)(1-p'_n)}} = O\left(\prod_{n=1}^{N} p_n\right)$$

and

$$\prod_{n=1}^{N} \frac{\sqrt{p_n p'_n}}{\sqrt{p_n p'_n} + \sqrt{(1-p_n)(1-p'_n)}} = O\left(\prod_{n=1}^{N} p'_n\right).$$

Multiplying these equations and taking the square root, we find that

$$\prod_{n=1}^{N} \left(\sqrt{p_n p_n'} + \sqrt{(1-p_n)(1-p_n')} \right)$$

is bounded below by a positive number; this implies

$$\sqrt{p_n p_n'} + \sqrt{(1-p_n)(1-p_n')} \to 1$$

as $n \to \infty$. So $p_n - p_n' \to 0$. ∎

This method of proof can also produce nonasymptotic game-theoretic versions of Jeffreys's law, similar to Theorem 1 in [37].

Game-Theoretic Probability and Finance Project
Defensive Forecasting Subseries Working Papers

7. *Good randomized sequential probability forecasting is always possible*, by Vladimir Vovk and Glenn Shafer, June 2003 (revised August 2007).

8. *Defensive forecasting*, by Vladimir Vovk, Akimichi Takemura, and Glenn Shafer, September 2004 (revised January 2005).

9. *Experiments with the K29 algorithm*, by Vladimir Vovk, October 2004.

10. *Defensive forecasting for linear protocols*, by Vladimir Vovk, Ilia Nouretdinov, Akimichi Takemura, and Glenn Shafer, February 2005 (revised September 2005).

11. *On-line regression competitive with reproducing kernel Hilbert spaces*, by Vladimir Vovk, November 2005 (revised January 2006).

13. *Non-asymptotic calibration and resolution*, by Vladimir Vovk, November 2004 (revised July 2006).

14. *Competitive on-line learning with a convex loss function*, by Vladimir Vovk, May 2005 (revised September 2005).

16. *Competing with wild prediction rules*, by Vladimir Vovk, December 2005 (revised January 2006).

17. *Predictions as statements and decisions*, by Vladimir Vovk, June 2006.

18. *Leading strategies in competitive on-line prediction*, by Vladimir Vovk, August 2007.

20. *Defensive forecasting for optimal prediction with expert advice*, by Vladimir Vovk, August 2007.

21. *Continuous and randomized defensive forecasting: unified view*, by Vladimir Vovk, August 2007.

22. *Game-theoretic probability and its uses, especially defensive forecasting*, by Glenn Shafer, August 2007.