

Mathematical and Computing Sciences

Tokyo Institute of Technology

Wednesday, June 2, 2004

Is everything stochastic?

Glenn Shafer

Rutgers University

Newark, New Jersey, USA

This document has been designed to be viewed two pages at a time—pp. 2-3 together, pp. 4-5 together, and so on. (Select "Continuous - Facing" from the "View" menu in Adobe Acrobat Reader.)

Is everything stochastic?

It is always possible, using randomization, to make sequential probability forecasts that will pass any battery of statistical tests.

We may interpret this by saying that everything not perfectly predictable is stochastic.

Alternatively, we may question whether probability forecasts say anything about the future.

Reference: “Good sequential probability forecasting is always possible”, by Vladimir Vovk and Glenn Shafer. Working Paper #6 at www.probabilityandfinance.com.

Outline of Lecture

1. A game between Forecaster and Reality
What Forecaster can achieve
2. Using martingales to test Forecaster
Adding Skeptic to the game
3. Skeptic's task
Skeptic has a more or less universal strategy
4. Forecaster's task
Reality can always stymie Forecaster, unless Forecaster randomizes
5. What does it mean?
Is Forecaster telling about the future or about the past?

1. Forecaster and Reality

We can divide phenomena into three classes:

1. That which we can predict perfectly. (Our predictions will be confirmed.)
2. That which we can predict probabilistically. (Our probabilities will be confirmed.)
3. That which we cannot predict even probabilistically. (We cannot even give probabilities that will be confirmed.)

The point of this lecture: The third class is empty. We can always give probabilities that will be confirmed. Everything, in this sense, is stochastic.

We are speaking, of course, of repetitive phenomena. Without repetition, it makes no sense to ask whether probabilities are confirmed.

Forecaster's task is to predict a sequence x_1, x_2, \dots , say a sequence of 1s and 0s.

Before each x_n is announced, we give a probability p_n for the event $x_n = 1$.

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

Forecaster announces $p_n \in [0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

We say that p_1, p_2, \dots are *confirmed* if they are not refuted by any statistical test based on the observed x_1, x_2, \dots .

The point of this lecture: We can always give p_1, p_2, \dots that are not refuted by any statistical test.

The game between Forecaster and Reality

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

Forecaster announces $p_n \in [0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

Forecaster can give p_1, p_2, \dots that are not refuted by any statistical test.

Clarifications:

1. The phenomena need not be binary. We assume $x_n \in \{0, 1\}$ only for simplicity.
2. It is not even necessary that Reality's move space be the same on every round.
3. The game is one of perfect information; everyone hears the announcements as they are made.
4. When he is about to announce p_n , Forecaster knows x_1, \dots, x_{n-1} . He may also have other newly acquired information.
5. To be fair to Forecaster, we do not consider statistical tests based on information he does not have.

Suppose Reality plays

$$x_1 = 1$$

$$x_2 = 0$$

...

If Forecaster begins the game with a probability distribution P for x_1, x_2, \dots , then he can set

$$p_1 := P(x_1 = 1)$$

$$p_2 := P(x_2 = 1 | x_1 = 1)$$

$$p_3 := P(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 0)$$

...

Alternatively, if Forecaster begins the game with a probability distribution P for everything he might see as the game proceeds, then he can set

$$p_1 := P(x_1 = 1 | \text{all info before round 1})$$

$$p_2 := P(x_2 = 1 | \text{all info before round 2})$$

$$p_3 := P(x_3 = 1 | \text{all info before round 3})$$

...

But Forecaster is not required to base his moves on an initial probability distribution P .

The game between Forecaster and Reality

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

Forecaster announces $p_n \in [0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

Forecaster is not required to base his moves on an initial probability distribution P .

Moreover, Forecaster's moves in the course of the game do not define a probability distribution for x_1, x_2, \dots

If Reality plays $x_1 = 1$, $x_2 = 0$, and so on, then Forecaster's moves p_1, p_2, \dots can be interpreted as conditional probabilities:

$$p_1 = P(x_1 = 1)$$

$$p_2 = P(x_2 = 1 | x_1 = 1)$$

$$p_3 = P(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 0)$$

...

But these conditional probabilities fall short of defining a probability distribution P for x_1, x_2, \dots . They leave unspecified the conditional probabilities

$$P(x_2 = 1 | x_1 = 0)$$

$$P(x_3 = 1 | x_1 = 0 \ \& \ x_2 = 0)$$

$$P(x_3 = 1 | x_1 = 0 \ \& \ x_2 = 1)$$

$$P(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 1)$$

...

Prequential testing

Fortunately, we can test the probabilities p_1, p_2, \dots statistically (using x_1, x_2, \dots) even though they do not define an entire probability distribution for x_1, x_2, \dots .

This point was articulated in the 1980s by Phil Dawid, who called such testing *prequential*.



A. Philip Dawid, in the foreground, with Glenn Shafer on Lake Huron, during an excursion from the Fields Institute in October 1999.

2. Martingale testing

To test Forecaster, try to get rich betting at Forecaster's probabilities.

Formalize this by adding a third player, Skeptic, who does the testing. Skeptic plays against both Forecaster and Reality, who may cooperate.

Players: Forecaster, Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$. Otherwise Forecaster and Reality win.

- \mathcal{K}_n is Skeptic's capital at the end of the n th round.
- Skeptic wins (refutes Forecaster's probabilities) if $\mathcal{K}_n \rightarrow \infty$.
- Skeptic must work with his limited initial capital ($\mathcal{K}_0 = 1$). He cannot borrow without limit. This forestalls the classical trick of doubling your money every time you lose.
- A strategy for Skeptic (a rule for choosing M_n depending on the other players' previous moves) determines Skeptic's capital as a function of those previous moves.
- The capital process $\mathcal{K}_n(p_1, x_1, \dots, p_n, x_n)$ determined by a strategy for Skeptic is called a *martingale*.

GAME FOR TESTING FORECASTER

Players: Forecaster, Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$. Otherwise Forecaster and Reality win.

Every strategy for Skeptic (every martingale) is a test of Forecaster.

The same idea can be used to test a probability distribution P for x_1, x_2, \dots . Just replace Forecaster with P .

This produces a game with only two players.

GAME FOR TESTING P

Players: Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

$$p_n := P(x_n = 1 | x_1, \dots, x_{n-1}).$$

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$. Otherwise Reality wins.

In 1939, Jean Ville showed that every property of a probability distribution P on $\{0, 1\}^\infty$ can be tested by a martingale.



Ville as student



Ville as professor

Assume, for simplicity, that P gives positive probability to every possible initial sequence x_1, \dots, x_n of 0s and 1s.

GAME FOR TESTING P

Players: Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

$$p_n := P(x_n = 1 | x_1, \dots, x_{n-1}).$$

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$. Otherwise Reality wins.

If $E \subseteq \{0, 1\}^\infty$ and $P(E) = 1$, we call E a *property* of P .

Ville showed that for every property E , there is a strategy for Skeptic that tests it.

Ville's Theorem. For every property E , there is a nonnegative martingale that diverges to infinity if Reality makes E fail.



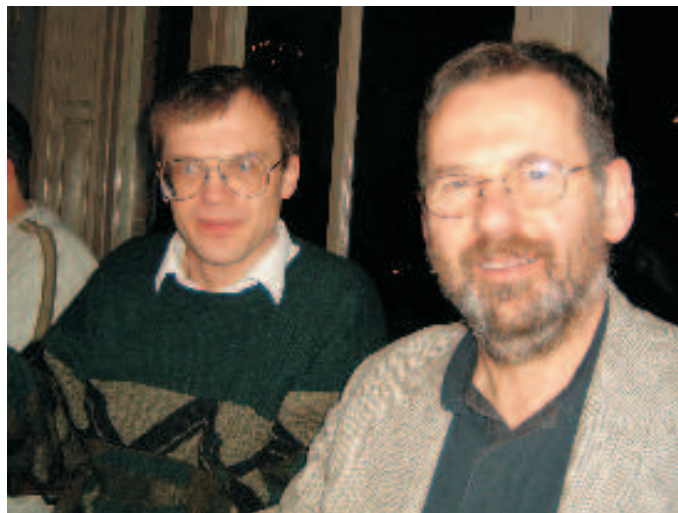
Richard von Mises
(1883–1957) in Berlin
in 1930.

Ville contrasted his martingale testing with von Mises's frequentist testing: select a subsequence of rounds on which to bet.

Von Mises's approach fails to enforce many properties, including the law of the iterated logarithm.

Vladimir Vovk was the first to make Ville's notion of martingale testing prequential.

As Vovk and I demonstrate in *Probability and Finance*, the classical limit theorems generalize from Ville's game (where a distribution P for x_1, x_2, \dots is specified) to the general game, where Forecaster announces his p_n as play proceeds.



Volodya Vovk and Glenn Shafer at Eindhoven, January 2003

GAME FOR TESTING FORECASTER

Players: Forecaster, Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$. Otherwise Forecaster and Reality win.

Example: Vovk's game-theoretic strong law of large numbers.

There is a nonnegative martingale that diverges to infinity unless

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

Example: Strong Law of Large Numbers

Players: Forecaster, Reality, Skeptic

Protocol:

$$\mathcal{K}_0 := 1.$$

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

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and either

$$\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$$

or

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - p_i) = 0$$

Theorem: Skeptic has a winning strategy in this game.

Any event E which has probability 1 in classical probability theory can be substituted for $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - p_i) = 0$.

Technical point

To prove a theorem of this type, it suffices to show that Skeptic has a strategy that makes his capital unbounded when event E fails.

Because Skeptic then has

- a strategy that makes his capital exceed an arbitrary integer 2^k (stop playing when the capital goes past 2^k), and therefore
- a strategy that makes his capital tend to infinity (put 2^{-k} of the \$1 initial capital on the strategy that makes 2^k from 1).

In 1933, Kolmogorov pointed out that probability's classical limit theorems can be proven rigorously within measure theory.

This persuaded mathematicians that probability measures (or distributions) are the proper starting point for probability.

Philosophical consequence:

explaining the meaning
of probability = explaining the meaning
of a probability
distribution



Andrei Kolmogorov (1903–1987) in his dacha at Komarovka

Vovk's demonstration that the classical limit theorems have game-theoretic formulations more general than their measure-theoretic formulations changes this philosophical landscape.

We may now argue that the meaning of probability is game-theoretic.

- **Empirical meaning of probabilities** p_1, p_2, \dots :
An opponent who is allowed to bet against them will not become infinitely rich without risking bankruptcy.
- **Subjective meaning of probabilities** p_1, p_2, \dots :
I believe that an opponent who is allowed to bet against them will not become infinitely rich without risking bankruptcy.



Vladimir Vovk (born 1960 in Ukraine) at home in Windsor, England

For simplicity, this lecture discusses only the infinitary version of the game-theoretic treatment of probability.

- **Empirical meaning of p_1, p_2, \dots :**

An opponent who is allowed to bet against them will not become infinitely rich without risking bankruptcy.

- **Subjective meaning of p_1, p_2, \dots :**

I believe that an opponent who is allowed to bet against them will not become infinitely rich without risking bankruptcy.

But there are also finitary versions.

- **Empirical meaning of p_1, \dots, p_N :**
An opponent who is allowed to bet against them will not multiply his capital by a large factor without risking bankruptcy.
- **Subjective meaning of p_1, \dots, p_N :**
I believe that an opponent who is allowed to bet against them will not multiply his capital by a large factor without risking bankruptcy.

The finitary versions are more meaningful but more complicated, just as the weak law of large numbers is more complicated than the strong law.

3. Skeptic's task

Skeptic's task is not so difficult, because he has a more or less universal strategy.

$$\mathcal{K}_0 := 1.$$

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

Forecaster announces $p_n \in [0, 1]$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in \{0, 1\}$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n).$$

For every property E probabilities should have (law of large numbers, iterated logarithm, etc.), Skeptic has a strategy \mathcal{S} that doesn't risk bankruptcy and makes him infinitely rich if E is violated. By playing \mathcal{S} , Skeptic *forces* Forecaster and Reality to obey E .

Suppose \mathcal{S}_1 forces E_1 and \mathcal{S}_2 forces E_2 . Can Skeptic force both properties by playing both strategies at once?

YES!! If neither \mathcal{S}_1 nor \mathcal{S}_2 risk bankruptcy starting with capital 1, then $\mathcal{S} := (\mathcal{S}_1 + \mathcal{S}_2)/2$ doesn't either, and \mathcal{S} makes Skeptic infinitely rich whenever \mathcal{S}_1 or \mathcal{S}_2 does.

This also works for a countable number of strategies.

So as a practical matter, Skeptic can force validity of probabilities using a single strategy.

Skeptic can reveal his strategy in advance to the other players.

Less radically, he can announce his strategy for a given round at the beginning of that round:

$$\mathcal{K}_0 := 1.$$

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

Skeptic announces $S_n : [0, 1] \mapsto \mathbb{R}$.

Forecaster announces $p_n \in [0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

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

If Forecaster moves p , Skeptic moves $S_n(p)$.

4. Forecaster's task

Players: Forecaster, Reality, Skeptic

Protocol:

$\mathcal{K}_0 := 1.$

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

Skeptic announces $S_n : [0, 1] \mapsto \mathbb{R}.$

Forecaster announces $p_n \in [0, 1].$

Reality announces $x_n \in \{0, 1\}.$

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

Adopting the game-theoretic philosophy just presented, we say that Forecaster's probabilities p_1, p_2, \dots are *valid* if Skeptic does not get infinitely rich.

Can Forecaster produce valid probabilities?

Reality can always stymie Forecaster.

But as we shall show, Forecaster can succeed against Skeptic if Reality is neutral.

To keep Reality neutral, we have Forecaster randomize and hide the result of the randomization from Reality.

Making Forecaster's move partly random

Instead of a probability p_n , Forecaster gives a probability distribution P_n on $[0, 1]$, from which p_n is chosen at random.

The randomization is game-theoretical.

$\mathcal{F}_0 := 1.$

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

Forecaster announces $P_n \in \mathcal{P}[0, 1].$

Forecaster announces $f_n : [0, 1] \rightarrow \mathbb{R}$

such that $\int f_n dP_n \leq 0.$

Random Number Generator announces $p_n \in [0, 1].$

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

We embed the randomization in Forecaster's game with Reality and Skeptic. (Reality sees P_n and f_n but not p_n .)

In the resulting game, Forecaster has a winning strategy.

The complete game

Protocol:

$$\mathcal{K}_0 := 1.$$

$$\mathcal{F}_0 := 1.$$

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

Skeptic announces continuous $S_n : [0, 1] \rightarrow \mathbb{R}$.

Forecaster announces $P_n \in \mathcal{P}[0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

Forecaster announces $f_n : [0, 1] \rightarrow \mathbb{R}$

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 Skeptic: Keep all $\mathcal{K}_n \geq 0$.

Restriction on Forecaster: Keep all $\mathcal{F}_n \geq 0$.

Winner: Forecaster wins if either (i) \mathcal{F}_n tends to infinity or (ii) \mathcal{K}_n stays bounded.

Theorem: Forecaster has a winning strategy.

In fact, Forecaster can guarantee $\mathcal{F}_n \geq \mathcal{K}_n$.

- As usual, we assume perfect information; all players see each others' moves.
- Reality makes her move x_n before seeing p_n .
- Random Number Generator should ignore everything except P_n and just try to make p_n a random draw from P_n .
- If Random Number Generator succeeds, \mathcal{F}_n should stay bounded.
- So by playing his winning strategy, Forecaster keeps \mathcal{K}_n bounded, vindicating the forecasts p_1, p_2, \dots
- Because Random Number Generator ignores everything except P_n when producing p_n , he can produce it and show it to people outside the game before Reality announces x_n . So for the outside world, p_n a probability forecast for x_n .
- Caveat: The theorem depends on Forecaster seeing x_n before he chooses f_n .

Proof of the theorem.

Suppose the following zero-sum game is played right after Skeptic's move S_n :

Protocol:

Simultaneously:

Forecaster chooses $p_n \in [0, 1]$.

Reality chooses $x_n \in \{0, 1\}$.

Payoffs: Forecaster loses $S_n(p_n)(x_n - p_n)$ to Reality.

A mixed strategy for Reality is a measure Q on $\{0, 1\}$. For any such Q , Forecaster can limit Reality's expected gain to 0 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 set $[0, 1]$ with a finite subset of $[0, 1]$. For each natural number i , we choose a finite subset dense enough in $[0, 1]$ that the value of the game is smaller than 2^{-i} . This is possible, because Skeptic's not being able to risk bankruptcy puts a bound on $|S_n(p_n)|$.

The minimax theorem then tells us that Forecaster has a mixed strategy $P \in \mathcal{P}[0, 1]$ such that

$$\int S_n(p)(x - p)P(dp) \leq 2^{-i} \quad (1)$$

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

Designate by E_i the subset of $\mathcal{P}[0, 1]$ consisting of all probability measures P satisfying (1). Endowed with the weak topology, $\mathcal{P}[0, 1]$ is compact. Since each E_i is closed, $\bigcap_i E_i$ is nonempty. So there exists $P_n \in \mathcal{P}[0, 1]$ such that

$$\int S_n(p)(x - p)P_n(dp) \leq 0$$

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

Returning now to the game with Skeptic, consider the strategy for Forecaster that tells him, on round n , to use the P_n just identified and

$$f_n(p) := S_n(p)(x_n - p).$$

This makes $\mathcal{F}_n = \mathcal{K}_n$.

A variation on the game

Protocol:

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

Skeptic announces $S_n : [0, 1] \rightarrow \mathbb{R}$.

Forecaster announces $P_n \in \mathcal{P}[0, 1]$.

Forecaster announces, for $x = 0$ and $x = 1$,

$f_n^x : [0, 1] \rightarrow \mathbb{R}$ such that $\int f_n^x dP_n \leq 0$.

Random Number Generator announces $p_n \in [0, 1]$.

Reality announces $x_n \in \{0, 1\}$.

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

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

Restriction on Skeptic: Keep all $\mathcal{K}_n \geq 0$.

Restriction on Forecaster: Keep all $\mathcal{F}_n \geq 0$.

Winner: Forecaster wins if either (i) \mathcal{F}_n tends to infinity or (ii) \mathcal{K}_n stays bounded.

Theorem: Forecaster has a winning strategy.

In fact, Forecaster can guarantee $\mathcal{F}_n \geq \mathcal{K}_n$.

- In this variation, Reality's move comes last, where it intuitively belongs.
- Because Forecaster's strategy for f_n depends on x_n , he announces in advance what he will do depending on x_n .
- Forecaster wins with the same strategy as in the preceding game. Because it is a winning strategy —wins no matter what the other players do—it makes no difference if he announces some of it in advance and changes the order in which other players move.
- Now Reality's neutrality is guaranteed by the assumption that she will not act so as to discredit the randomization. (Equivalently, we assume that effective randomization is possible regardless of Reality's choice of x_1, x_2, \dots .)

5. What does it mean?

What information is Forecaster using?

All players may have information from outside the game. Forecaster and Skeptic may have different information, but Forecaster may gain some insight into Skeptic's information when he sees S_n .

What is P_n telling us?

Forecaster presumably chooses P_n so as to bias against continuing any past regularity in the relation between p and x that might be continued with $x_n = 0$ but also so to bias against continuing any past regularity in the relation between p and x that might be continued with $x_n = 1$. So perhaps P_n is telling us more about the past than about the future.

Does p_n predict the future?

As a product of randomization, p_n cannot be taken too seriously as a prediction.

What do these results tell us about stochasticity?

Some mathematicians think stochasticity has an objective meaning, independent of anyone's knowledge. The results here suggest that probabilities mean something only if someone knows them. Otherwise, you can make them up.

The idea that valid probability forecasts can always be achieved by randomization was apparently first advanced by Dean P. Foster and Rakesh V. Vohra, in an article finally published in *Biometrika* in 1997, after circulating for many years. Its publication was apparently hindered by disbelief.



Dean Foster



Rakesh Vohra

Our result goes beyond the results of Foster and Vohra and subsequent authors, because we use the full game-theoretic concept of validity. Foster and Vohra had a more limited concept of validity—calibration. The contrast is analogous to the contrast between von Mises's frequentist interpretation of probability and Ville's more powerful game-theoretic interpretation.