Is everything stochastic?

Glenn Shafer
Rutgers University

Cournot Centre – 13 October 2010

1. The question
2. The game
3. Hilary Putnam's counterexample
4. Defensive forecasting
5. Philosophical implications
6. Complements

1. The question

Is everything stochastic?

Does every event have an objective probability?

- Andrei Kolmogorov said no.
- Karl Popper said yes.
- I will say yes.

Bien sûr, chaque réponse donne un sens différent à la question.

Does every event have an objective probability?

Kolmogorov said **NO.**

Not every event has a definite probability. The assumption that a definite probability in fact exists for a given event under given conditions is a *hypothesis* which must be verified or justified in each individual case.

Great Soviet Encyclopedia, 1951
(quotation abridged)

Andrei Kolmogorov (1903-1987)

Does every event have an objective probability?

Popper said **YES.**

I suggest *a new physical hypothesis*: every experimental arrangement generates propensities which can sometimes be tested by frequencies.

Realism and the Aim of Science, 1983
(quotation abridged)

Karl Popper (1902-1994)

Does every event have an objective probability?

Three ways of framing the question:

- Kolmogorov considered repeatable conditions. He thought the frequency might not be stable.
- Popper imagined repetitions. He asserted the existence of a stable "virtual" frequency even if the imagined repetition is impossible.
- I assume only that the event is embedded in a sequence of events. We can successively assign the events probabilities that will pass all statistical tests.

Giving probabilities for successive events.

Think “stochastic process, unknown probabilities”, not “iid”.

Can I assign probabilities that will pass statistical tests?

1. If you insist that I announce all probabilities before seeing any outcomes, NO.

2. If you always let me see the preceding outcomes before I announce the next probability, then YES.

2. The game

A game between Forecaster and Reality

Forecaster gives probabilities for a sequence $x_1, x_2, \ldots$ of 1s and 0s.

Before Reality announces x_n, Forecaster announces probability p_n for $x_n = 1$.

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

Theorem: Forecaster can give $p_1, p_2, \ldots$ that are not refuted by statistical tests.

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

Clarifications:

1. The phenomena need not be binary. We assume $x_n \in \{0, 1\}$ only for simplicity.
2. Reality's move space may change from round to round.
3. Perfect information: All players hear announcements as they are made.
4. In addition to $x_1, \ldots, x_{n-1}$, Forecaster may have other newly acquired information.
5. To be fair to Forecaster, we do not consider statistical tests based on information he does not have.

Bayesian Forecaster

Suppose Reality plays

$$x_1 = 1$$
$$x_2 = 0$$
$$\ldots$$

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

$$p_1 := \mathrm{P}(x_1 = 1)$$
$$p_2 := \mathrm{P}(x_2 = 1 | x_1 = 1)$$
$$p_3 := \mathrm{P}(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 0)$$
$$\ldots$$

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

$$p_1 := \mathrm{P}(x_1 = 1 | \text{all info before round 1})$$
$$p_2 := \mathrm{P}(x_2 = 1 | \text{all info before round 2})$$
$$p_3 := \mathrm{P}(x_3 = 1 | \text{all info before round 3})$$
$$\ldots$$

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

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

Forecaster's moves do not define a probability distribution for $x_1, x_2, \ldots$.

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

$$
\begin{aligned}
p_1 &= \mathrm{P}(x_1 = 1) \\
p_2 &= \mathrm{P}(x_2 = 1 | x_1 = 1) \\
p_3 &= \mathrm{P}(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 0) \\
&\ldots
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathrm{P}(x_2 = 1 | x_1 = 0) \\
&\mathrm{P}(x_3 = 1 | x_1 = 0 \ \& \ x_2 = 0) \\
&\mathrm{P}(x_3 = 1 | x_1 = 0 \ \& \ x_2 = 1) \\
&\mathrm{P}(x_3 = 1 | x_1 = 1 \ \& \ x_2 = 1) \\
&\ldots
\end{aligned}
$$

Forecaster is tested by a third player, Skeptic, who tries to get rich from Forecaster's betting offers.

Players: Forecaster, Reality, Skeptic

Protocol:

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

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

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 $\lim_{n\to\infty} \mathcal{K}_n = \infty$. Otherwise Forecaster and Reality win.

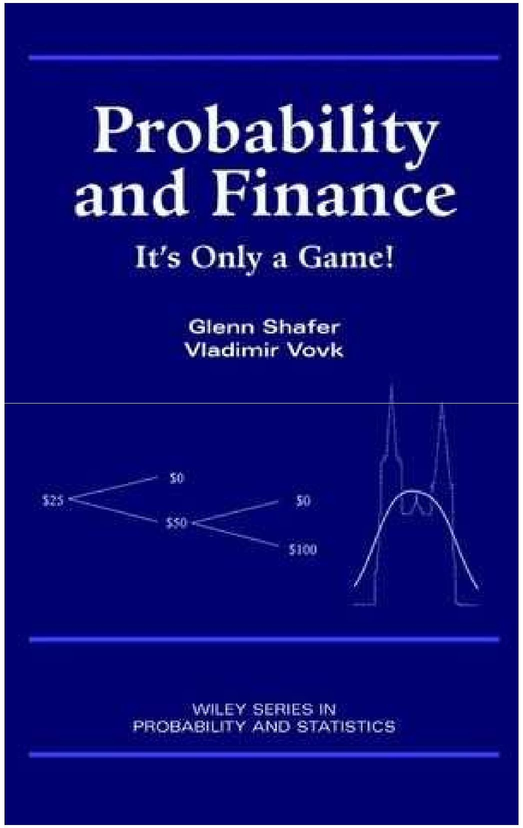
Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n\to\infty} \mathcal{K}_n = \infty$. Otherwise Forecaster and Reality win.

The thesis that statistical testing can be always be carried out by <u>strategies that attempt to multiply the capital risked</u> goes back to Ville.

Jean André Ville, 1910-1989.

At home at 3, rue Campagne Première, shortly after the Liberation.

For more on statistical testing by martingales, see my 2001 book with Kolmogorov's student Volodya Vovk.



www.probabilityandfinance.com

Vladimir Vovk, born 1960

3. Hilary Putnam's counterexample

With Bruno Latour

Putnam thought probability prediction is impossible:

FOR $n = 1, 2, \dots$
  Forecaster announces $p_n \in [0, 1]$.
  Skeptic announces $s_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0, 1\}$.
  Skeptic's profit $:= s_n(y_n - p_n)$.

Reality makes Forecaster look as bad as possible:

$$y_n := \begin{cases} 1 & \text{if } p_n < 0.5 \\ 0 & \text{if } p_n \geq 0.5 \end{cases}$$

Skeptic then makes steady money:

$$s_n := \begin{cases} 1 & \text{if } p_n < 0.5 \\ -1 & \text{if } p_n \geq 0.5 \end{cases}$$

Reality makes Forecaster look as bad as possible:

$$y_n = \begin{cases} 1 & \text{if } p_n < 0.5 \\ 0 & \text{if } p_n \geq 0.5 \end{cases}$$

Skeptic then makes steady money:

$$\text{bet} = \begin{cases} \text{50 cents on 1} & \text{if } p_n < 0.5 \\ \text{50 cents on 0} & \text{if } p_n \geq 0.5 \end{cases}$$

But the example is artificial, because the testing strategy is discontinuous in the forecast p_n.

Two paths to successful probability forecasting

1. Insist that tests be continuous. Conventional tests can be implemented with continuous betting strategies (Shafer & Vovk, 2001). Only continuous functions are constructive (L. E. J. Brouwer).

Leonid Levin, born 1948

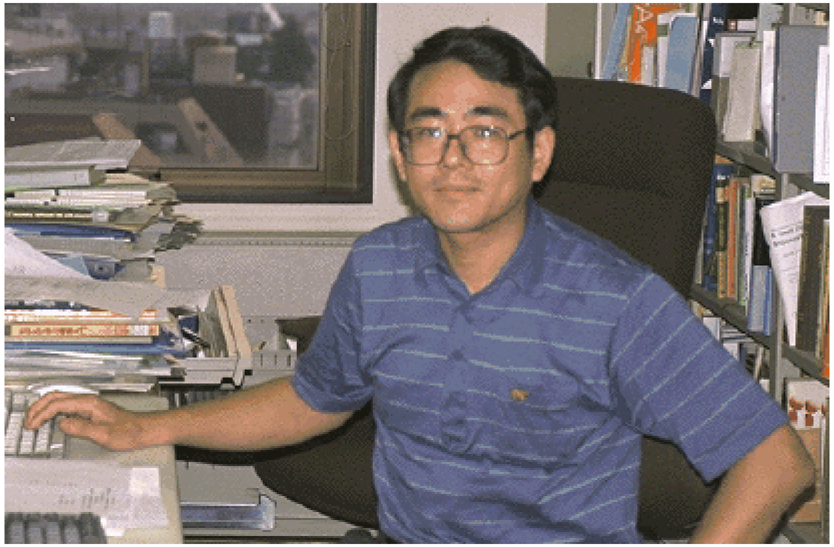
2. Allow Forecaster to hide his precise prediction from Reality using a bit of randomization.

Dean Foster

Rakesh Vohra

4. Defensive forecasting

The name was introduced in Working Paper 8 at www.probabilityandfinance, by Vovk, Takemura, and Shafer (September 2004). See also Working Papers 7, 9, 10, 11, 13, 14, 16, 17, 18, 20, 21, 22, and 30.



Akimichi Takemura in 1994

Crucial idea: all the tests (betting strategies for Skeptic) Forecaster needs to pass can be merged into a single portmanteau test for Forecaster to pass.

1. If you have two strategies for multiplying capital risked, divide your capital between them.

2. Formally: average the strategies.

3. You can average countably many strategies.

4. As a practical matter, there are only countably many tests (Abraham Wald, 1937).

5. **I will explain how Forecaster can beat any single test (including the portmanteau test).**

A. How Forecaster beats any single test

B. How to construct a portmanteau test for binary probability forecasting

- **Use law of large numbers to test calibration for each probability p.**
- **Merge the tests for different p.**

C. How the idea generalizes

How Forecaster can beat the any single test *S*

Skeptic adopts a continuous strategy $\mathcal{S}$.
FOR $n = 1, 2, \ldots$
Reality announces $x_n \in \mathbf{X}$.
Forecaster announces $p_n \in [0, 1]$.
Skeptic makes the move s_n specified by $\mathcal{S}$.
Reality announces $y_n \in \{0, 1\}$.
Skeptic's profit $:= s_n(y_n - p_n)$.

Theorem Forecaster can guarantee that Skeptic never makes money.

We actually prove a stronger theorem. Instead of making Skeptic announce his entire strategy in advance, only make him reveal his strategy for each round in advance of Forecaster's move.

FOR $n = 1, 2, \ldots$
Reality announces $x_n \in \mathbf{X}$.
Skeptic announces continuous $S_n : [0, 1] \to \mathbb{R}$.
Forecaster announces $p_n \in [0, 1]$.
Reality announces $y_n \in \{0, 1\}$.
Skeptic's profit $:= S_n(p_n)(y_n - p_n)$.

Theorem. Forecaster can guarantee that Skeptic never makes money.

FOR $n = 1, 2, \ldots$
Reality announces $x_n \in \mathbf{X}$.
Skeptic announces continuous $S_n : [0, 1] \to \mathbb{R}$.
Forecaster announces $p_n \in [0, 1]$.
Reality announces $y_n \in \{0, 1\}$.
Skeptic's profit $:= S_n(p_n)(y_n - p_n)$.

Theorem Forecaster can guarantee that Skeptic never makes money.

Proof:

- If $S_n(p) > 0$ for all p, take $p_n := 1$.
- If $S_n(p) < 0$ for all p, take $p_n := 0$.
- Otherwise, choose p_n so that $S_n(p_n) = 0$.

The game between Forecaster and Reality

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

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

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

Constructing a portmanteau test

In practice, we want to test

1. **calibration (x=1 happens 30% of the times you say p=.3)**
2. **resolution (also true just for times when it rained yesterday)**

For simplicity, consider only calibration.

1. **Use law of large numbers to test calibration for each p.**
2. **Merge the tests for different p.**

FOR $n = 1, 2, \ldots$
  Reality announces $x_n \in \mathbf{X}$.
  Forecaster announces $p_n \in [0, 1]$.
  Reality announces $y_n \in \{0, 1\}$.

1. Fix $p^* \in [0, 1]$. Look at n for which $p_n \approx p^*$. If the frequency of $y_n = 1$ always approximates p^*, Forecaster is *properly calibrated*.

2. Fix $x^* \in \mathbf{X}$ and $p^* \in [0, 1]$. Look at n for which $x_n \approx x^*$ and $p_n \approx p^*$. If the frequency of $y_n = 1$ always approximates p^*, Forecaster is properly calibrated and has *good resolution*.

Skeptic can easily multiply the capital he risks when he bets against an uncalibrated constant probability.

Ville's strong law of large numbers.

(Special case where probability is always 1/2.)

$\mathcal{K}_0 = 1.$
FOR $n = 1, 2, \ldots$:
Skeptic announces $s_n \in \mathbb{R}$.
Reality announces $y_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - \frac{1}{2})$.
Skeptic wins if
(1) $\mathcal{K}_n$ is never negative and
(2) either $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{2}$ or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.

Theorem Skeptic has a winning strategy.

Ville's strategy

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \dots$:
    Skeptic announces $s_n \in \mathbb{R}$.
    Reality announces $y_n \in \{0, 1\}$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - \frac{1}{2})$.

Ville suggested the strategy

$$s_n(y_1, \dots, y_{n-1}) = \frac{4}{n+1}\mathcal{K}_{n-1}\left(r_{n-1} - \frac{n-1}{2}\right), \text{ where } r_{n-1} := \sum_{i=1}^{n-1} y_i.$$

It produces the capital

$$\mathcal{K}_n = 2^n \frac{r_n!(n-r_n)!}{(n+1)!}.$$

From the assumption that this remains bounded, you can easily prove, using Stirling's formula, that $r_n/n \to \frac{1}{2}$.

Fundamental idea: Average strategies for Skeptic for a grid of values of p^*. (The p^*-strategy makes money if calibration fails for p_n close to p^*.) The derived strategy for Forecaster guarantees good calibration everywhere.

Example of a resulting strategy for Skeptic:

$$S_n(p) := \sum_{i=1}^{n-1} e^{-C(p-p_i)^2}(y_i - p_i)$$

Any kernel $K(p, p_i)$ can be used in place of $e^{-C(p-p_i)^2}$.

Skeptic's strategy:

$$S_n(p) := \sum_{i=1}^{n-1} e^{-C(p-p_i)^2}(y_i - p_i)$$

Forecaster's strategy: Choose p_n so that

$$\sum_{i=1}^{n-1} e^{-C(p_n-p_i)^2}(y_i - p_i) = 0.$$

The main contribution to the sum comes from i for which p_i is close to p_n. So Forecaster chooses p_n in the region where the $y_i - p_i$ average close to zero.

On each round, choose as p_n the probability value where calibration is the best so far.

Defensive forecasting is not Bayesian

TWO APPROACHES TO FORECASTING

FOR $n = 1, 2, \ldots$
  Forecaster announces $p_n \in [0, 1]$.
  Skeptic announces $s_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0, 1\}$.

1. Start with strategies for Forecaster. Improve by averaging (Bayes, prediction with expert advice).

2. Start with strategies for Skeptic. Improve by averaging (defensive forecasting).

5. Philosophical implications

We knew that a probability can be estimated from a random sample. But this depends on the idd assumption.

Defensive forecasting tells us something new.

1. Our opponent is Reality rather than Nature.
   (Nature follows laws; Reality plays as he pleases.)

2. Defensive forecasting gives probabilities that pass statistical tests regardless of how Reality behaves.

3. I conclude that the idea of an unknown inhomogeneous stochastic process has no empirical content.

Three settings for probability

1. **Causal theory.** I must give probabilities for the whole sequence $x_1, x_2, \ldots$ at the outset. (I do not observe $x_1, \ldots, x_{n-1}$ before predicting x_n.) I can succeed only if I have a valid theory. The valid theory may give only upper and lower probabilities, as in Shafer & Vovk (2001).

2. **On-line prediction.** For each n, I must predict x_n after observing $x_1, \ldots, x_{n-1}$. Using defensive forecasting, I can succeed without knowing anything about Reality. (This is why I say that the idea of an unknown inhomogeneous stochastic process has no empirical content). My predictions will be additive probabilities, not merely upper and lower probabilities.

3. **Probability judgement.** I must predict x without having specified a sequence $x_1, \ldots, x_{n-1}$ preceding it. Perhaps I face advocates with different choices for $x_1, \ldots, x_{n-1}$. It may be difficult to provide even upper and lower probabilities. (Dempster-Shafer is one method for this situation.)

Does every event have an objective probability?

- Kolmogorov considered repeatable conditions. He thought the frequency might not be stable.

I agree.

- Popper imagined repetitions. He asserted the existence of a stable "virtual" frequency even if the imagined repetition is impossible.

A major blunder, Most probabilists, statisticians, and econometricians make the same blunder.

- I assume only that the event is embedded in a sequence of events. We can successively assign probabilities that will pass all statistical tests.

Success in online prediction does not demonstrate knowledge of reality. The statistician's skill resides in the choice of the sequence and the kernel, not in modeling.

6. Complements

Karl Popper

1. Published *Logik der Forschung* in Vienna in 1935. Translated into English in 1959.
2. Sought a position in Britain, then left Vienna definitively for New Zealand in 1937.
3. Finally obtained a position in Britain in 1946, after becoming celebrated for *The Open Society*.
4. Wrote his lengthy *Postscript* to the *Logik der Forschung* in the 1950s. It was published in three volumes in 1982-1983.

The *Postscript* was published as three books:

1. *Realism and the Aim of Science.* A philosophical foundation for Kolmogorov's measure-theoretic framework for probability.
   My evaluation: Flawed and ill-informed. But important, because the notion of propensities is extremely popular.
2. *The Open Universe: An Argument for Indeterminism.*
   My evaluation: effective and insufficiently appreciated.
3. *Quantum Mechanics and the Schism in Physics*.

References

- *Probability and Finance, It's Only a Game*, Glenn Shafer and Vladimir Vovk, Wiley, 2001.

- Many working papers at www.probabilityandfinance.com