Institut d'histoire et de philosophie des sciences et des techniques
Sminaire de philosophie des probabilités, dirigé par Thierry Martin & Jacques Dubucs

Implications philosophiques de la prévision défensive

Glenn Shafer

la grande salle de l'IHPST, 13 rue du Four, 75006 Paris, 18 mai 2006

Part I. Les probabilités ludiques:
La théorie de mesure remplacée par la théorie des jeux.

Part II. Resultat étonnant:
Bonnes prévisions sont toujours possible, quelque soit le comportement de la réalité.

Replace measure theory with game theory as a framework for probability

Classical theorems in probability become theorems about games where betting odds are specified but *no stochasticity is assumed*.

Probability and Finance: It's Only a Game!
Glenn Shafer and Volodya Vovk, Wiley 2001
http://www.probabilityandfinance.com

MEASURE-THEORETIC FRAMEWORK

- Start with prices for everything.
- Basic framework (measure space) is static. Add filtration to model time.
- Draw conclusions except for a set of measure zero.

GAME-THEORETIC FRAMEWORK

- Limited prices (betting offers).
- Sequential perfect-information game.
- Prices may be set in the course of the game.
- Draw conclusions unless player becomes infinitely rich.

Classical theorems (law of large numbers, central limit theorem, etc.) become theorems about a two-player perfect-information game.

On each round of the game:		
Player I	(Skeptic)	bets on what Reality will do.
Player II	(Reality)	decides what to do.

Each theorem says that Skeptic has a strategy guaranteed to achieve a certain goal.

Example: Coin Tossing

On each round, Skeptic bets as much as he wants on heads or tails, at even odds. Skeptic wins if

1. he never goes broke, and

2. either he becomes infinitely rich or else the proportion of heads converges to one-half.

Theorem. Skeptic has a winning strategy.

Players: Skeptic, Reality

Protocol:

$\mathcal{K}_0 = 1$.

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

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

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

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

Winner:

Skeptic wins if

(1) $\mathcal{K}_n$ is never negative and

(2) either $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = 0$ or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.

Otherwise Reality wins.

Theorem. Skeptic has a winning strategy.

Generalize by letting Skeptic choose any number in the interval $[-1, 1]$. Then we have a sequence of variables $x_1, x_2, \ldots$. (Don't call them "random variables", because they have no probability distribution—just a price of zero.)

Players: Skeptic, Reality
Protocol:
  $\mathcal{K}_0 = 1$.
  FOR $n = 1, 2, \ldots$:
    Skeptic announces $M_n \in \mathbb{R}$.
    Reality announces $x_n \in [-1, 1]$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.
Winner:
Skeptic wins if
  (1) $\mathcal{K}_n$ is never negative and
  (2) either $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0$ or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.
Otherwise Reality wins.

Theorem. Skeptic has a winning strategy.

Generalize further by letting another player set the prices on each round.

Players: Forecaster, Skeptic, Reality
Protocol:
  $\mathcal{K}_0 = 1$.
  FOR $n = 1, 2, \ldots$:
    Forecaster announces $m_n \in \mathbb{R}$.
    Skeptic announces $M_n \in \mathbb{R}$.
    Reality announces $x_n \in [m_n - 1, m_n + 1]$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n)$.
Winner:
Skeptic wins if
  (1) $\mathcal{K}_n$ is never negative and
  (2) either $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_i - m_i) = 0$ or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.
Otherwise Reality wins.

Theorem. Skeptic has a winning strategy.

The Idea of the Proof

Idea 1 Establish an account for betting on heads. On each round, bet ϵ of the account on heads. Then Reality can keep the account from getting indefinitely large only by eventually holding the cumulative proportion of heads at or below $\frac{1}{2}(1+\epsilon)$. It does not matter how little money the account starts with.

Idea 2 Establish infinitely many accounts. Use the kth account to bet on heads with $\epsilon = 1/k$. This forces the cumulative proportion of heads to stay at 1/2 or below.

Idea 3 Set up similar accounts for betting on tails. This forces Reality to make the proportion converge exactly to one-half.

Averaging strategies replaces intersecting sets!

In the standard proofs, you are able to average a countable number of strategies.

Insight into a century-old problem: We do not need the axiom of continuity (countable additivity) reluctantly adopted by Borel and Kolmogorov.

Game theory replaces measure theory.

- Mathematics: Classical probability theorems become theorems in game theory (someone has a winning strategy).

- Philosophy: Cournot's principle (an event of small probability does not happen) becomes game-theoretic (you do not get rich without risking bankruptcy).

Jean Ville, 1910–1988, on entering the *École Normale Supérieure*.

If you never bet more than you have, you will not get infinitely rich.

As Ville showed, this is equivalent to the principle that events of small probability will not happen.

We call both principles Cournot's principle.

Infinitary and finitary versions of the theory

- The strong law of large numbers. Infinite and impractical: You will not get infinitely rich in an infinite number of trials.

- The weak law of large numbers. Finite and practical: You will not multiply your capital by a large factor in N trials.

Part II. Defensive forecasting.

Good probability forecasting is possible.

- We call it defensive forecasting because it defends against a portmanteau (quasi-universal) test.

- Your probability forecasts will pass this portmanteau test even if reality plays against you.

Defensive forecasting is a radically new method, not encountered in classical or measure-theoretic probability.

Part II. Defensive Forecasting

1. Definition. A strategy for Skeptic is a test.

2. Theorem. Forecaster can beat any test.

3. Practical message. We can make a test that is passed only if the probability forecasts are well calibrated.

4. Thesis. Good (well calibrated) probability forecasting is possible.

THESIS

Good probability forecasting is possible.

We can always give probabilities with good calibration and resolution.

PERFECT INFORMATION PROTOCOL

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

There exists a strategy for Forecaster that gives p_n with good calibration and resolution.

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*.

FOR $n = 1, 2, \ldots$

Reality announces $x_n \in \mathbf{X}$.

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

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

Forecaster can give ps with good calibration and resolution *no matter what Reality does.*

Philosophical implications:

- To a good approximation, everything is stochastic.
- Getting the probabilities right means describing the past well, not having insight into the future.

THEOREM. Forecaster can beat any test.

FOR $n = 1, 2, \ldots$

Reality announces $x_n \in \mathbf{X}$.

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

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

- Theorem. Given a test, Forecaster has a strategy guaranteed to pass it.

- Thesis. There is a test of Forecaster universal enough that passing it implies the ps have good calibration and resolution. (Not a theorem, because "good calibration and resolution" is fuzzy.)

The probabilities are tested by another player, Skeptic.

FOR $n = 1, 2, \dots$

Reality announces $x_n \in \mathbf{X}$.

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

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

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

Skeptic's profit $:= s_n(y_n - p_n)$.

A test of Forecaster is a strategy for Skeptic that is continuous in the ps. **If Skeptic does not make too much money, the ps pass the test.**

Theorem If Skeptic plays a known continuous strategy, Forecaster has a strategy guaranteeing that Skeptic never makes money.

Why insist on continuity? Why count only strategies for Skeptic that are continuous in the ps as tests of Forecaster?

1. *Brouwer's thesis*: A computable function of a real argument is continuous.

2. Classical statistical tests (e.g., reject if LLN fails) correspond to continuous strategies.

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 $S_n(p_n) = 0$.

Research agenda. Use proof to translate tests of Forecaster into forecasting strategies.

- Example 1: Use a strategy for Sceptic that makes money if Reality does not obey the LLN (frequency of $y_n = 1$ overall approximates average of p_n). The derived strategy for Forecaster guarantees the LLN—i.e., its probabilities are calibrated "in the large".

- Example 2: Use a strategy for Skeptic that makes money if Reality does not obey the LLN for rounds where p_n is close to p^*. The derived strategy for Forecaster guarantees calibration for p_n close to p^*.

- Example 3: Average the preceding strategies for Skeptic for a grid of values of p^*. The derived strategy for Forecaster guarantees good calibration everywhere.

- Example 4: Average over a grid of values of p^* and x^*. Then you get good resolution too.

Example 3: 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.

Example 4: Average over a grid of values of p^* and x^*. (The (p^*, x^*)-strategy makes money if calibration fails for n where (p_n, x_n) is close to (p^*, x^*).) Then you get good calibration and good resolution.

- Define a metric for $[0,1] \times \mathbf{X}$ by specifying an inner product space H and a mapping
$$\Phi : [0,1] \times \mathbf{X} \to H$$
continuous in its first argument.

- Define a kernel $K : ([0,1] \times \mathbf{X})^2 \to \mathbb{R}$ by
$$K((p,x)(p',x')) := \Phi(p,x) \cdot \Phi(p',x').$$

The strategy for Skeptic:

$$S_n(p) := \sum_{i=1}^{n-1} K((p,x_n)(p_i,x_i))(y_i - p_i).$$

Skeptic's strategy:

$$S_n(p) := \sum_{i=1}^{n-1} K((p, x_n)(p_i, x_i))(y_i - p_i).$$

Forecaster's strategy: Choose p_n so that

$$\sum_{i=1}^{n-1} K((p_n, x_n)(p_i, x_i))(y_i - p_i) = 0.$$

The main contribution to the sum comes from i for which (p_i, x_i) is close to (p_n, x_n). So we need to choose p_n to make (p_n, x_n) close (p_i, x_i) for which $y_i - p_i$ average close to zero.

Choose p_n to make (p_n, x_n) look like (p_i, x_i) for which we already have good calibration/resolution.

References

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

- www.probabilityandfinance.com: Chapters from book, reviews, many working papers.

- *Statistical Science*, forthcoming: The sources of Kolmogorov's *Grundebegriffe*.

- Journal of the Royal Statistical Society, Series B **67** 747-764. 2005: Good randomized sequential probability forecasting is always possible.

Continuity rules out Dawid's counterexample

FOR $n = 1, 2, \ldots$
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 $:= S_n(p_n)(y_n - p_n)$.

Reality can make Forecaster uncalibrated by setting

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

Skeptic can then make steady money with

$$S_n(p) := \begin{cases} 1 & \text{if } p < 0.5 \\ -1 & \text{if } p \geq 0.5, \end{cases}$$

But if Skeptic is forced to approximate S_n by a continuous function of p, then the continuous function will have a zero close to $p = 0.5$, and so Forecaster will set $p_n \approx 0.5$.

THREE 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 (prediction with expert advice).

2. Start with strategies for Skeptic. Improve by averaging (approach of this talk).

3. Start with strategies for Reality (probability disributions). Improve by averaging (Bayesian theory).