LIP 6

Defensive Forecasting

Glenn Shafer

May 18, 2006

Part I. A new mathematical foundation for probability theory
Game theory replaces measure theory.

Part II. Application to statistics: Defensive forecasting.
Good probability forecasting is possible.

Part I. A new mathematical foundation for probabil

Game theory replaces measure theory.

- Mathematics: Classical probability theorems bec theorems in game theory (someone has a winnin
- Philosophy: Cournot's principle (an event of sm probability does not happen) becomes game-the do not get rich without risking bankruptcy).

Part II. Application to statistics: Defensive forecast

Good probability forecasting is possible.

- We call it defensive forecasting because it defen portmanteau (quasi-universal) test.

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

Defensive forecasting is a radically new method, no encountered in classical or measure-theoretic probal

Part I. Basics of Game-Theoretic Probabil

1. Pascal & Ville. Pascal assumed no arbitrage (yo make money for sure) in a sequential game. Vil Cournot's principle (you will not get rich withou bankruptcy).

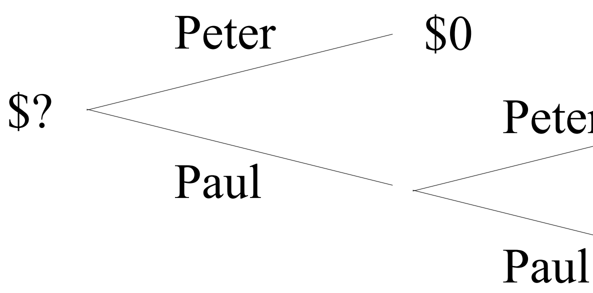
2. The strong law of large numbers

3. The weak law of large numbers

Blaise Pascal (1623–1662), as imagined in the 19th century by Hippolyte Flandrin.

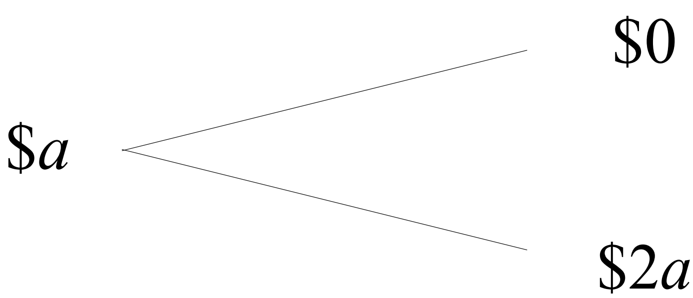
Pascal: Fair division

Peter and Paul play for $
behind. Paul needs 2 po
and Peter needs only 1.

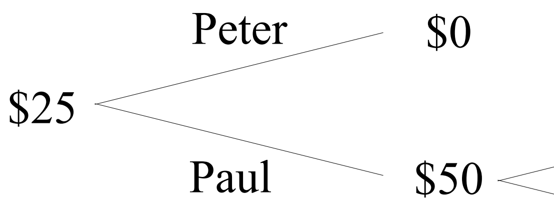


If the game must be bro
much of the $100 shoul

It is fair for Paul to pay $\$a$ in order to get $\$2a$ if he defeats Peter and $0 if he loses to Peter.



So Paul should



Modern formulation:
on the left is availab
above are forced by
of no arbitrage.

Binary probability game.

(Here $\mathcal{K}_n$ is Skeptic's capital and s_n is the total stal

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

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

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

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

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

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

No Arbitrage: If Forecaster announces a strategy in the strategy must obey the rules of probability to k from making money for sure.

In other words, the p_n should be conditional probab some probability distribution for $y_1, y_2, \ldots$.

Blaise Pascal

Probability is about fair prices in a sequential game

Pascal's concept of fairness: no arbitrage.

Jean Ville

A second concept of fairness: you will not get rich
risking bankruptcy.

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

In 1939, Ville showed th
of probability can be der
principle of market efficie

If you never bet mo
you have, you will no
finitely rich.

As Ville showed, this is
to the principle that ever
probability will not happe
both principles Cournot's

Binary probability game when Forecaster uses the s
given by a probability distribution P.

$\mathcal{K}_0 := 1.$
FOR $n = 1, 2, \ldots$:
  Skeptic announces $s_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0, 1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - \mathsf{P}\{Y_n = 1 | Y_1 = y_1, \ldots, Y_{n-1} =$

Restriction on Skeptic: Skeptic must choose the s_n
$\mathcal{K}_n \geq 0$ for all n no matter how Reality moves.

Two sides of fairness in game-theoretic probability.

Pascal Constraint on Forecaster: Don't let Skeptic
money for sure. (No arbitrage.)

Ville Constraint on Skeptic: Do not risk bankruptc
(Cournot's principle say's he will then not make
money.)

Part I. Basics of Game-Theoretic Probabil

1. Pascal & Ville

2. The strong law of large numbers (Borel). The c
version says the proportion of heads converges t
on a set of measure zero. The game-theoretic v
it converges to $\frac{1}{2}$ unless you get infinitely rich.

3. The weak law of large numbers

Fair-coin game. (Skeptic announces the amount
losing rather than the total stakes s_n.)

$\mathcal{K}_0 = 1$.

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

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

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

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

Skeptic wins if

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

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

Otherwise Reality wins.

Theorem Skeptic has a winning strategy.

Who wins? Skeptic wins if (1) $\mathcal{K}_n$ is never negative
either

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i = 0 \qquad \text{or} \qquad \lim_{n\to\infty} \mathcal{K}_n = \infty.$$

So the theorem says that Skeptic has a strategy th
not risk bankruptcy and (2) guarantees that either
of the y_i converges to 0 or else Skeptic becomes in

Loosely: The average of the y_i converges to 0 unle
becomes infinitely rich.

The Idea of the Proof

Idea 1 Establish an account for betting on heads. O
round, bet ϵ of the account on heads. Then Reality
the account from getting indefinitely large only by e
holding the cumulative proportion of heads at or bel
It does not matter how little money the account sta

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

Idea 3 Set up similar accounts for betting on tails.
Reality to make the proportion converge exactly to

Definitions

- A *path* is an infinite sequence $y_1 y_2 \ldots$ of moves
- An *event* is a set of paths.
- A *situation* is a finite initial sequence of moves say $y_1 y_2 \ldots y_n$.
- $\Box$ is the *initial situation*, a sequence of length ze
- When ξ is a path, say $\xi = y_1 y_2 \ldots$, write ξ^n for t $y_1 y_2 \ldots y_n$.

Game-theoretic processes and martingale

- A real-valued function on the situations is a *pro*

- A process $\mathcal{P}$ can be used as a strategy for Skep
buys $\mathcal{P}(y_1 \ldots y_{n-1})$ of y_n Skeptic in situation y_1

- A strategy for Skeptic, together with a particula
capital for Skeptic, also defines a process: Skep
process $\mathcal{K}(y_1 \ldots y_n)$.

- We also call a capital process for Skeptic a *mar*

Notation for Martingales

Skeptic begins with capital 1 in our game, but we c
the rules so he begins with α.

Write $\mathcal{K}^{\mathcal{P}}$ for his capital process when he begins wit
follows strategy $\mathcal{P}$: $\mathcal{K}^{\mathcal{P}}(\Box) = 0$ and

$$\mathcal{K}^{\mathcal{P}}(y_1 y_2 \dots y_n) := \mathcal{K}^{\mathcal{P}}(y_1 y_2 \dots y_{n-1}) + \mathcal{P}(y_1 y_2 \dots$$

When he starts with α, his capital process is $\alpha + \mathcal{K}^{\mathcal{}}$

The capital processes that begin with zero form a l
for

$$\beta\mathcal{K}^{\mathcal{P}} = \mathcal{K}^{\beta\mathcal{P}} \quad \text{and} \quad \mathcal{K}^{\mathcal{P}_1} + \mathcal{K}^{\mathcal{P}_2} = \mathcal{K}^{\mathcal{P}_1+\mathcal{P}_2}$$

So the martingales also form a linear space.

Convex Combinations of Martingales

If $\mathcal{P}_1$ and $\mathcal{P}_2$ are strategies, and $\alpha_1 + \alpha_2 = 1$, then

$$\alpha_1(1 + \mathcal{K}^{\mathcal{P}_1}) + \alpha_2(1 + \mathcal{K}^{\mathcal{P}_2}) = 1 + \mathcal{K}^{\alpha_1\mathcal{P}_1+\alpha_2}$$

—LHS is the convex combination of two martingale
begin with capital 1.

—RHS is the martingale produced by the same con
combination of strategies, also beginning with capit

Conclusion: In the game where we begin with capit
obtain a convex combination of $1 + \mathcal{K}^{\mathcal{P}_1}$ and $1 + \mathcal{K}^{\mathcal{P}}$
splitting our capital into two accounts, one with ini
α_1 and one with initial capital α_2. Apply $\alpha_1\mathcal{P}_1$ to th
account and $\alpha_2\mathcal{P}_2$ to the second.

Infinite Convex Combinations: Suppose $\mathcal{P}_1, \mathcal{P}_2, \ldots$ a
and $\alpha_1, \alpha_2, \ldots$ are nonnegative real numbers adding

- If $\sum_{k=1}^{\infty} \alpha_k \mathcal{P}_k$ converges, then $\sum_{k=1}^{\infty} \alpha_k \mathcal{K}^{\mathcal{P}_k}$ also c

- $\sum_{k=1}^{\infty} \alpha_k \mathcal{K}^{\mathcal{P}_k}$ is the capital process from $\sum_{k=1}^{\infty} \alpha_k$

- You can prove this by induction on

$$\mathcal{K}^{\mathcal{P}}(y_1 y_2 \ldots y_n) := \mathcal{K}^{\mathcal{P}}(y_1 y_2 \ldots y_{n-1}) + \mathcal{P}(y_1 y_2 \ldots$$

In game-theoretic probability, you can usually get an infinite c
combination of martingales, but you have to check on the con
the infinite convex combination of strategies. In a sense, this
historical confusion about countable additivity in measure-theo
probability (see Working Paper #4).

The greater power of game-theoretic probability

Instead of a probability distribution for $y_1, y_2, \ldots$, maybe you h
prices. Instead of giving them at the outset, maybe your mak
you go along. Instead of

Skeptic announces $M_n \in \mathbb{R}$. Reality announces $y_n \in \{-1, 1\}$. $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$.

use

Skeptic announces $M_n \in \mathbb{R}$. Reality announces $y_n \in [-1, 1]$. $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$.

or

Forecaster announces $m_n \in \mathbb{R}$. Skeptic announces $M_n \in \mathbb{R}$. Reality announces $y_n \in [m_n - 1, m_n + 1]$. $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n (y_n - m_n)$.

Part I. Basics of Game-Theoretic Probabil

1. Pascal & Ville

2. The strong law of large numbers. Infinite and in
You will not get infinitely rich in an infinite num

3. The weak law of large numbers. Finite and prac
will not multiply your capital by a large factor in

The weak law of large numbers (Bernoulli)

$\mathcal{K}_0 := 1$.
FOR $n = 1, \ldots, N$:
    Skeptic announces $M_n \in \mathbb{R}$.
    Reality announces $y_n \in \{-1, 1\}$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$.

Winning: Skeptic wins if $\mathcal{K}_n$ is never negative and $\mathcal{K}_N \geq C$ or $|\sum_{n=1}^{N} y_n/N| < \epsilon$.

Theorem. Skeptic has a winning strategy if $N \geq C/$

Part II. Defensive Forecasting

1. Thesis. Good probability forecasting is possible.

2. Theorem. Forecaster can beat any test.

3. Research agenda. Use proof to translate tests c
into forecasting strategies.

4. Example. Forecasting using LLN (law of large n

THESIS

Good probability forecasting is possibl

We can always give probabilities with good calibrati
resolution.

PERFECT INFORMATION PROTOCOL

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

There exists a strategy for Forecaster that give
good calibration and resolution.

FOR $n = 1, 2, \dots$
  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 th of $y_n = 1$ always approximates p^*, Forecaster is *calibrated*.

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

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

matter what Reality does.

Philosophical implications:

- To a good approximation, everything is stochas
- Getting the probabilities right means describing 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 strateg guaranteed to pass it.

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

The probabilities are tested by another player, Skep

FOR $n = 1, 2, \ldots$
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
in the ps. **If Skeptic does not make too much m**
ps pass the test.

Theorem If Skeptic plays a known continuous strate
Forecaster has a strategy guaranteeing that Skeptic
makes money.

This concept of test generalizes the standard stoch
concept.

Stochastic setting:

- There is a probability distribution P for the xs
- Forecaster uses P's conditional probabilities as
- Reality chooses her xs and ys from P.

Standard concept of statistical test:

- Choose an event A whose probability under P
- Reject P if A happens.

In 1939, Jean Ville showed that in the stochastic se
standard concept is equivalent to a strategy for Ske

Why insist on continuity? Why count only strategie
Skeptic that are continuous in the ps as tests of Fo

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

2. Classical statistical tests (e.g., reject if LLN fail
   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 make

We actually prove a stronger theorem. Instead of making Ske his entire strategy in advance, only make him reveal his strate 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 mak

FOR $n = 1, 2, \dots$

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 make

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 Fo
into forecasting strategies.

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

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

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

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

Example 3: Average strategies for Skeptic for a gri
of p^*. (The p^*-strategy makes money if calibration
close to p^*.) The derived strategy for Forecaster gu
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 $e^{-C(p-p_i}$

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 close to p_n. So Forecaster chooses p_n in the region $y_i - p_i$ average close to zero.

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

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

- Define a metric for $[0,1] \times \mathbf{X}$ by specifying an inner produ
  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 fo
(p_i, x_i) is close to (p_n, x_n). So we need to choose p_n
(p_n, x_n) close (p_i, x_i) for which $y_i - p_i$ average close

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

References

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

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

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

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

More talks in Paris

- 19 May, 10:00. Why did Cournot's principle disa
  EHESS, Séminaire de histoire du calcul des prob
  de la statistique, 54 boulevard Raspail

- 19 May, 14:00. Philosophical implications of de
  forecasting. Séminaire de philosophie des proba
  l'IHPST, la grande salle de l'IHPST, 13 rue du

- 5 July, 9:00-10:00. The game-theoretic framew
  probability. Plenary lecture, 11th IPMU Internat
  Conference, Les Cordeliers, 15 rue de l'Ecole de

Standard stochastic concept of statistical test:

- Choose an event A whose probability under P is
- Reject P if A happens.

Ville's Theorem: In the stochastic setting. . .

- Given an event of probability less than $1/C$, ther for Skeptic that turns \$1 into \$$C$ without riskin
- Given a strategy for Skeptic that starts with \$1 risk bankruptcy, the probability that it turns \$ more is no more than $1/C$.

So the concept of a strategy for Skeptic generalizes concept of testing with events of small probability.

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 fur
then the continuous function will have a zero close to $p = 0.5$
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 averagin
   with expert advice).

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

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