SIPTA School 08

July 8, 2008, Montpellier

What is risk? What is probability? Game-theoretic answers.

Glenn Shafer

- For 170 years: objective vs. subjective probability

- Game-theoretic probability (Shafer & Vovk, 2001) asks more concrete question:
  Is there a repetitive structure?

Distinction first made by Simon-Dénis Poisson in 1837:

- objective probability = frequency = stochastic uncertainty = aleatory probability
- subjective probability = belief = epistemic probability

Our more concrete question:

Is there a repetitive structure for the question and the data?

- If yes, we can make good probability forecasts. No model, probability assumption, or underlying stochastic reality required.
- If no, we must weigh evidence. Dempster-Shafer can be useful here.

Who is Glenn Shafer?

A Mathematical Theory of Evidence (1976) introduced the Dempster-Shafer theory for weighing evidence when the repetitive structure is weak.

The Art of Causal Conjecture (1996) is about probability when repetitive structure is very strong.

Probability and Finance: It's Only a Game! (2001) provides a unifying game-theoretic framework.
www.probabilityandfinance.com

I. Game-theoretic probability

New foundation for probability

II. Defensive forecasting

Under repetition, good probability forecasting is possible.

III. Objective vs. subjective probability

The important question is how repetitive your question is.

Part I. Game-theoretic probability

- Mathematics: The law of large numbers is a theorem about a game (a player has a winning strategy).

- Philosophy: Probabilities are connected to the real world by the principle that you will not get rich without risking bankruptcy.

Basic idea of game-theoretic probability

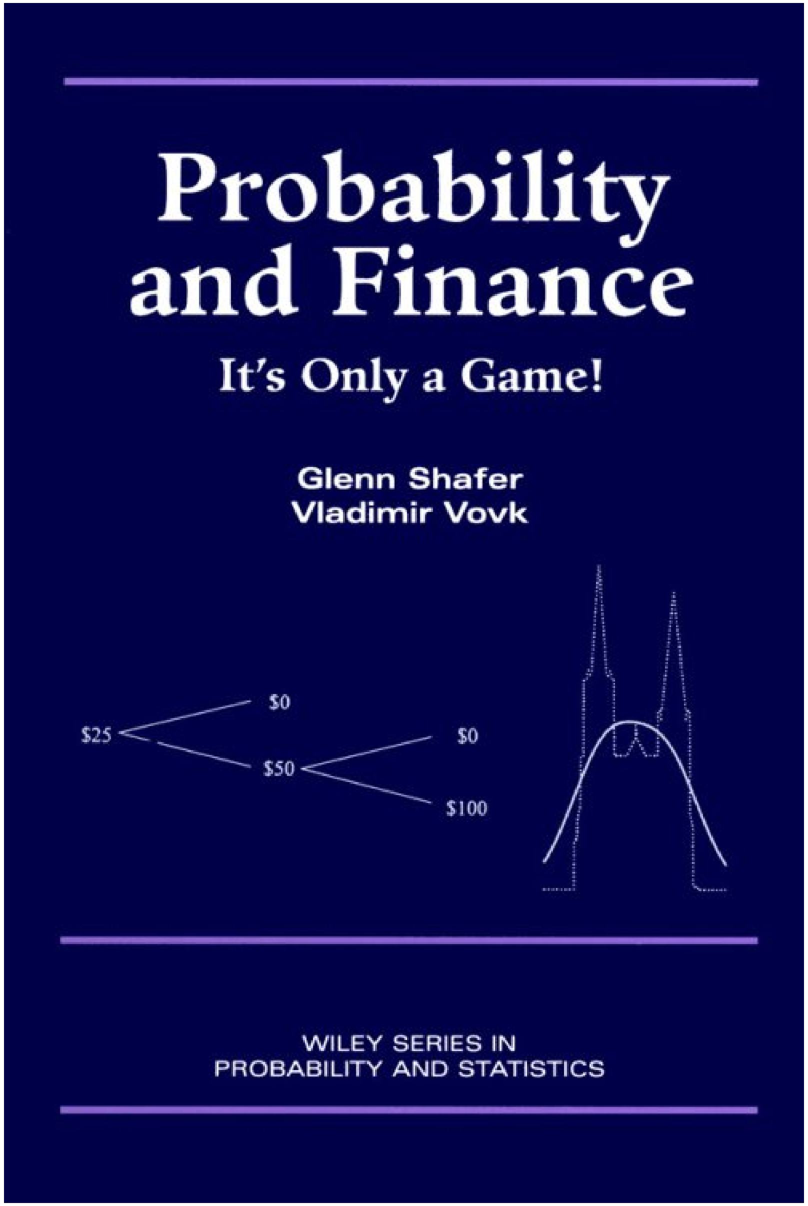
- Classical statistical tests reject if an event of small probability happens.
- But an event of small probability is equivalent to a strategy for multiplying capital you risk. (Markov's inequality.)
- So generalize by replacing event of small probability will not happen with you will not multiply capital you risk by large factor.

Game-Theoretic Probability

Online at www.probabilityandfinance.com:

- 3 chapters
- 34 working papers

Working paper 22:
Game-theoretic probability and its uses, especially defensive forecasting



Wiley 2001

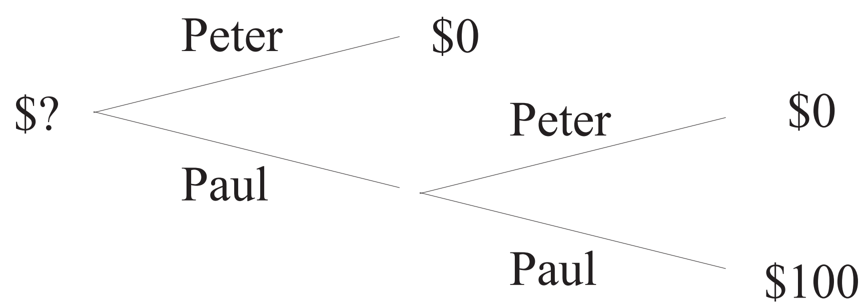
Three heroes of game-theoretic probability

		
Blaise Pascal (1623–1662)	Antoine Augustin Cournot (1801–1877)	Jean Ville (1910–1988)

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

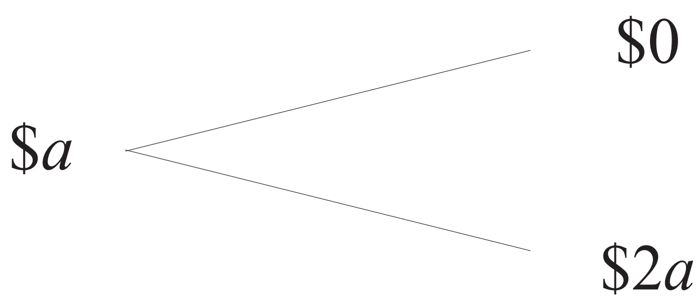
Pascal: Fair division

Peter and Paul play for \$100. Paul is behind. Paul needs 2 points to win, and Peter needs only 1.

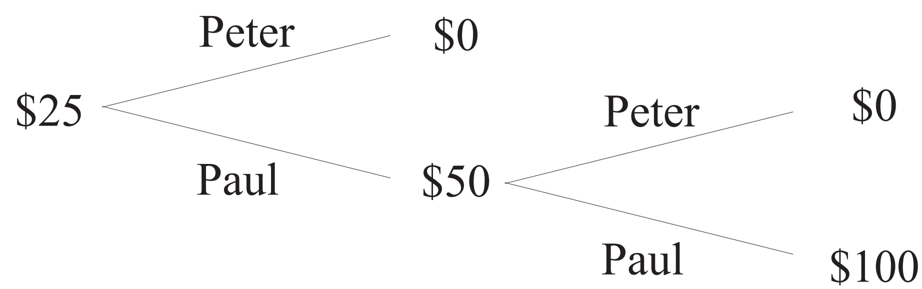


If the game must be broken off, how much of the \$100 should Paul get?

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 get $25.



Modern formulation: If the game on the left is available, the prices above are forced by the principle of no arbitrage.

Antoine Cournot (1801–1877)

"A physically impossible event is one whose probability is infinitely small. This remark alone gives substance—an objective and phenomenological value—to the mathematical theory of probability." (1843)

Agreeing with Cournot:

- Émile Borel
- Maurice Fréchet
- Andrei Kolmogorov

Fréchet dubbed the principle that an event of small probability will not happen *Cournot's principle*.

Émile Borel
1871–1956

Inventor of measure theory.

Minister of the French navy in 1925.

Borel was emphatic: the principle that an event with very small probability will not happen is the only law of chance.

- Impossibility on the human scale: $p < 10^{-6}$.
- Impossibility on the terrestrial scale: $p < 10^{-15}$.
- Impossibility on the cosmic scale: $p < 10^{-50}$.

Andrei Kolmogorov
1903–1987

Hailed as the Soviet Euler, Kolmogorov was credited with establishing measure theory as the mathematical foundation for probability.

In his celebrated 1933 book, Kolmogorov wrote:

> When P(A) very small, we can be practically certain that the event A will not happen on a single trial of the conditions that define it.

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

In 1939, Ville showed that the laws of probability can be derived from a game-theoretic principle:

> 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.

Jean André Ville (1910-1989)

Born 1910

Hometown: Mosset, in Pyrenees

Mother's family: priests, schoolteachers

Father's family: farmers

Father worked for PTT.

#1 on written entrance exam for Ecole Normale Supérieure in 1929.

Ville’s family went back 8 generations in Mosset, to the shepherd Miguel Vila.

The basic protocol for game-theoretic probability

$\mathcal{K}_0 = 1$.

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

  Reality announces x_n.

  Forecaster announces a price f_n for a ticket that pays y_n.

  Skeptic decides how many tickets to buy.

  Reality announces y_n.

  $\mathcal{K}_n := \mathcal{K}_{n-1}$ + Skeptic's net gain or loss.

Goal for Skeptic: Make K_N very large without risking K_n ever negative.

Ville showed that every statistical test of Forecaster's prices can be expressed as a strategy for Skeptic.

Example of a game-theoretic probability theorem.

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

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

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

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

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

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

Skeptic wins if

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

(2) either $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (y_i - p_i) = 0$

or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.

Theorem Skeptic has a winning strategy.

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.

Who wins? Skeptic wins if (1) $\mathcal{K}_n$ is never negative and (2) either

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{2} \qquad \text{or} \qquad \lim_{n\to\infty} \mathcal{K}_n = \infty.$$

So the theorem says that Skeptic has a strategy that (1) does not risk bankruptcy and (2) guarantees that either the average of the y_i converges to 0 or else Skeptic becomes infinitely rich.

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

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 by some constant C, you can easily derive the strong law of large numbers using Stirling's formula.

The weak law of large numbers (Bernoulli)

$\mathcal{K}_0 := 1$.
FOR $n = 1, \dots, N$:
  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 either $\mathcal{K}_N \geq C$ or $|\sum_{n=1}^{N} y_n / N| < \epsilon$.

Theorem. Skeptic has a winning strategy if $N \geq C/\epsilon^2$.

Definition of upper price and upper probability

$\mathcal{K}_0 := \alpha.$
FOR $n = 1, \dots, N$:
  Forecaster announces $p_n \in [0,1]$.
  Skeptic announces $s_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0,1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n (y_n - p_n)$.

For any real-valued function X on $([0,1] \times \{0,1\})^N$,

$$\overline{\mathbb{E}}\, X := \inf\{\alpha \mid \text{Skeptic has a strategy guaranteeing } \mathcal{K}_N \ge X(p_1, y_1, \dots, p_N, y_N)\}$$

For any subset $A \subseteq ([0,1] \times \{0,1\})^N$,

$$\overline{\mathbb{P}}\, A := \inf\{\alpha \mid \text{Skeptic has a strategy guaranteeing } \mathcal{K}_N \ge 1 \text{ if } A \text{ happens and } \mathcal{K}_N \ge 0 \text{ otherwise}\}.$$

$$\underline{\mathbb{E}}\, X = -\overline{\mathbb{E}}(-X) \qquad \underline{\mathbb{P}}\, A = 1 - \overline{\mathbb{P}}\, \overline{A}$$

Put it in terms of upper probability

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

Theorem. $\overline{\mathbb{P}}\left\{\frac{1}{N}\left|\sum_{n=1}^{N}(y_n - p_n)\right| \geq \epsilon\right\} \leq \frac{1}{4N\epsilon^2}$.

Part II. Defensive forecasting

Under repetition, good probability forecasting is possible.

- We call it defensive because it defends against a quasi-universal test.

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

Why Phil Dawid thought good probability prediction is impossible. . .

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

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

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

But if Skeptic is forced to approximate s_n by a continuous function of p_n, then the continuous function will be zero close to $p = 0.5$, and Forecaster can set p_n equal to this point.

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 of Forecaster into forecasting strategies.

4. **Example.** Forecasting using LLN (law of large numbers).

We can always give probabilities with good calibration and 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 gives p_n with good calibration and resolution.

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 $S_n(p_n) = 0$.

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 $n = 1, 2, \dots$

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 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.)

TWO APPROACHES TO FORECASTING

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

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).

The probabilities are tested by another player, Skeptic.

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 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.

Example: 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.

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.

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:

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

Part III. Aleatory (objective) vs. epistemic (subjective)

From a 1970s perspective:

- Aleatory probability is the irreducible uncertainty that remains when knowledge is complete.
- Epistemic probability arises when knowledge is incomplete.

New game-theoretic perspective:

- Under a repetitive structure you can make make good probability forecasts relative to whatever state of knowledge you have.
- If there is no repetitive structure, your task is to combine evidence rather than to make probability forecasts.

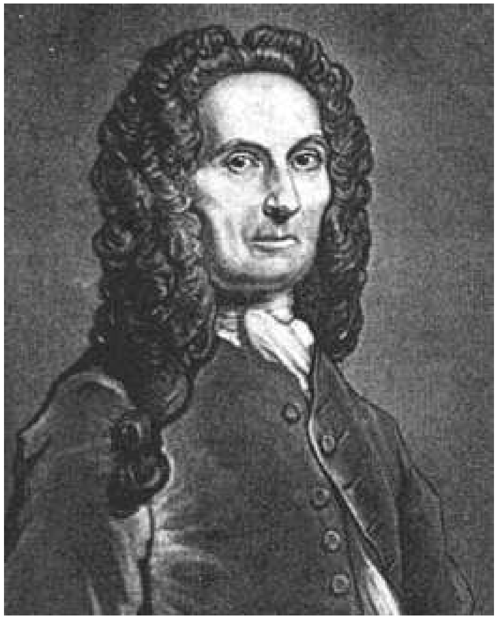
Three betting interpretations:

- De Moivre: $P(E)$ is the value of a ticket that pays 1 if E happens. (No explanation of what "value" means.)

- De Finetti: $P(E)$ is a price at which YOU would buy or sell a ticket that pays 1 if E happens.

- Shafer: The price $P(E)$ cannot be beat—i.e., a strategy for buying and selling such tickets at such prices will not multiply the capital it risks by a large factor.

De Moivre's argument for $P(A\&B) = P(A)P(B|A)$



Abraham de Moivre
1667–1754

Gambles available:

- pay $P(A)$ for 1 if A happens,
- pay $P(A)x$ for x if A happens, and
- after A happens, pay $P(B|A)$ for 1 if B happens.

To get 1 if $A\&B$ if happens, pay

- $P(A)P(B|A)$ for $P(B|A)$ if A happens,
- then if A happens, pay the $P(B|A)$ you just got for 1 if B happens.

De Finetti's argument for
$P(A\&B) = P(A)P(B|A)$

Suppose you are required to announce...

- prices $P(A)$ and $P(A\&B)$ at which you will buy or sell \$1 tickets on these events.
- a price $P(B|A)$ at which you will buy or sell \$1 tickets on B if A happens.

Opponent can make money for sure if you announce $P(A\&B)$ different from $P(A)P(B|A)$.

Bruno de Finetti
(1906–1985)

Cournotian argument for $P(B|A) = P(A\&B)/P(A)$

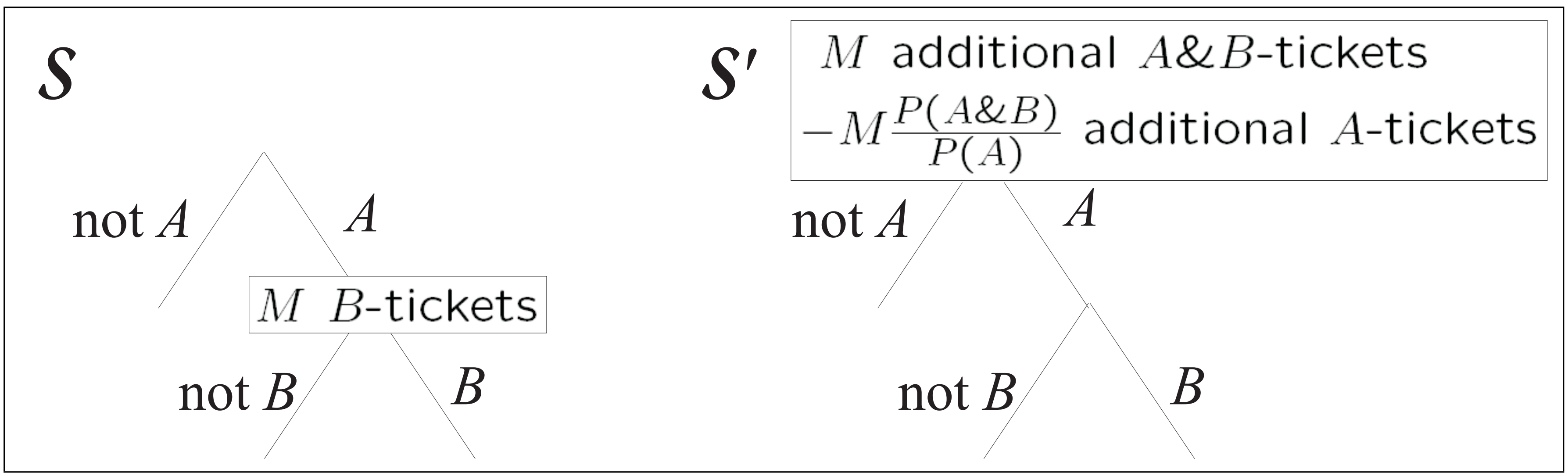
Claim: Suppose $P(A)$ and $P(A\&B)$ cannot be beat. Suppose we learn A happens and nothing more. Then we can include $P(A\&B)/P(A)$ as a new probability for B among the probabilities that cannot be beat.

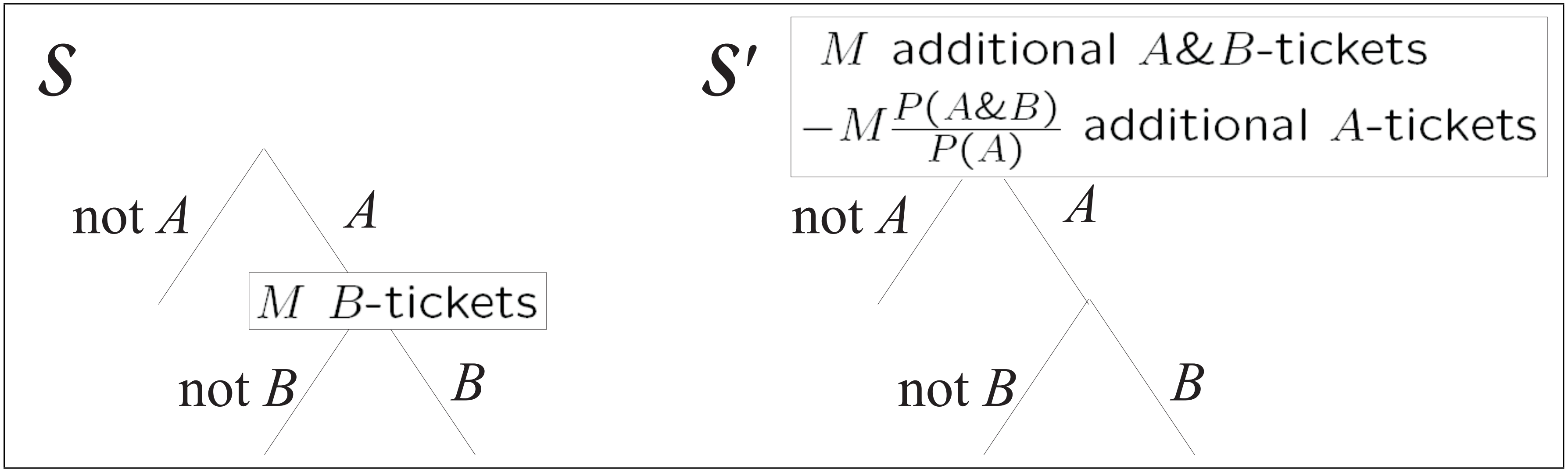
Structure of proof:

- Consider a bankruptcy-free strategy $\mathcal{S}$ against probabilities $P(A)$ and $P(A\&B)$ and $P(A\&B)/P(A)$. We want to show that $\mathcal{S}$ does not get rich.
- Do this by constructing a strategy $\mathcal{S}'$ against $P(A)$ and $P(A\&B)$ alone that does the same thing as $\mathcal{S}$.

Given: Bankruptcy-free strategy $\mathcal{S}$ that deals in A-tickets and $A\&B$-tickets in the initial situation and B-tickets in the situation where A has just happened.

Construct: Strategy $\mathcal{S}'$ that agrees with $\mathcal{S}$ except that it does not buy the B-tickets but instead initially buys additional A- and $A\&B$-tickets.





1. A's happening is the only new information used by $\mathcal{S}$. So $\mathcal{S}'$ uses only the initial information.
2. Because the additional initial tickets have net cost zero, $\mathcal{S}'$ and $\mathcal{S}$ have the same cash on hand in the initial situation.
3. In the situation where A happens, they again produce the same cash position, because the additional A-tickets require $\mathcal{S}'$ to pay $M\frac{P(A\&B)}{P(A)}$, which is the cost of the B tickets that $\mathcal{S}$ buys.
4. They have the same payoffs if not A happens (0), if $A\&$(not B) happens (0), or if $A\&B$ happens (M).
5. By hypothesis, $\mathcal{S}$ is bankruptcy-free. So $\mathcal{S}'$ is also bankruptcy-free.
6. Therefore $\mathcal{S}'$ does not get rich. So $\mathcal{S}$ does not get rich either.

Crucial assumption for conditioning on A: You learn A and nothing more that can help you beat the probabilities.

In practice, you always learn more than A.

- But you judge that the other things don't matter.

- Probability judgement is always in a small world. We judge knowledge outside the small world irrelevant.

Cournotian understanding of Dempster-Shafer

- Fundamental idea: transferring belief
- Conditioning
- Independence
- Dempster's rule

Fundamental idea: transferring belief

- Variable ω with set of possible values Ω.

- Random variable $\mathbf{X}$ with set of possible values $\mathcal{X}$.

- We learn a mapping $\Gamma : \mathcal{X} \to 2^{\Omega}$ with this meaning:

$$\text{If } \mathbf{X} = x, \text{ then } \omega \in \Gamma(x).$$

- For $A \subseteq \Omega$, our belief that $\omega \in A$ is now

$$\mathbb{B}(A) = \mathbb{P}\{x | \Gamma(x) \subseteq A\}.$$

Cournotian judgement of independence: Learning the relationship between $\mathbf{X}$ and ω does not affect our inability to beat the probabilities for $\mathbf{X}$.

Example: The sometimes reliable witness

- Joe is reliable with probability 30%. When he is reliable, what he says is true. Otherwise, it may or may not be true.

  $\mathcal{X} = \{\text{reliable}, \text{not reliable}\}$ $\quad$ $\mathbb{P}(\text{reliable}) = 0.3$ $\quad$ $\mathbb{P}(\text{not reliable}) = 0.7$

- Did Glenn pay his dues for coffee? $\quad$ $\Omega = \{\text{paid}, \text{not paid}\}$

- Joe says "Glenn paid."

  $$\Gamma(\text{reliable}) = \{\text{paid}\} \qquad \Gamma(\text{not reliable}) = \{\text{paid}, \text{not paid}\}$$

- New beliefs:

  $$\mathbb{B}(\text{paid}) = 0.3 \qquad \mathbb{B}(\text{not paid}) = 0$$

Cournotian judgement of independence: Hearing what Joe said does not affect our inability to beat the probabilities concerning his reliability.

Example: The more or less precise witness

- Bill is absolutely precise with probability 70%, approximate with probability 20%, and unreliable with probability 10%.

  $\mathcal{X} = \{\text{precise}, \text{approximate}, \text{not reliable}\}$

  $$\mathbb{P}(\text{precise}) = 0.7 \qquad \mathbb{P}(\text{approximate}) = 0.2 \qquad \mathbb{P}(\text{not reliable}) = 0.1$$

- What did Glenn pay? $\quad \Omega = \{0, \$1, \$5\}$

- Bill says "Glenn paid \$ 5."

  $$\Gamma(\text{precise}) = \{\$5\} \qquad \Gamma(\text{approximate}) = \{\$1, \$5\} \qquad \Gamma(\text{not reliable}) = \{0, \$1, \$5\}$$

- New beliefs:

  $$\mathbb{B}\{0\} = 0 \qquad \mathbb{B}\{\$1\} = 0 \qquad \mathbb{B}\{\$5\} = 0.7 \qquad \mathbb{B}\{\$1, \$5\} = 0.9$$

Cournotian judgement of independence: Hearing what Bill said does not affect our inability to beat the probabilities concerning his precision.

Conditioning

- Variable ω with set of possible values Ω.
- Random variable $\mathbf{X}$ with set of possible values $\mathcal{X}$.
- We learn a mapping $\Gamma : \mathcal{X} \to 2^{\Omega}$ with this meaning:

$$\text{If } \mathbf{X} = x, \ \text{ then } \omega \in \Gamma(x).$$

- $\boxed{\Gamma(x) = \emptyset \text{ for some } x \in \mathcal{X}.}$

- For $A \subseteq \Omega$, our belief that $\omega \in A$ is now

$$\mathbb{B}(A) = \frac{\mathbb{P}\{x | \Gamma(x) \subseteq A \ \ \& \ \ \Gamma(x) \neq \emptyset\}}{\mathbb{P}\{x | \Gamma(x) \neq \emptyset\}}.$$

Cournotian judgement of independence: Aside from the impossibility of the x for which $\Gamma(x) = \emptyset$, learning Γ does not affect our inability to beat the probabilities for $\mathbf{X}$.

Example: The witness caught out

- Tom is absolutely precise with probability 70%, approximate with probability 20%, and unreliable with probability 10%.

$$\mathcal{X} = \{\text{precise}, \text{approximate}, \text{not reliable}\}$$

$$\mathbb{P}(\text{precise}) = 0.7 \qquad \mathbb{P}(\text{approximate}) = 0.2 \qquad \mathbb{P}(\text{not reliable}) = 0.1$$

- What did Glenn pay? $\qquad \Omega = \{0, \$1, \$5\}$

- Tom says "Glenn paid \$ 10."

$$\Gamma(\text{precise}) = \emptyset \qquad \Gamma(\text{approximate}) = \{\$5\} \qquad \Gamma(\text{not reliable}) = \{0, \$1, \$5\}$$

- New beliefs:

$$\mathbb{B}\{0\} = 0 \qquad \mathbb{B}\{\$1\} = 0 \qquad \mathbb{B}\{\$5\} = 2/3 \qquad \mathbb{B}\{\$1, \$5\} = 2/3$$

Cournotian judgement of independence: Aside ruling out his being absolutely precise, what Tom said does not help us beat the probabilities for his precision.

Independence

$$\mathcal{X}_{\text{Bill}} = \{\text{Bill precise}, \text{Bill approximate}, \text{Bill not reliable}\}$$
$$\mathbb{P}(\text{precise}) = 0.7 \qquad \mathbb{P}(\text{approximate}) = 0.2 \qquad \mathbb{P}(\text{not reliable}) = 0.1$$

$$\mathcal{X}_{\text{Tom}} = \{\text{Tom precise}, \text{Tom approximate}, \text{Tom not reliable}\}$$
$$\mathbb{P}(\text{precise}) = 0.7 \qquad \mathbb{P}(\text{approximate}) = 0.2 \qquad \mathbb{P}(\text{not reliable}) = 0.1$$

Product measure:

$$\begin{aligned} \mathcal{X}_{\text{Bill \& Tom}} &= \mathcal{X}_{\text{Bill}} \times \mathcal{X}_{\text{Tom}} \\ \mathbb{P}(\text{Bill precise}, \text{Tom precise}) &= 0.7 \times 0.7 = 0.49 \\ \mathbb{P}(\text{Bill precise}, \text{Tom approximate}) &= 0.7 \times 0.2 = 0.14 \\ \text{etc.} \end{aligned}$$

Cournotian judgements of independence: Learning about the precision of one of the witnesses will not help us beat the probabilities for the other.

Nothing novel here. Dempsterian independence = Cournotian independence.

Example: Independent contradictory witnesses

- Joe and Bill are both reliable with probability 70%.
- Did Glenn pay his dues? $\Omega = \{\text{paid}, \text{not paid}\}$
- Joe says, "Glenn paid." Bill says, "Glenn did not pay."

$$\Gamma_1(\text{Joe reliable}) = \{\text{paid}\} \qquad \Gamma_1(\text{Joe not reliable}) = \{\text{paid}, \text{not paid}\}$$
$$\Gamma_2(\text{Bill reliable}) = \{\text{not paid}\} \qquad \Gamma_2(\text{Bill not reliable}) = \{\text{paid}, \text{not paid}\}$$

- The pair (Joe reliable,Bill reliable), which had probability 0.49, is ruled out.

$$\mathbb{B}(\text{paid}) = \frac{0.21}{0.51} = 0.41 \qquad \mathbb{B}(\text{not paid}) = \frac{0.21}{0.51} = 0.41$$

Cournotian judgement of independence: Aside from learning that they are not both reliable, what Joe and Bill said does not help us beat the probabilities concerning their reliability.

Dempster's rule (independence + conditioning)

- Variable ω with set of possible values Ω.
- Random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ with sets of possible values $\mathcal{X}_1$ and $\mathcal{X}_2$.
- Form the product measure on $\mathcal{X}_1 \times \mathcal{X}_2$.
- We learn mappings $\Gamma_1 : \mathcal{X}_1 \to 2^\Omega$ and $\Gamma_2 : \mathcal{X}_2 \to 2^\Omega$:

  If $\mathbf{X_1} = x_1$, then $\omega \in \Gamma_1(x_1)$. If $\mathbf{X_2} = x_2$, then $\omega \in \Gamma_2(x_2)$.
- So if $(\mathbf{X_1}, \mathbf{X_2}) = (x_1, x_2)$, then $\omega \in \Gamma_1(x_1) \cap \Gamma_2(x_2)$.
- Conditioning on what is not ruled out,

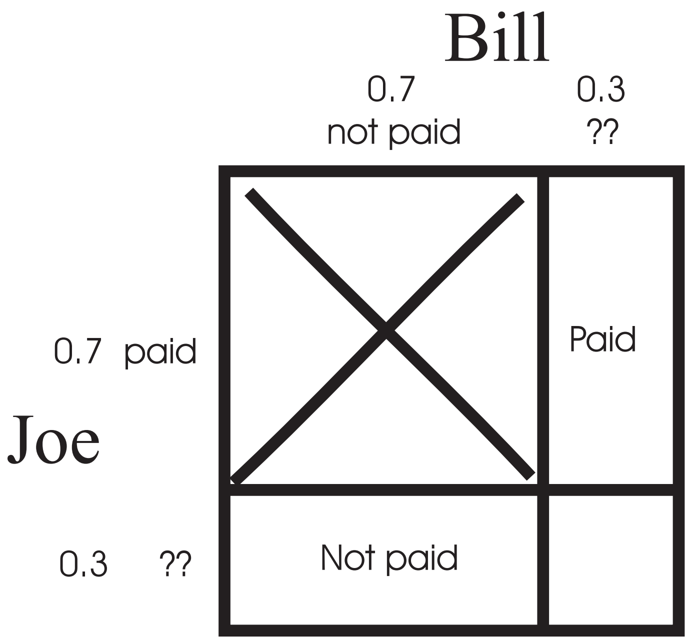
$$\mathbb{B}(A) = \frac{\mathbb{P}\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2) \subseteq A\}}{\mathbb{P}\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2)\}}$$

Cournotian judgement of independence: Aside from ruling out some (x_1, x_2), learning the Γ_i does not help us beat the probabilities for $\mathbf{X}_1$ and $\mathbf{X}_2$.

You can suppress the Γs and describe Dempster's rule in terms of the belief functions

Joe: $\mathbb{B}_1\{\text{paid}\} = 0.7$ $\mathbb{B}_1\{\text{not paid}\} = 0$

Bill: $\mathbb{B}_2\{\text{not paid}\} = 0.7$ $\mathbb{B}_2\{\text{paid}\} = 0$



$$\mathbb{B}(\text{paid}) = \frac{0.21}{0.51} = 0.41$$

$$\mathbb{B}(\text{not paid}) = \frac{0.21}{0.51} = 0.41$$

Dempster's rule is unnecessary. It is merely a composition of Cournot operations: formation of product measures, conditioning, transferring belief.

But Dempster's rule is a unifying idea. Each Cournot operation is an example of Dempster combination.

- Forming product measure is Dempster combination.

- Conditioning on A is Demspter combination with a belief function that gives belief one to A.

- Transferring belief is Dempster combination of (1) a belief function on $\mathcal{X} \times \Omega$ that gives probabilities to cylinder sets $\{x\} \times \Omega$ with (2) a belief function that gives probability one to $\{(x, \omega) | \omega \in \Gamma(x)\}$.

Parametric models are not the starting point!

- Mathematical statistics departs from probability by standing outside the protocol.

- Classical example: the error model

- Parametric modeling

- Dempster-Shafer modeling

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.

- www.glennshafer.com: Most of my published articles.

- *Statistical Science*, **21** 70–98, 2006: 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.

Art Dempster (born 1929) with his Meng & Shafer hatbox.

Retirement dinner at Harvard, May 2005.

See http://www.stat.purdue.edu/ chuanhai/projects/DS/ for Art's D-S papers.

Volodya Vovk atop the World Trade Center in 1998.

- Born 1960.
- Student of Kolmogorov.
- Born in Ukraine, educated in Moscow, teaches in London.
- Volodya is a nickname for the Ukrainian Volodimir and the Russian Vladimir.

Wiki for On-Line Prediction

http://onlineprediction.net

Main topics

1. Competitive online prediction
2. Conformal prediction
3. Game-theoretic probability
4. Prequential statistics
5. Stochastic prediction