Fusion 2007

July 9, 2007, Quebec City

What is risk? What is probability?

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

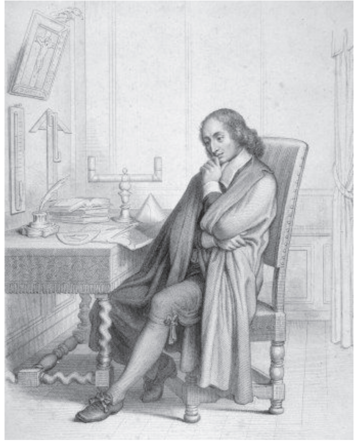
The important question is how repetitive your question is.

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.

Three heroes of game-theoretic probability



Blaise Pascal
(1623–1662)

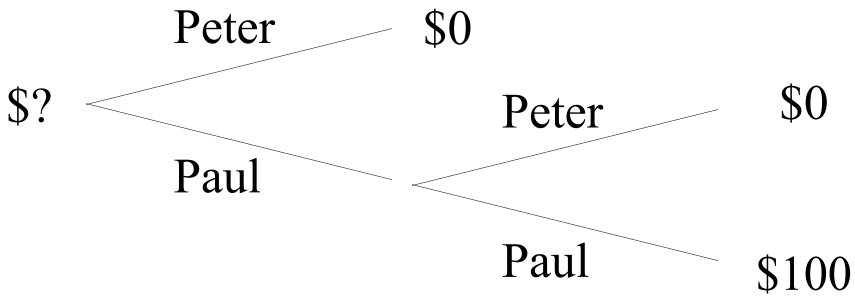
Antoine Augustin
Cournot
(1801–1877)

Jean Ville
(1910–1988)

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

Pascal: Fair division

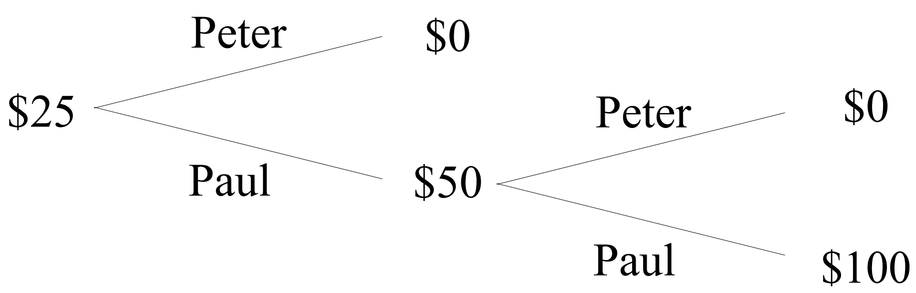
Peter and Paul play for \$100. Paul is behind. Paul needs 2 points to win, and Peter needs only 1.

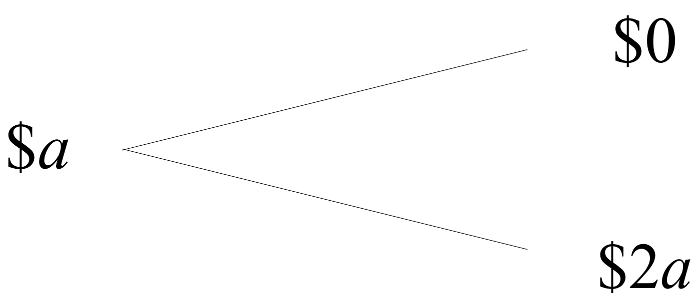


If the game must be broken off, how much of the \$100 should Paul get?

So Paul should get \$25.

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





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.

Game-theoretic law of large numbers (Shafer & Vovk 2001):

Simplest case: binary outcomes, even odds

$\mathcal{K}_0 = 1$.

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

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

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

$\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - \frac{1}{2})$.

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

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

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.

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.

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 $x_n \in \mathbf{X}$.

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 $n = 1, 2, \dots$

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.

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.

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.

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.

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

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

- 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.

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.

EXTRA SLIDES

Think about the result of a gamble that does not risk more than one's initial capital.

The resulting wealth is a nonnegative random variable X with expected value $E(X)$ equal to the initial capital.

Markov's inequality says

$$P(X \geq \frac{E(X)}{\epsilon}) \leq \epsilon.$$

You have probability ϵ or less of multiplying your initial capital by $1/\epsilon$ or more.

Ville proved what is now called *Doob's inequality*, which generalizes Markov's inequality to a sequence of bets.

I.1. Cournotian understanding of conditional probability

Three explanations of conditional probability

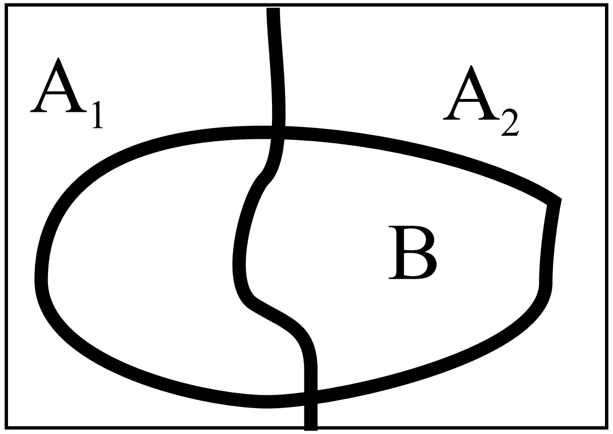
- Abraham De Moivre: Rule of compound probability
- Bruno de Finetti: Coherence
- Antoine-Augustin Cournot: Excluded gambling strategy

The Cournotian explanation extends to Dempster-Shafer.

Conditional probability is a way to update probabilities.

Probability updating

- You have probabilities for A_1, A_2, B, and their intersections.
- You learn A_2 and nothing more.
- What is your new probability for B?



Conventional answer:

$$P(B|A_2) = \frac{P(A_2 \& B)}{P(A_2)}$$

Why?

Rule of compound probability: $P(A_2 \& B) = P(A_2)P(B|A_2)$

Abraham de Moivre
(1667–1754)

De Moivre's proof assumes $P(A)$ is the fair price of a ticket that pays 1 if A happens.

In general,

- you pay $P(A_2)$ for 1 if A_2 happens,
- you pay $P(A_2)x$ for x if A_2 happens, and
- after A_2 happens, you pay $P(B|A_2)$ for 1 if B happens.

To get 1 if $A_2 \& B$ if happens, pay

- $P(A_2)P(B|A_2)$ for $P(B|A_2)$ if A_2 happens,
- then if A_2 happens, pay the $P(B|A_2)$ you just got for 1 if B happens.

De Moivre's argument for the rule of compound probability involves undefined concepts: fair price, knowledge. Many people criticized this aspect of classical probability.

There are two ways of resolving the problem:

- Coherence. De Finetti took the viewpoint of the person offering bets. This person should avoid sure loss.

- Cournot's principle. Take the viewpoint of the person deciding what bets to accept. Prices for gambles are fair when you don't think any strategy will beat them.

Coherence

Suppose you must announce. . .

- $P(A_2)$ and $P(A_2 \& B)$, interpreted as prices at which you are offering to buy or sell £1 tickets on these events.

- $P(B|A_2)$, interpreted as prices at which you will offer to buy or sell £1 tickets on B if A_2 happens.

Bruno de Finetti (1906–1985)

Opponents can make money for sure if you announce $P(A_2 \& B)$ different from $P(A_2)P(B|A_2)$.

- You announce current prices for tickets on all relevant events.

- You also announce the price for B-tickets you will offer in the future if A_2 happens.

Critique: This eliminates explicit talk about “fair price” and “knowledge”. But the plausibility of the assumptions depends on these ideas.

- Who would offer bets to people who know more? Surely the offers are only to people with the same knowledge as you.

- You will not offer to take either side of a bet unless you think the odds are fair given what you and others know.

- We must also assume that if A_2 happens, its happening will be the only new knowledge you and the others have. Otherwise your new price for tickets on B may vary depending on the other new knowledge.

Cournot's principle

Take the viewpoint of the person deciding what bets to accept. Prices being fair means you will not beat them.

More precisely: No matter how you bet, provided you avoid risk of bankruptcy, you will not multiply your capital by a large factor.

The proviso avoid risk of bankruptcy is motivated by Markov's inequality: If X is a nonnegative random variable with expected value 1, then

$$\mathbb{P}\{X \geq K\} \leq \frac{1}{K}.$$

History of the principle

Antoine Cournot
1801–1877

Cournot located the meaning of probability in the idea that an event of tiny probability will not happen.

Many 20th century giants agreed with Cournot:

- Émile Borel (1871–1956)
- Maurice Fréchet (1878–1973)
- Paul Lévy (1886–1971)
- Andrei Kolmogorov (1903–1987)

The game-theoretic form of Cournot's principle says a bettor will not multiply his capital substantially without risking bankruptcy.

More general. Applies to games with limited betting offers.

Cournot's principle is about knowledge and fair price.

Order of play:
   Forecaster gives prices for gambles.
   Skeptic chooses gamble.
   Reality decides outcome.

Skeptic is testing Forecaster. If Skeptic multiplies his capital substantially without risking bankruptcy, we reject Forecaster. His prices are not fair.

Forecaster's prices express Skeptic's knowledge when Skeptic believes he has no strategy for beating these prices.

Laplace: Probability is relative, in part to our ignorance, in part to our knowledge.

Cournot's principle is about the long run

We seldom test a single forecast, because a single bet that does not risk bankruptcy usually has only modest payoffs.

Instead we test series of forecasts.

Cournotians are often called frequentists. But successive forecasts are not necessarily identical and independent.

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 $:= s_n(y_n - p_n)$.

Game-theoretic law of large numbers (Shafer & Vovk 2001):

Simplest case: fair coin (binary outcomes, even odds)

$\mathcal{K}_0 = 1$.

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

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

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

$\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - \frac{1}{2})$.

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.

Cournot's principle is about strategies for Skeptic.

- A strategy for Skeptic tells him how to move on every round, depending on what other players have done so far.

- Cournot's principle: If we pick out a bankruptcy-free strategy $\mathcal{S}$ for Skeptic, we can be confident Reality will move so that $\mathcal{S}$ does not make Skeptic rich.

- This is analogous (equivalent in the fully probabilized case) to saying that if we pick out an event E of small probability in advance, E will not happen. Reality will respect the probabilities.

Abbreviate

these probabilities are in a series for which Cournot's principle holds

to

we will not beat these probabilities.

Cournotian argument for $P(A_2\&B)/P(A_2)$ as the new probability for B

Claim: Suppose we will not beat $P(A_2)$ and $P(A_2\&B)$. Suppose we learn A_2 happens and nothing more. Then we can include $P(A_2\&B)/P(A_2)$ as a new probability for B among the probabilities we will not beat.

Argument: Consider a bankruptcy-free strategy $\mathcal{S}$ against the three probabilities. Let M, positive or negative, be the amount of B tickets $\mathcal{S}$ buys after learning A_2. Let $\mathcal{S}'$ be the strategy against $P(A_2)$ and $P(A_2\&B)$ alone obtained by adding

$$M \text{ tickets on } A_2\&B \qquad \text{and} \qquad -M\frac{P(A_2\&B)}{P(A_2)} \text{ tickets on } A_2 \tag{1}$$

to $\mathcal{S}$'s purchases of these tickets.

1. The tickets in (1) have net cost zero.
2. $\mathcal{S}'$ and $\mathcal{S}$ have the same payoffs.
3. So $\mathcal{S}'$ is bankruptcy-free like $\mathcal{S}$.
4. A_2's happening is the only new information used by $\mathcal{S}$. So $\mathcal{S}'$ uses only the initial information.
5. By hypothesis, $\mathcal{S}'$ does not get rich. So $\mathcal{S}$ does not get rich either.

Conditional probability in practice

In practice, you always learn more than A_2.

- But you make a judgement that the other things you learn do not matter.

- Not we learn A_2 and nothing else.

- Rather we learn A_2 and nothing else that can help us beat Forecaster's prices.

- Probability argument is always in a small world. We judge initial and new knowledge outside this small world irrelevant.

Advantages of Cournot

1. Role of knowledge and fairness explicit.

2. No need to know in advance you will learn A_1 or A_2 and nothing else.

3. Generalizes to case where knowledge is not expressed by additive probabilities.

Disadvantages of Cournot

1. Not inside the mathematics. Uses soft concepts such as knowledge.

2. Repetition required.

Cournot's greatest advantage is its generality.

- *Probability and Finance* generalizes unconditional limit theorems to the case of limited betting offers.

- probabilityandfinance.com Working Paper #3 generalizes the argument for conditioning to Walley's imprecise probabilities.

- This talk generalizes the argument for conditioning to Dempster-Shafer.

I.2. Probabilistic independence

- De Moivre: The happening of A does not change the probability of B.

- de Finetti: The happening of A does not change the price at which you offer to buy or sell tickets on B.

- Cournot: The happening of A does not change the price you will not beat for tickets on B. In practice: Neither the happening of A nor your other new information will help you beat the probability for B.

Uncorrelated vs. independent

1. **Y uncorrelated with X** means the event $\mathbf{X} = x$ does not affect the price of $\mathbf{Y}$.

2. **Y independent of X** means the event $\mathbf{X} = x$ does not affect probabilities for $\mathbf{Y}$.

$\mathbf{Y}$ can be priced without being fully probabilized. So uncorrelatedness is the wider notion.

I.3. 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.

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

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

- 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? $\qquad \Omega = \{0, \pounds 1, \pounds 5\}$

- Bill says "Glenn paid £5."

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

- New beliefs:

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

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

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? $\qquad \Omega = \{0, \pounds 1, \pounds 5\}$

- Tom says "Glenn paid £10."

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

We have a problem: Hearing what Tom said does help us beat the probabilities concerning his precision. We now know he is not precise.

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? $\Omega = \{0, £1, £5\}$

- 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.

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 $\mathbf{X}_2$.

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.

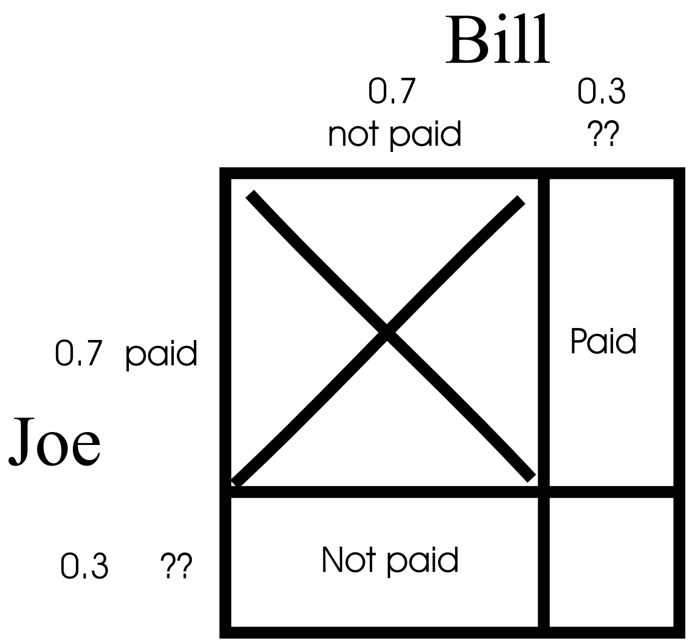
$$\mathbb{B}(\text{paid}) = \frac{0.21}{0.51} = 0.41 \qquad \mathbb{B}(\text{not paid}) = \frac{0.21}{0.51} = 0.41$$

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.

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 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 $\{(x, \omega) | \omega \in \Gamma(x)\}$.

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

Retirement dinner at Harvard, May 2005.

Part II. Richer understanding of statistical modeling

- The perfect-information protocol for probability
- Mathematical statistics departs from probability by standing outside the protocol.
- Classical example: the error model
- Parametric modeling
- Dempster-Shafer modeling

The perfect-information protocol for probability

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, N$:

Forecaster announces prices for various payoffs.

Skeptic decides which payoffs to buy.

Reality determines the payoffs.

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

Mathematical statistics departs from probability by standing outside the protocol in various ways.

Forecaster, Skeptic, and Reality see each others' moves, but we do not.

- Usually Skeptic is not really there. We can take this player's role if we see the other players' moves.

- Perhaps we do not see Forecaster's moves. We infer what we can about them from Reality's moves. Or perhaps it is our job to make the forecasts.

- Perhaps we see only a noisy or distorted version of Reality's moves. We infer what we can about them from Forecaster's moves.

Classical example: errors in measurement

A measuring instrument makes errors obeying some probability distribution.

You do not see the errors $e_1, \ldots, e_N$.

You only see measurements $x_1, \ldots, x_N$, where

$$x_n = \theta + e_n.$$

How do you make inferences about θ?

Parametric modeling. The parametric model P_θ is a class of strategies for Forecaster.

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, N$:
Forecaster gives prices p_n following a strategy P_θ.
Skeptic makes purchases M_n following a strategy $\mathcal{S}_\theta$.
Reality announces y_n.
$\mathcal{K}(\theta)_n := \mathcal{K}(\theta)_{n-1} +$ Skeptic's net gain or loss.
Cournot's principle: Not all the $\mathcal{K}(\theta)$ get very large.

We see y_n, and we know the strategies, but we do not know θ and do not see p_n and M_n.

If all the $\mathcal{K}(\theta)_N \geq K$ for all θ, we reject the model. Otherwise, those θ for which $\mathcal{K}(\theta)_N < K$ form a $1 - \frac{1}{K}$ confidence interval for θ.

Errors in measurement as a parametric model

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, N$:
  Forecaster announces (but not to us) the price θ.
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $y_n \in \mathbb{R}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \theta)$.
Winner: Skeptic wins if $\mathcal{K}_n$ is never negative and either $\mathcal{K}_N \geq K$ or $|\overline{y} - \theta| < \epsilon$, where $\overline{y} := \sum_{n=1}^{N} y_n$.

According to *Probability and Finance* (p. 125), if $N \geq KC^2/\epsilon^2$ and Reality is constrained to obey $y_n \in [\theta - C, \theta + C]$, then Skeptic has a winning strategy.

Dempster-Shafer modeling. We see the moves by Forecaster and Skeptic, but not those by Reality.

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, N$:
  Forecaster announces prices p_n.
  Skeptic makes purchases M_n.
  Reality announces (but not to us) x_n.
  $\mathcal{K}_n := \mathcal{K}_{n-1} +$ Skeptic's net gain or loss.
Cournot's principle: With probability $1 - \frac{1}{K}$, $\mathcal{K}_N < K$.

We see only $y_n = \omega(x_n)$ for some function ω. The mapping

$$\Gamma(x_1, \ldots, x_N) = \{\omega | \omega(x_n) = y_n, n = 1, \ldots, N\}$$

allows us to transfer the probabilities about $x_1, \ldots, x_N$ to beliefs about ω.

Errors in measurement as a Dempster-Shafer model.

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, N$:
    Forecaster announces the standard Gaussian distribution.
    Skeptic chooses a function f_n of the payoff x_n.
    Reality announces (but not to us) $x_n \in \mathbb{R}$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(x_n) - \mathbb{E}(f_n(x_n))$.

We see only $y_n = \omega + x_n$ for some $\omega \in \mathbb{R}$. Conditioning on the configuration $x_1 - \overline{x}, \ldots, x_N - \overline{x}$, we get probabilities for ω.

Functions of configuration can be used to test the model.

Part III. Extended examples. Elementary examples are inadequate guides for dealing with the complications that arise in real problems.

I would like to develop some examples over the next few years:

- Combining information in continuous auditing

- Fusing competing computational models (weather forecasting, etc.)

- Fusing information in electronic defense (deciding who is monitoring you with radar, etc.)

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.

Definitions

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

Game-theoretic processes and martingales

- A real-valued function on the situations is a *process*.

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

- A strategy for Skeptic, together with a particular initial capital for Skeptic, also defines a process: Skeptic's *capital process* $\mathcal{K}(y_1 \ldots y_n)$.

- We also call a capital process for Skeptic a *martingale*.

Notation for Martingales

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

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

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

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

The capital processes that begin with zero form a linear space, 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

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

—LHS is the convex combination of two martingales that each begin with capital 1.

—RHS is the martingale produced by the same convex combination of strategies, also beginning with capital 1.

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

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

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

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

- You can prove this by induction on

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

In game-theoretic probability, you can usually get an infinite convex combination of martingales, but you have to check on the convergence of the infinite convex combination of strategies. In a sense, this explains the historical confusion about countable additivity in measure-theoretic 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 have only a few prices. Instead of giving them at the outset, maybe your make them up as 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 Probability

1. Pascal & Ville

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

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

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

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

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).

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 $n = 1, 2, \dots$
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.)

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.

This concept of test generalizes the standard stochastic concept.

Stochastic setting:

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

Standard concept of statistical test:

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

In 1939, Jean Ville showed that in the stochastic setting, the standard concept is equivalent to a strategy for Skeptic.

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 $n = 1, 2, \dots$
  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 $n = 1, 2, \dots$
  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 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.

More talks in Paris

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

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

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

Standard stochastic concept of statistical test:

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

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

- Given an event of probability less than $1/C$, there is a strategy for Skeptic that turns \$1 into \$$C$ without risking bankruptcy.
- Given a strategy for Skeptic that starts with \$1 and does not risk bankruptcy, the probability that it turns \$1 into \$$C$ or more is no more than $1/C$.

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

Continuity rules out Dawid's counterexample

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