Workshop on Game-Theoretic Probability and Related Topics

Tokyo University February 28, 2008

Using game-theoretic probability for probability judgment

Glenn Shafer

- For 170 years, people have asked whether probabilities are objective or subjective.

- Game-theoretic probability asks a simpler question: Is there a repetitive structure?

- For probability judgement outside repetitive structures we need judgements of independence.

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

- Sample chapters and working papers at www.probabilityandfinance.com.

Abstract

The success of defensive forecasting shows that the crucial question is not whether there is a valid objective model but simply whether there is an agreed-on repetitive structure for the questions to be answered and the data for answering them.

Once we place the current example in a sequence of other examples, defensive forecasting can produce valid probability forecasts-valid insofar as they resist statistical tests—without any assumptions.

But if no single sequence of previous examples imposes itself, then we must go outside probability theory to weigh evidence.

- Theory To prove a probabilistic prediction, construct a betting strategy that makes you rich if the prediction fails.

- Testing To test a probabilistic theory, bet against its predictions.

- Forecasting To make probability forecasts in a repetitive structure, construct a forecasting strategy that defeats all reasonable betting strategies.

- Probability judgement Judgements of evidential independence are needed to extend reasoning to non-repetitive situations (Dempster-Shafer, etc.).

Part I. Game-theoretic probability

Blaise Pascal
(1623–1662)

Probability is about betting.

Antoine Cournot
(1801–1877)

Events of small probability do not happen.

Jean Ville
(1910–1988)

Pascal + Cournot:

If the probabilities are right, you don't get rich.

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)

This is more basic than frequentism.

Émile Borel
1871–1956

Inventor of measure theory.

Minister of the French navy in 1925.

Borel: 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 $\mathrm{P}(A)$ 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 this principle:

> You will not multiply the capital you risk by a large factor.

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

We call both principles Cournot's principle.

Suppose you gamble without risking more than your initial capital.

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

Markov's inequality says

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

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

Game-theoretic probability generalizes classical probability to the case of limited betting offers (less than a probability distribution) by taking the inability to multiply initial capital as basic.

Perfect-information protocol for predicting and testing even without a probability model

$\mathcal{K}_0 = 1$.

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.

In Ville's theory, Forecaster gives prices based on a probability distribution. He uses conditional probabilities for Reality's next move given her past moves.

In Vovk's generalization, (1) Forecaster does not necessarily use a known probability distribution, and (2) he may give less than a probability distribution for Reality's next move.

Ville's strong law of large numbers.

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

$\mathcal{K}_0 = 1$.

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.

Ville's strategy

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

Ville suggested the strategy

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

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.

Ville's more general game.

Ville started with a probability distribution for P for $y_1, y_2, \ldots$.
The conditional probability for $y_n = 1$ given $y_1, \ldots, y_{n-1}$ is not necessarily 1/2.

> $\mathcal{K}_0 := 1$.
> FOR $n = 1, 2, \ldots$:
> Skeptic announces $s_n \in \mathbb{R}$.
> Reality announces $y_n \in \{0, 1\}$.
> $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n \left(y_n - \mathrm{P}(y_n = 1 | y_1, \ldots, y_{n-1})\right)$.
>
> Skeptic wins if
> (1) $\mathcal{K}_n$ is never negative and
> (2) either $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathrm{P}(y_i = 1 | y_1, \ldots, y_{i-1})\right) = 0$
> or $\lim_{n\to\infty} \mathcal{K}_n = \infty$.

Theorem Skeptic has a winning strategy.

Vovk's generalization: Replace P with a forecaster.

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

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

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

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

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

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

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.

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 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 be zero close to $p = 0.5$, and Forecaster can set p_n equal to this point.

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

Theorem Forecaster can guarantee that Skeptic never makes money.

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

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

Theorem. Forecaster can guarantee that Skeptic never makes money.

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

Theorem Forecaster can guarantee that Skeptic never makes money.

Proof:

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

TWO APPROACHES TO FORECASTING

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

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

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

We can always give probabilities with good calibration and resolution.

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*.

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

Example of a resulting strategy for Skeptic:

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

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

Skeptic's strategy:

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

Forecaster's strategy: Choose p_n so that

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

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

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

Objective vs. 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.

Part 2. Cournotian justification of Bayesian updating

We update probabilities using the rule

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

- In 1738, De Moivre justified this rule under the classical betting interpretation.
- In 1937, de Finetti recast the justification in terms of his betting interpretation.
- I will recast it in Cournotian terms.

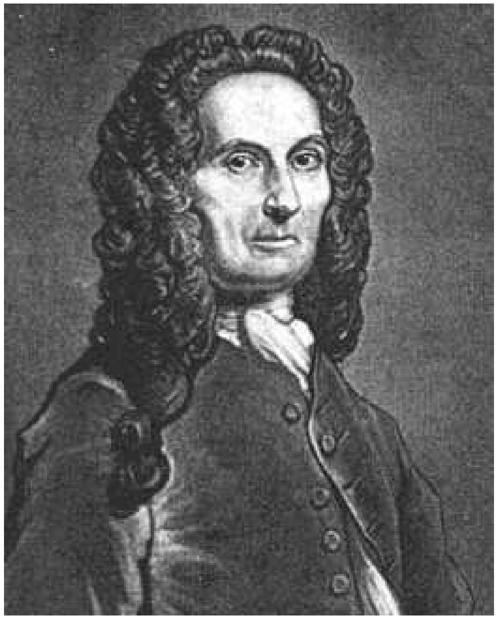
Three betting interpretations:

- 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 $P(B|A) = P(A\&B)/P(A)$

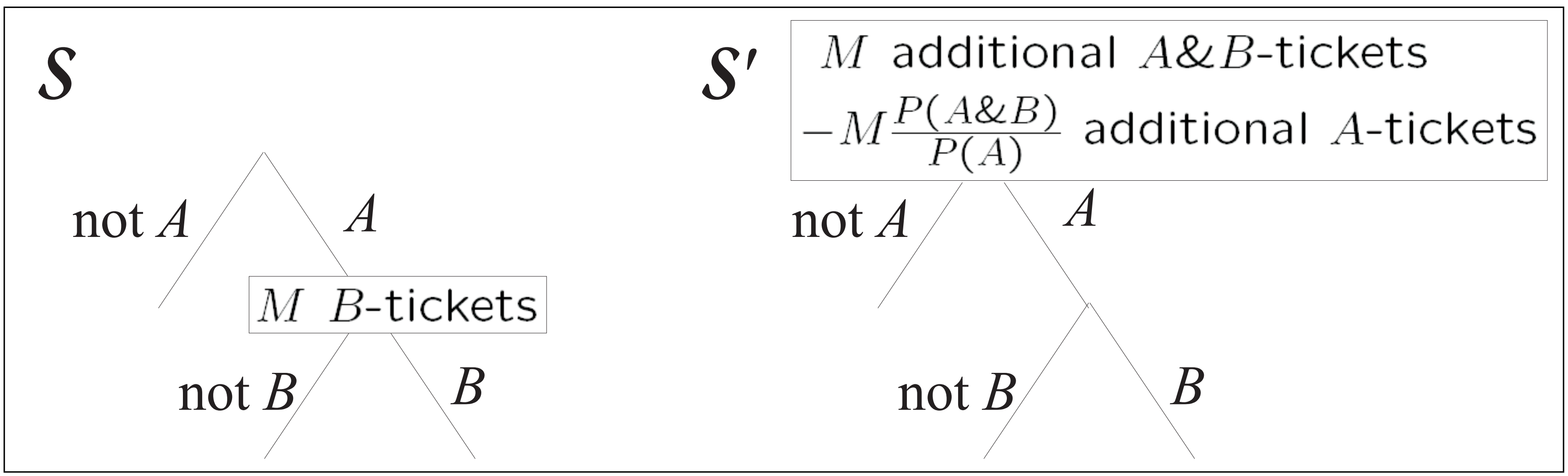
Claim: Suppose $P(A)$ and $P(A\&B)$ cannot be beat. Suppose we learn A happens and nothing more. Then we can include $P(A\&B)/P(A)$ as a new probability for B among the probabilities that cannot be beat.

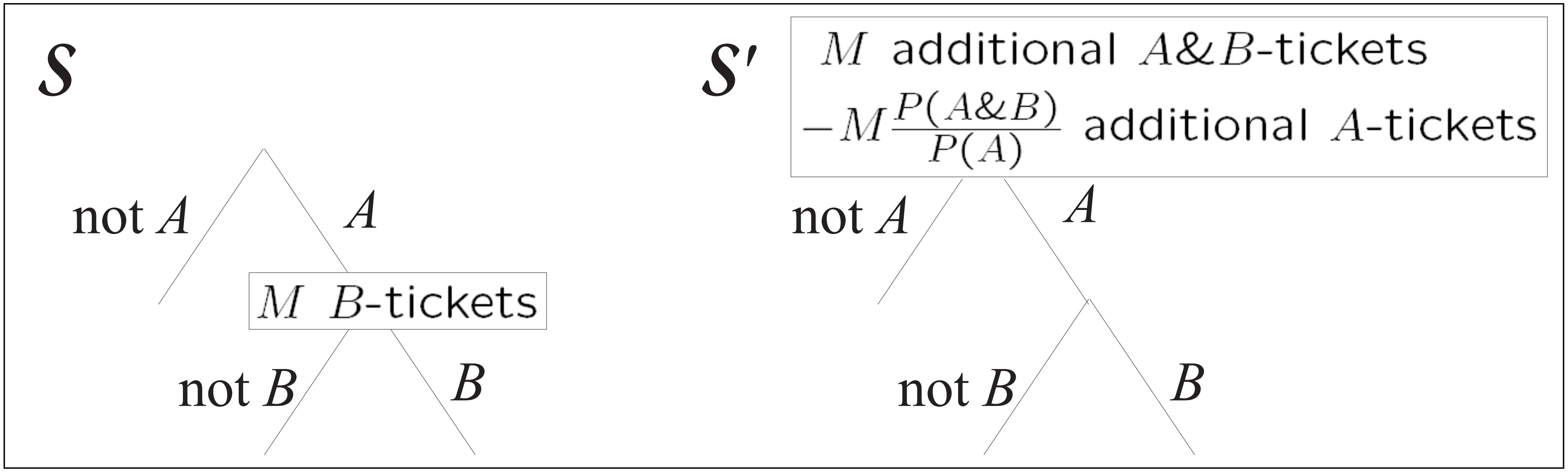
Structure of proof:

- 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.

Part 3. Cournotian justification of Dempster-Shafer operations

Dempster-Shafer has three fundamental operations:

- Transferring belief
- Independence
- Conditioning (same as Bayesian updating)

All three are justified by a judgement that certain information does not help us beat certain probabilities.

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.

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.

EXTRA SLIDES

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

- 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.

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.

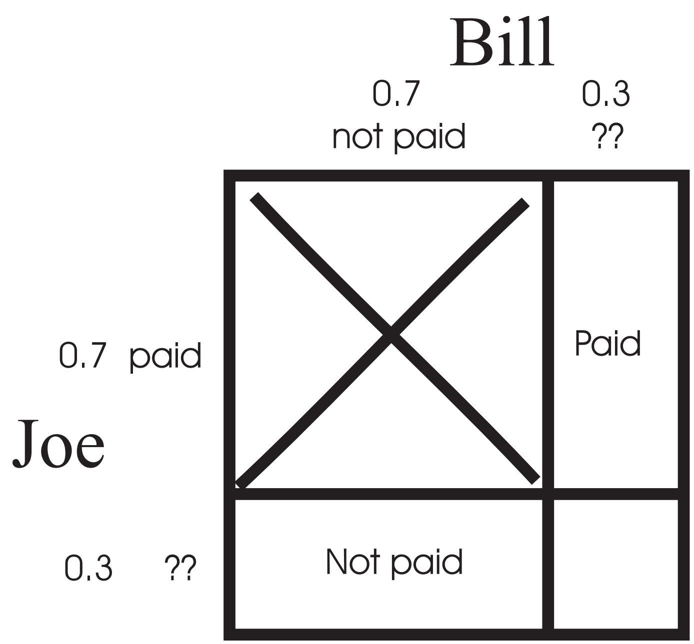
$$\mathbb{B}(\text{paid}) = \frac{0.21}{0.51} = 0.41 \qquad \mathbb{B}(\text{not paid}) = \frac{0.21}{0.51} = 0.41$$

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 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

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