

Testing with non-negative martingales, measure-theoretically and game-theoretically

Glenn Shafer

March 21, 2013
Center for Research in Mathematics
Guanajuato

Goals of this talk

- Introduce an idea well known in algorithmic randomness:
using a nonnegative martingale as a dynamic measure of evidence.
(In the culture of mathematical statistics, testing is not dynamic.)
- How to **adjust** a p-value.
- How to **insure** against losing all the money you have made so far.
- Introduce **game-theoretic probability** (as alternative to measure-theoretic probability) and illustrate its power.

Testing with non-negative martingales

Statistical testing can be framed as a repetitive game between two players, **Forecaster** and **Skeptic**.

1. On each round, Forecaster sets prices for various gambles, and Skeptic chooses which gambles to make.
2. **The factor by which Skeptic multiplies the capital he puts at risk** is a measure of his evidence against Forecaster.
3. This factor can go up and then back down. If you report **the maximum so far instead of the current value**, you are exaggerating the evidence against Forecaster.
4. The exaggeration corresponds to using a p-value rather than a Bayes factor for testing. There are ways to **correct the exaggeration**, which also apply to p-values.

References

- Test martingales, Bayes factors, and p-values
Shafer, Shen, Vereshchagin, & Vovk
Statistical Science 26:84-101, 2011
- Insuring against loss of evidence in game-theoretic probability
Dawid, de Rooij, Shafer, Shen, Vereshchagin, & Vovk
Statistics and Probability Letters 81:157-162, 2011
- Probability-free pricing of adjusted American lookbacks
Dawid, de Rooij, Grünwald, Koolen, Shen, Vereshchagin, & Vovk
Working Paper 37, www.probabilityandfinance.com, August 2011

The modern concept of a **martingale** was introduced into probability theory by Jean André Ville in 1939.

Ville's definition:

A **martingale** is a capital process – i.e., a process whose values are the current wealth for a betting strategy in a sequential game where you can buy any payoff for its current expected value.

Equivalent definition: A process L_1, L_2, \dots is a **martingale** if and only if $E_n(L_{n+1}) = L_n$ for all n .



A **martingale** is the capital process for a betting strategy.

Therefore, a **non-negative martingale** is the capital process for a betting strategy that does not risk more than its initial capital – i.e., does not risk bankruptcy.

Consider non-negative processes L_1, L_2, \dots with $L_1 = 1$. We have two ways of characterizing those that are martingales:

1. They are capital processes for strategies that start with unit capital and do not risk bankruptcy.
2. $L_1 = 1$ and $E_n(L_{n+1}) = L_n$ for all n .

Now a third way:

3. They are likelihood ratios.

Exercise 1: For a sequence of binary outcomes, verify that a likelihood ratio Q/P is a non-negative martingale with respect to P , with initial value 1.

Solution:

$$\begin{aligned}
 E \left(\frac{Q(y_1, \dots, y_n, Y_{n+1})}{P(y_1, \dots, y_n, Y_{n+1})} \middle| y_1, \dots, y_n \right) \\
 &= \frac{Q(y_1, \dots, y_n, 1)}{P(y_1, \dots, y_n, 1)} P(Y_{n+1} = 1 | y_1, \dots, y_n) + \frac{Q(y_1, \dots, y_n, 0)}{P(y_1, \dots, y_n, 0)} P(Y_{n+1} = 0 | y_1, \dots, y_n) \\
 &= \frac{Q(y_1, \dots, y_n, 1)}{P(y_1, \dots, y_n, 1)} \frac{P(y_1, \dots, y_n, 1)}{P(y_1, \dots, y_n)} + \frac{Q(y_1, \dots, y_n, 0)}{P(y_1, \dots, y_n, 0)} \frac{P(x_1, \dots, x_n, 0)}{P(x_1, \dots, x_n)} \\
 &= \frac{Q(y_1, \dots, y_n, 1)}{P(y_1, \dots, y_n)} + \frac{Q(y_1, \dots, y_n, 0)}{P(y_1, \dots, y_n)} \\
 &= \frac{Q(y_1, \dots, y_n)}{P(y_1, \dots, y_n)}
 \end{aligned}$$

The initial value of Q/P , before Y_1 is observed, is $Q(\square)/P(\square) = 1/1 = 1$.

Exercise 2: For a sequence of binary outcomes, verify that if L is a non-negative martingale under P with initial value 1, then LP is a probability distribution.

Solution: A function R is a binary probability distribution if and only if $R(\square) = 1$, $R(1) + R(0) = 1$, $R(y_1, 1) + R(y_1, 0) = R(y_1)$ for $y_1 = 1$ and $y_1 = 0$, etc. So we merely need to verify that

$$\begin{aligned} L(y_1, \dots, y_n)P(y_1, \dots, y_n) &= E(L(y_1, \dots, y_n, Y_{n+1})|y_1, \dots, y_n)P(y_1, \dots, y_n) \\ &= L(y_1, \dots, y_n, 1)\frac{P(y_1, \dots, y_n, 1)}{P(y_1, \dots, y_n)}P(y_1, \dots, y_n) \\ &\quad + L(y_1, \dots, y_n, 0)\frac{P(y_1, \dots, y_n, 0)}{P(y_1, \dots, y_n)}P(y_1, \dots, y_n) \\ &= L(y_1, \dots, y_n, 1)P(y_1, \dots, y_n, 1) + L(y_1, \dots, y_n, 0)P(y_1, \dots, y_n, 0) \end{aligned}$$

It follows that L is a likelihood ratio. Namely, the likelihood ratio LP/P .

Definition: A **test martingale** is a nonnegative martingale with initial value = 1.

- In game-theoretic probability, **test martingale** = capital process for a strategy that starts with unit capital and does not risk bankruptcy.
- In measure-theoretic probability, **test martingale** = likelihood ratio.

A test martingale can be used as a **statistical test**: if it gets very large, you reject the probabilistic null hypothesis (or the null hypothesis of market efficiency).

- In game-theoretic probability,
test martingale = capital process for a strategy that starts with unit capital and does not risk bankruptcy.
- In measure-theoretic probability,
test martingale = likelihood ratio.

The game-theoretic view is more general than measure-theoretic view

It applies to partial models and financial markets with no pretense about a *meaningful* full probability model.

Example: Testing whether a coin is fair

Notation

- Binary random variables x_1, x_2, \dots taking values in $\{0, 1\}$
- P_θ = distribution of x_1, x_2, \dots when they are IID with prob θ of $x_1 = 1$
- Null hypothesis: $\theta = 1/2$

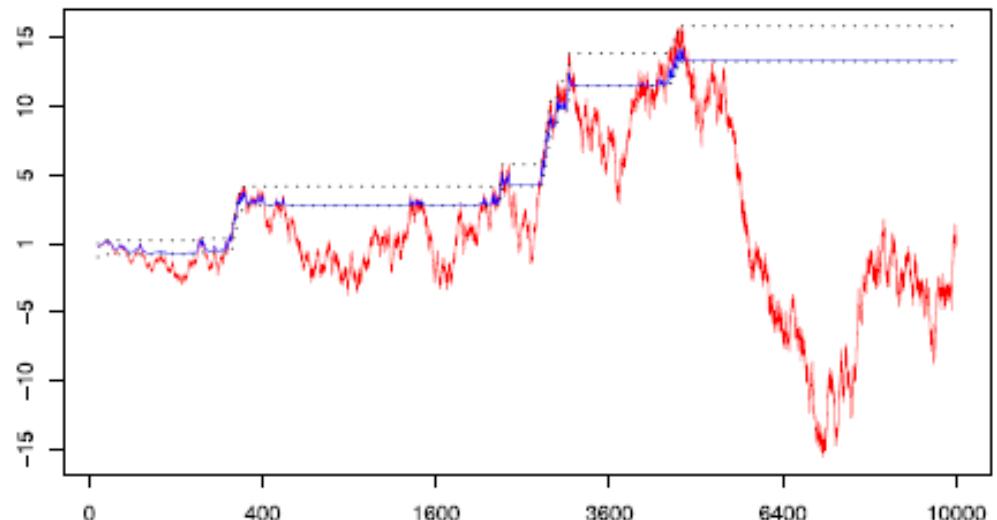
Two tests of $\theta = 1/2$

1. Against $\theta = 3/4$
2. Against $\theta \neq 1/2$, represented by the uniform distribution for θ

To test $\theta = 1/2$ against $\theta = 3/4$, use likelihood ratio

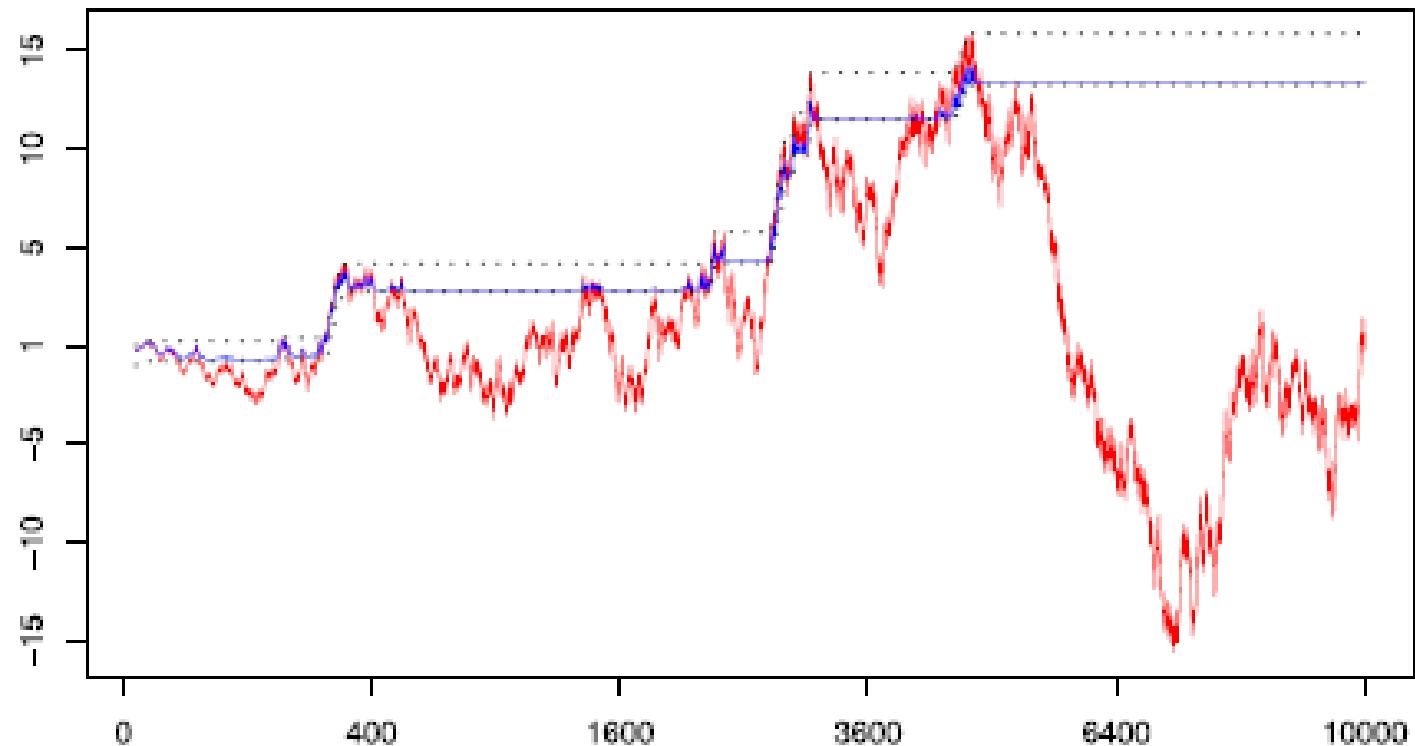
$$X_t := \frac{P_{3/4}(x_1, \dots, x_t)}{P_{1/2}(x_1, \dots, x_t)} = \frac{(3/4)^{k_t} (1/4)^{t-k_t}}{(1/2)^t} = \frac{3^{k_t}}{2^t}$$

where k_t is the number of 1s in x_1, \dots, x_t .



- Red line is likelihood ratio for 10,000 trials, with true $\theta = \ln 2 / \ln 3 \approx 0.63$.
- Both axes are logarithmic.
- Horizontal axis: number of observations so far.
- Vertical axis: powers of 10.
- Likelihood ratio varies wildly. Continued indefinitely, it would be unbounded in both directions.

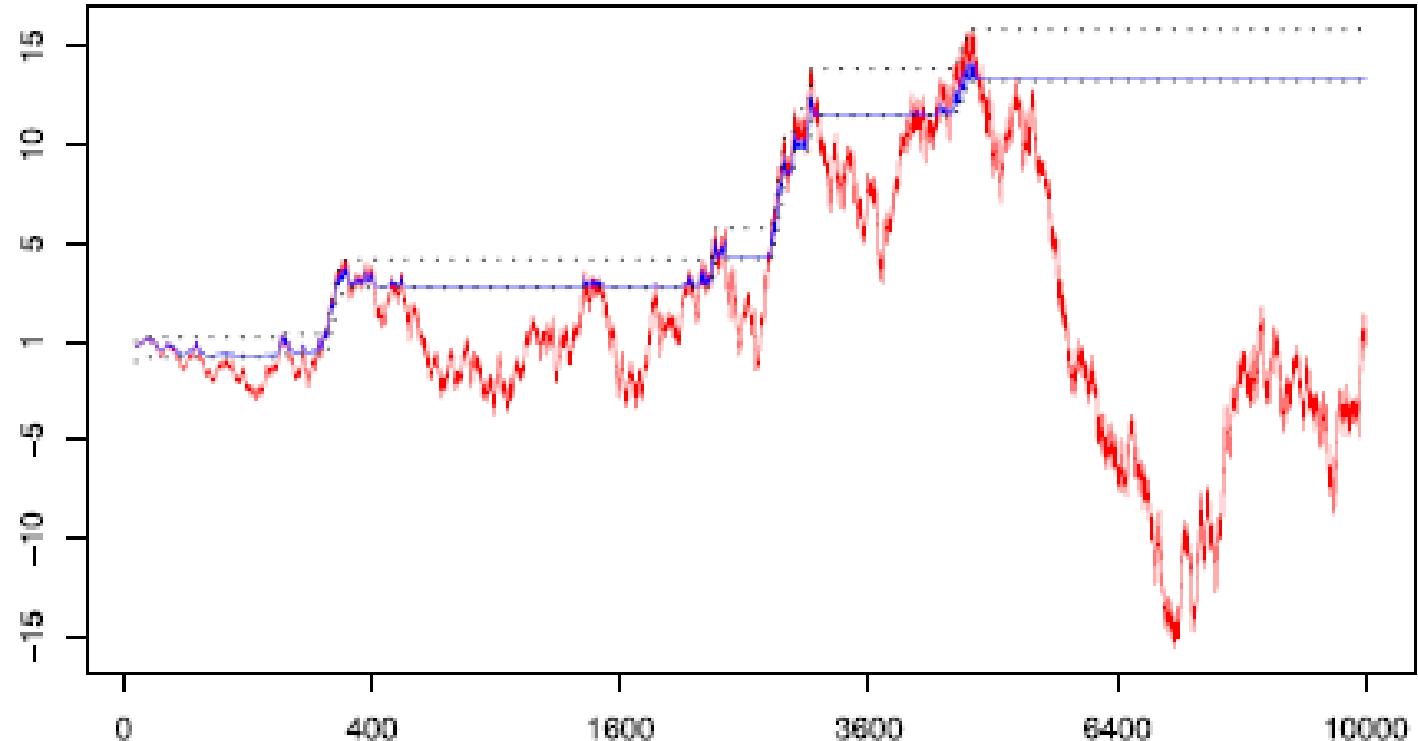
Testing $\theta = 1/2$ against $\theta = 3/4$ over 10,000 trials



- Red line = X_t = likelihood ratio (generated with true $\theta = \ln 2 / \ln 3 \approx 0.63$)
- Upper dotted line = $\max_{n=1,\dots,t} X_n$ = maximum likelihood ratio so far
- Lower dotted line = $\max_{n=1,\dots,t} 0.1X_n^{0.9}$
- Blue line = a test martingale Y_t satisfying $Y_t \geq \max_{n=1,\dots,t} 0.1X_n^{0.9}$

Lower dotted line = $\max_{n=1,\dots,t} 0.1X_n^{0.9}$

Blue line = test martingale Y_t satisfying $Y_t \geq \max_{n=1,\dots,t} 0.1X_n^{0.9}$



Definition. An increasing $f : [1, \infty) \rightarrow [0, \infty)$ is a *martingale calibrator* if for any test martingale X_t there exists a test martingale Y_t such that $Y_t \geq f(\max_{n=1,\dots,t} X_n)$ for all t .

Fact The function $f(y) := 0.1y^{0.9}$ is a martingale calibrator.

F is a martingale calibrator iff it satisfies $\int_1^\infty \frac{F(y)}{y^2} dy \leq 1$.

Admissible (best you can do), if the inequality is equality.

$$F(y) := \alpha y^{1-\alpha}$$

Examples

$$F(y) := \begin{cases} \alpha(1 + \alpha)^\alpha \frac{y}{\ln^{1+\alpha} y} & \text{if } y \geq e^{1+\alpha} \\ 0 & \text{otherwise,} \end{cases}$$

F is a martingale calibrator iff it satisfies $\int_1^\infty \frac{F(y)}{y^2} dy \leq 1$.

Admissible (best you can do), if the inequality is equality.

- This can be proven game-theoretically or measure-theoretically.
- The game-theoretic proof is more intuitive.
- The game-theoretic proof tells you how to construct the test martingale Y .

Main idea of game-theoretic proof.

- For every threshold u , consider the strategy that holds 1 unit of X , selling it when Investor's capital reaches (or exceeds) u .
- This achieves the capital $F_u(y) := u \mathbf{1}_{\{y \geq u\}}$.
- Now mix these strategies according to some probability measure P on u .
- It is easily shown that every increasing function F satisfying $\int_1^\infty \frac{F(y)}{y^2} dy = 1$. can be represented as such a mixture: $F(y) = \int_1^\infty F_u(y) P(du) = \int_1^y u P(du)$.

Lemma 2.2. *An increasing right-continuous function $F : [1, \infty) \rightarrow [0, \infty)$ satisfies*

$$\int_1^\infty \frac{F(y)}{y^2} dy = 1$$

if and only if

$$F(y) = \int_{[1,y]} u P(du), \quad \forall y \in [1, \infty),$$

holds for some probability measure P on $[1, \infty)$.

Idea of proof

$$\begin{aligned} \int_{[1, \infty)} \frac{F(y)}{y^2} dy &= \int_{[1, \infty)} \int_{[1,y]} \frac{u}{y^2} P(du) dy \\ &= \int_{[1, \infty)} \int_{[u, \infty)} \frac{u}{y^2} dy P(du) = \int_{[1, \infty)} P(du) = 1. \end{aligned}$$

Main theorem in *Test martingales, Bayes factors, and p-values*

Theorem 4. 1. An increasing function $f : [1, \infty) \rightarrow [0, \infty)$ is a martingale calibrator if and only if

$$\int_0^1 f(1/x)dx \leq 1. \quad (10)$$

2. Any martingale calibrator is dominated by an admissible martingale calibrator.
3. A martingale calibrator is admissible if and only if it is right-continuous and

$$\int_0^1 f(1/x)dx = 1. \quad (11)$$

We can calibrate the capital, which is larger than 1, or the corresponding p-value, which is between 0 and 1.

So we have two different ways of saying what functions work as calibrators.

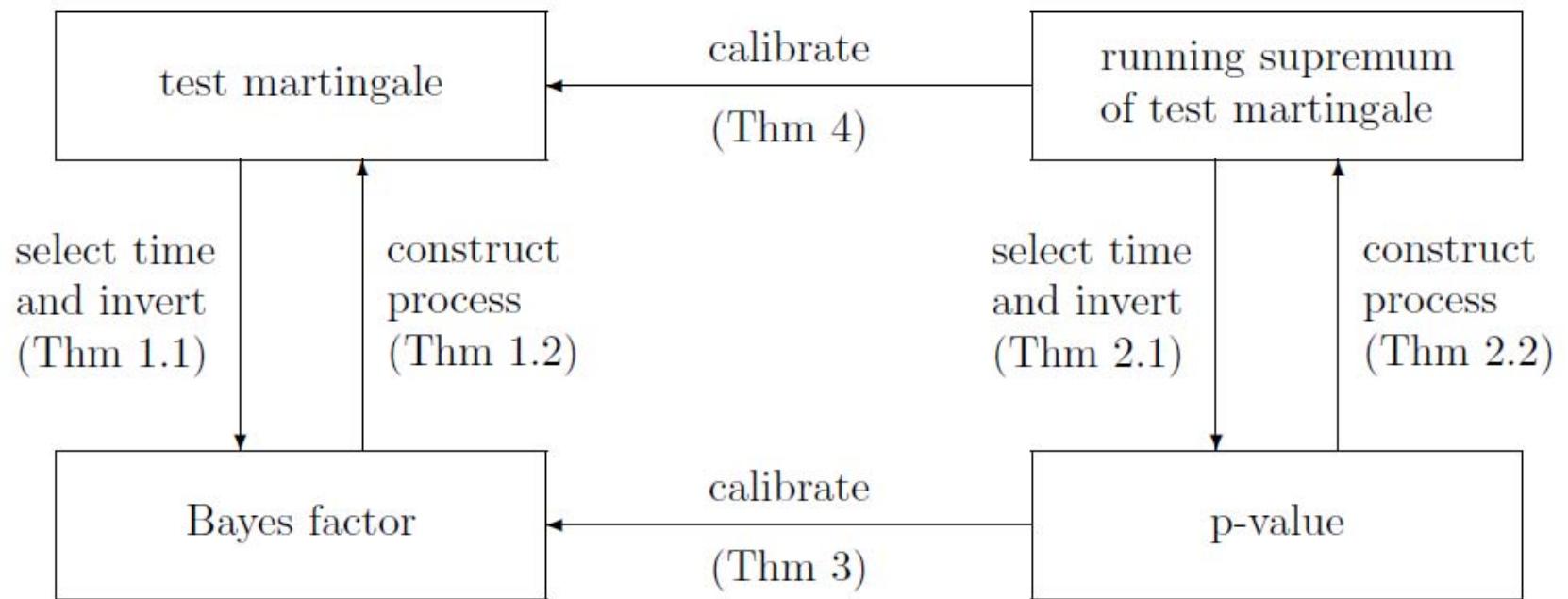
In game-theoretic paper,

$$\int_1^\infty \frac{F(y)}{y^2} dy \leq 1. \quad (2)$$

In measure-theoretic paper,

$$(6) \quad \int_0^1 \frac{dx}{f(x)} \leq 1.$$

Exercise 2: Prove that (2) holds for F if and only if (6) holds for f , where $f(x) = 1/F(1/x)$.



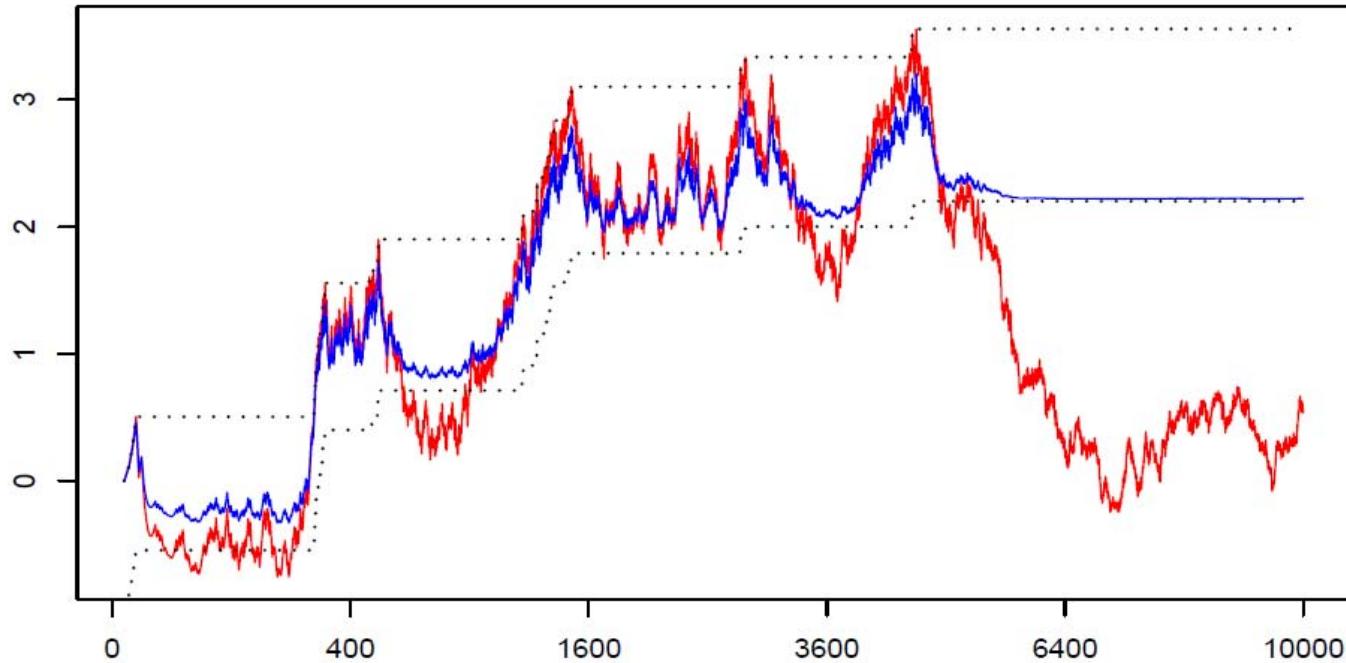
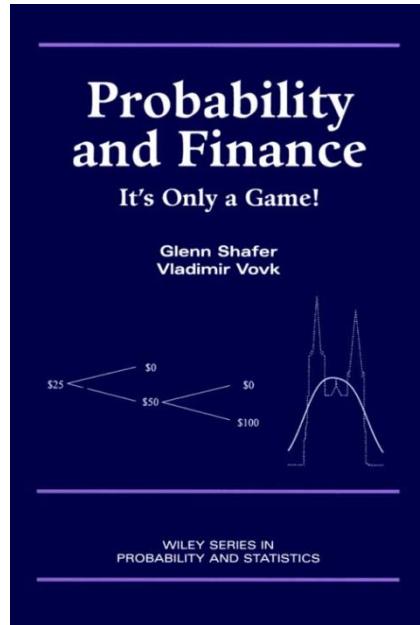


FIG 3. A realization over 10,000 trials of the likelihood ratio for testing $\theta = 1/2$ against the probability distribution Q obtained by averaging P_θ with respect to the uniform distribution for θ . The vertical axis is again logarithmic. As in Figure 2, the oscillations would be unbounded if trials continued indefinitely.

$$X_t := \frac{Q(x_1, \dots, x_t)}{P_{1/2}(x_1, \dots, x_t)} = \frac{\int_0^1 \theta^{k_t} (1-\theta)^{t-k_t} d\theta}{(1/2)^t} = \frac{k_t! (t-k_t)! 2^t}{(t+1)!}$$

The 0s and 1s in the sequence $x_1, \dots, x_{10,000}$ are independent with a probability for $x_t = 1$ that slowly converges to $1/2$: $\frac{1}{2} + \frac{1}{4}\sqrt{\ln t/t}$.



Glenn Shafer
Vladimir Vovk

Wiley
2001

2nd edition in preparation

1. Game theoretic probability
2. Game theoretic upper and lower probability
3. Defensive forecasting
4. Probability judgment

- **Probabilities derive from betting offers.**
Not from the measure of sets.
- **Test probabilities by betting.**
Refute alleged probabilities by making money.
- **Prove “probability one” by betting strategy.**
 - Do not say that the property will fail on a set of measure zero.
 - Say that its failure will mean the success of a betting strategy.

Probabilities derive from **betting** offers.

The offers may determine less than a probability distribution.

1. The stock market gives a price but not a probability distribution for tomorrow's value of a share of Google.
2. A forecaster who gives a probability for rain tomorrow every day for a year does not give a joint probability distribution for the 365 outcomes.

In such cases, we get only upper and lower probabilities.

- In the game-theoretic framework, it can be shown that **good probability forecasting is possible**.
- Once a sequence of events is fixed, **you can give probabilities that pass statistical tests**.
- The forecasting **defends** against the tests.

Defensive forecasting continued

- **Once a sequence of events is fixed**, you can give probabilities that pass statistical tests.
- **The only role of the observer** is to place the event in a sequence.
- Advance knowledge is not needed.
- The sequence need not be “iid”; this concept is not even defined.

1. Game theoretic probability
2. Game theoretic upper and lower probability
3. Defensive forecasting
4. Probability judgment

Jeyzy Neyman's **inductive behavior**

A statistician who makes predictions with 95% confidence has two goals:

be informative
be right 95% of the time

Question: Why isn't this good enough for **probability judgment?**

Answer: Because two statisticians who are right 95% of the time may tell the court different and even contradictory things.

They are placing the current event in different sequences.

Probability judgment continued

- Good probability forecasting requires a sequence.
- It does not require repetition of the “same” event.
- Each event remains unique.

Probability judgment: Assessment of the relevance or irrelevance of experience from different sequences for which we have good probability forecasters.

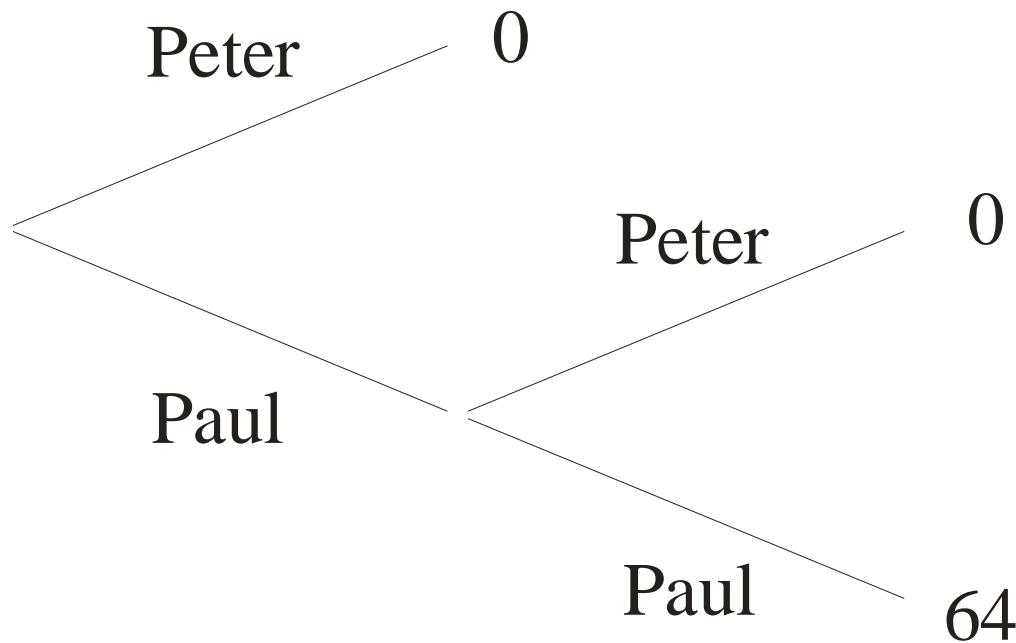
1. Game theoretic probability

The contrast between measure-theoretic & game-theoretic probability began in 1654.

Pascal = game theory

Fermat = measure theory

Pascal's question to Fermat in 1654



Paul needs 2 points to win.

Peter needs only one.

If the game must be broken off,
how much of the stake should
Paul get?



Blaise Pascal (1623–1662).

Fermat's answer (measure theory)

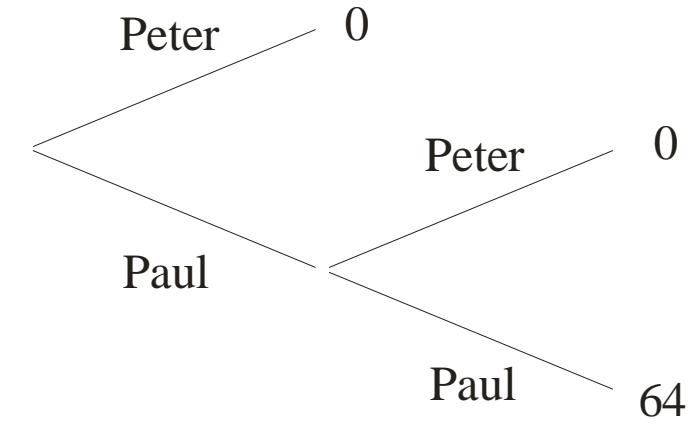
Count the possible outcomes.

Suppose they play two rounds. There are 4 possible outcomes:

1. Peter wins first, Peter wins second
2. Peter wins first, Paul wins second
3. Paul wins first, Peter wins second
4. Paul wins first, Paul wins second

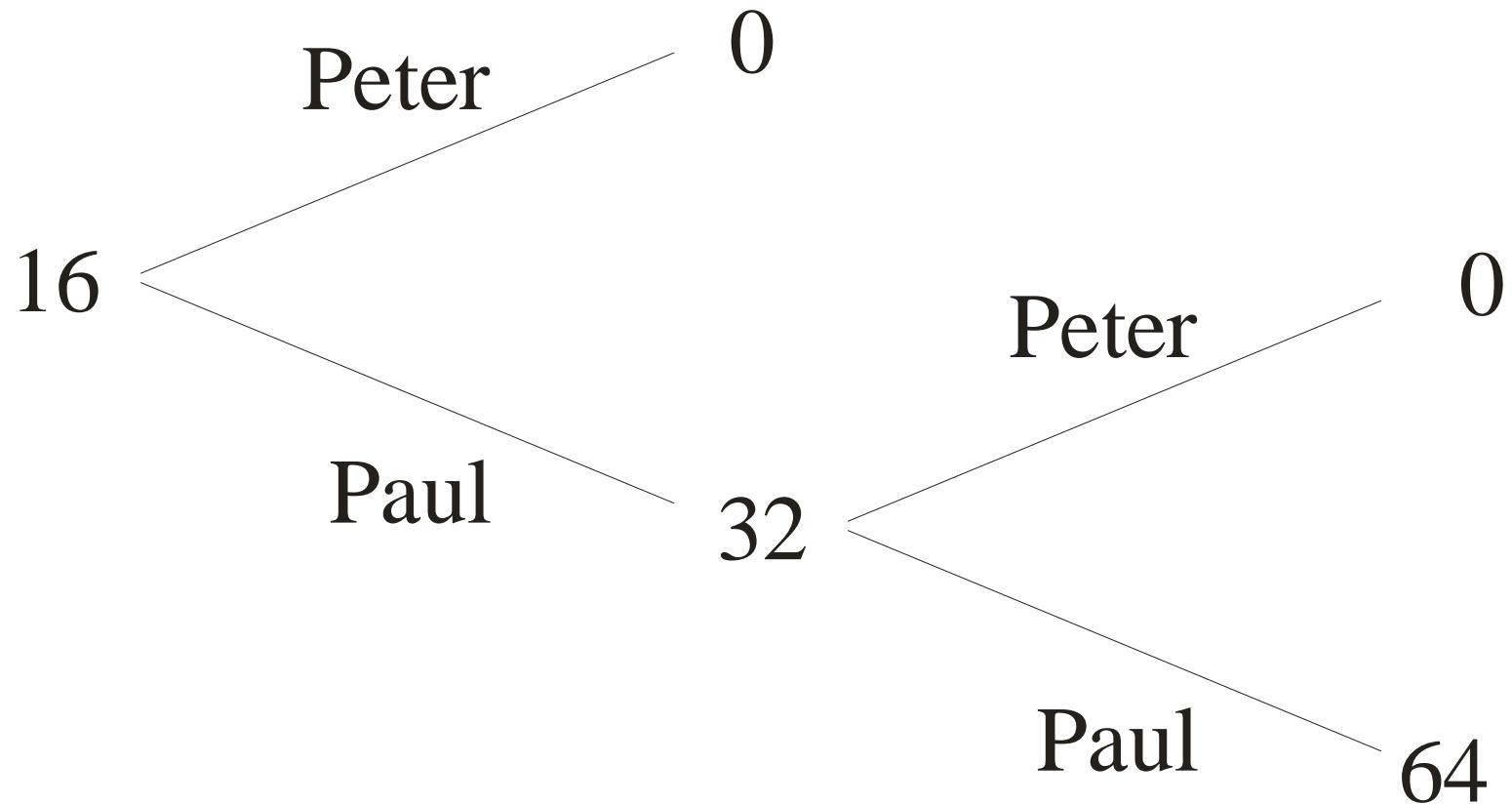
Paul wins only in outcome 4.
So his share should be $\frac{1}{4}$, or 16 pistoles.

Pascal didn't like the argument.



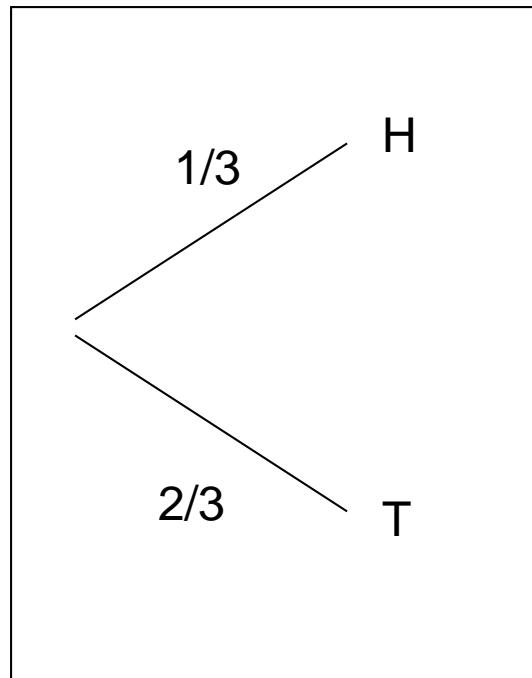
Pierre Fermat, 1601-1665 32

Pascal's answer (game theory)

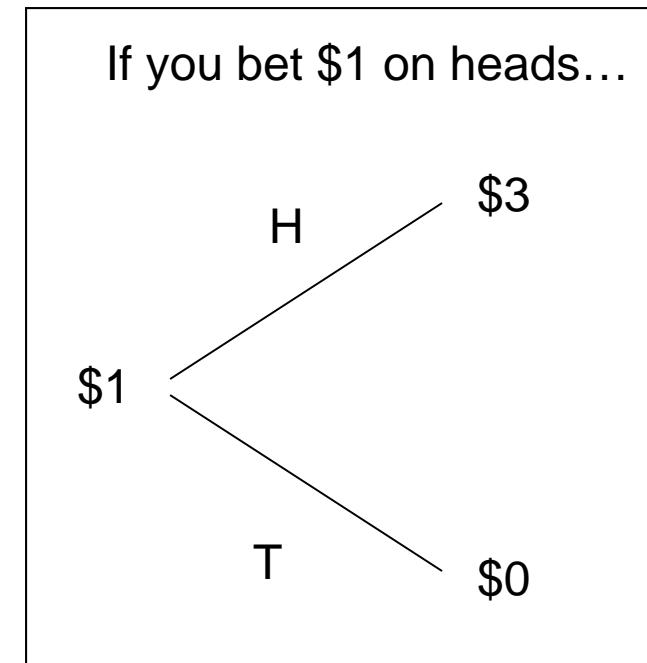


Another probability problem: HH before TT

PROBABILITIES



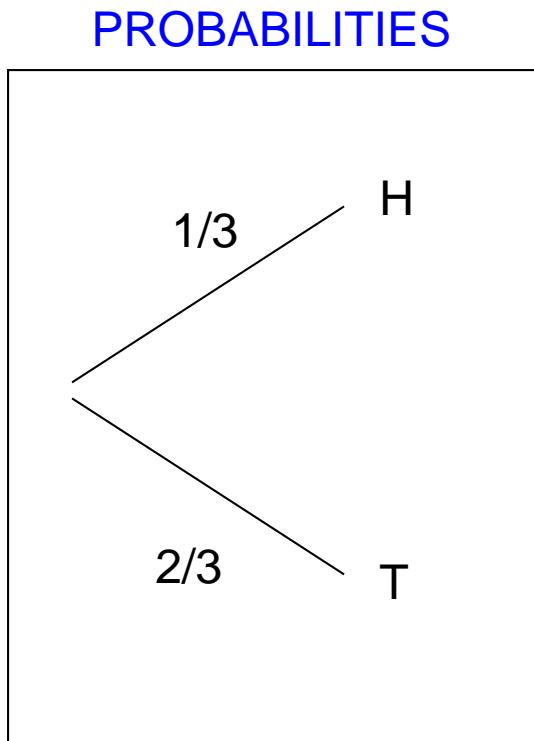
PAYOUTS



Toss the biased coin repeatedly.

What is the probability of
HH before TT?

Measure-theoretic question



What is the probability
of HH before TT?

Measure-theoretic solution

| Probability | Outcome |
|-----------------------------------|---------|
| $(\frac{1}{3})^2$ | HH |
| $(\frac{1}{3})^2 \frac{2}{3}$ | THH |
| $(\frac{1}{3})^3 \frac{2}{3}$ | HTHH |
| $(\frac{1}{3})^3 (\frac{2}{3})^2$ | THTHH |
| $(\frac{1}{3})^4 (\frac{2}{3})^2$ | HTHTHH |
| etc. | |

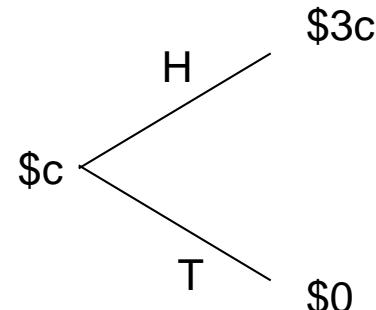
Summing the series, we find

$$\text{probability} = \frac{5}{21}$$

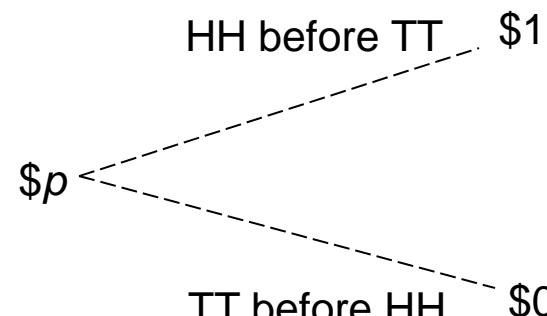
Game-theoretic question

What is p , the price of a \$1 payoff conditional on HH before TT?

Payoff on each toss



\$1 conditional on HH before TT

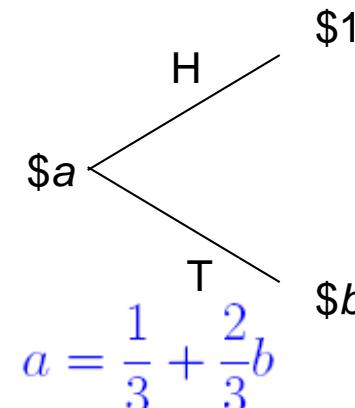


Game-theoretic solution

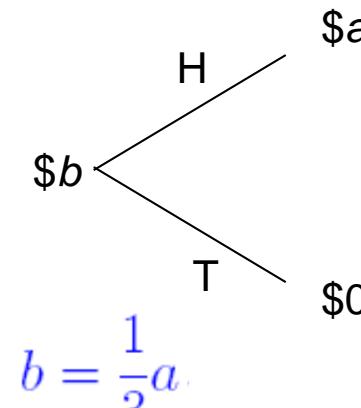
a = price when last toss was H

b = price when last toss was T

Last toss was H



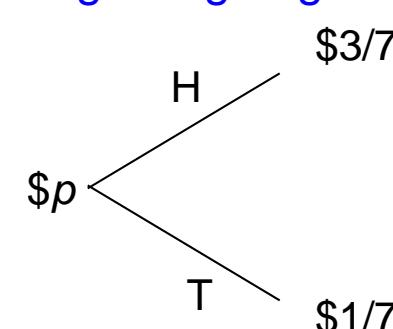
Last toss was T



$$a = \frac{3}{7}$$

$$b = \frac{1}{7}$$

Beginning of game



$$\begin{aligned}
 p &= \frac{1}{3} \times \frac{3}{7} + \frac{2}{3} \times \frac{1}{7} \\
 &= \frac{5}{21}
 \end{aligned}$$

Caveat

In Pascal's problem, the prices are not necessarily assertions about the world. Perhaps the players have different levels of skill but have nevertheless agreed to play at even odds.

This anticipates the modern theory of option pricing, where so-called “risk-neutral” probabilities are merely prices derived from other prices, not assertions about whether what will make money in the future.

Fermat's combinatorial or measure-theoretic approach leads to a metaphysics of possibility. By the 1690s, Jacob Bernoulli was writing about "equally possible cases" that happen equally often.

Pascal's betting approach is more flexible. Betting prices have many meanings. If we choose to use giving betting prices as a theory about the world, the meaning of this theory lies in the way it is tested, not in any metaphysics about possibilities that precede realities.

Game theoretic upper and lower probability

Measure-theoretic probability:

- Classical: elementary events with probabilities adding to one.
- Modern: space with filtration and probability measure.

Probability of A = total of probabilities for elementary events favoring A

Game-theoretic probability:

- One player offers prices for uncertain payoffs.
- Another player decides what to buy.

Probability of A = initial stake needed to obtain the payoff

[1 if A happens and 0 otherwise]

If no strategy delivers exactly the 0/1 payoff:

Upper probability of A = initial stake needed to obtain at least the payoff

[1 if A happens, 0 otherwise]

Objective interpretation of probability

Cournot's principle

Commonly used by mathematicians before WWII

An event of very small probability will not happen.

To avoid lottery paradox, consider only events with simplest descriptions.
(Wald, Schnorr, Kolmogorov, Levin)

Ville's principle

Equivalent to Cournot's principle when upper probabilities are probabilities

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

Mathematical definition of probability:

$P(A) = \text{stake needed to obtain \$1 if } A \text{ happens, \$0 otherwise}$

Objective interpretation of game-theoretic probability:

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

Subjective interpretation of game-theoretic probability:

I don't think you will multiply the capital you risk by a large factor.

Unlike de Finetti, we do not need behavioral assumptions (e.g., people want to bet or can be forced to do so).

To make Pascal's theory part of modern game theory, we must **define the game precisely.**

- Rules of play
- Each player's information
- Rule for winning

A game between Forecaster and Reality

Forecaster gives probabilities for a sequence x_1, x_2, \dots of 1s and 0s.

Before Reality announces x_n , Forecaster announces probability p_n for $x_n = 1$.

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

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

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

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

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

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

Clarifications:

1. The phenomena need not be binary. We assume $x_n \in \{0, 1\}$ only for simplicity.
2. Reality's move space may change from round to round.
3. Perfect information: All players hear announcements as they are made.
4. In addition to x_1, \dots, x_{n-1} , Forecaster may have other newly acquired information.

Forecaster is tested by a third player, Skeptic, who tries to get rich from Forecaster's betting offers.

Players: Forecaster, Reality, Skeptic

Protocol:

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

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

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

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

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

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

Winner: Skeptic wins if $\mathcal{K}_n \geq 0$ for all n and $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$.

Otherwise Forecaster and Reality win.

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 $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (y_i - p_i) = 0$
or $\lim_{n \rightarrow \infty} \mathcal{K}_n = \infty$.

Theorem Skeptic has a winning strategy.

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

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

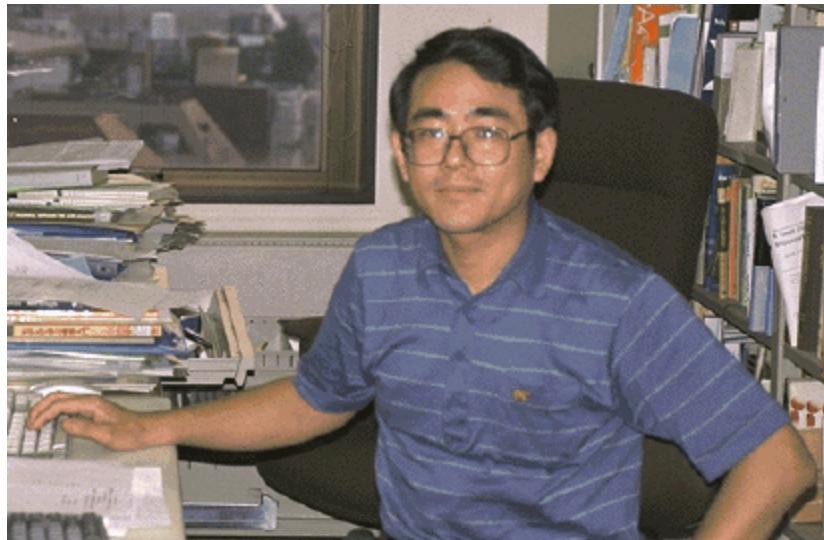
It produces the capital

$$\mathcal{K}_n = 2^n \frac{r_n!(n-r_n)!}{(n+1)!}.$$

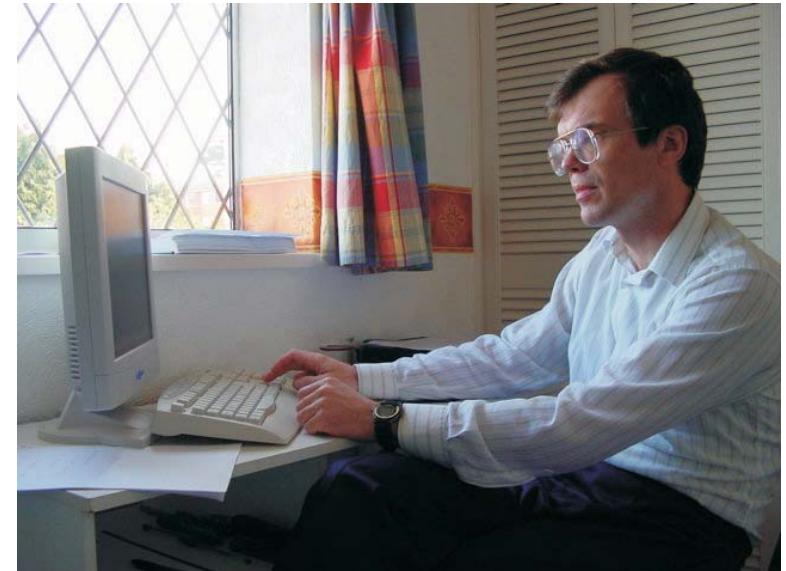
From the assumption that this remains bounded by some constant C , you can easily derive the strong law of large numbers using Stirling's formula.

Defensive forecasting

The name was introduced in Working Paper 8 at www.probabilityandfinance, by Vovk, Takemura, and Shafer (September 2004). See also Working Papers 7, 9, 10, 11, 13, 14, 16, 17, 18, 20, 21, 22, and 30.



Akimichi Takemura



Volodya Vovk

Crucial idea: all the tests (betting strategies for Skeptic) Forecaster needs to pass can be merged into a single **portmanteau test** for Forecaster to pass.

1. If you have two strategies for multiplying capital risked, divide your capital between them.
2. Formally: average the strategies.
3. You can average countably many strategies.
4. As a practical matter, there are only countably many tests (Abraham Wald, 1937).
5. I will explain how Forecaster can beat any single test (including the portmanteau test).

A. How Forecaster beats any single test

B. How to construct a portmanteau test for binary probability forecasting

- Use law of large numbers to test calibration for each probability p .**
- Merge the tests for different p .**

How Forecaster can beat any single test S

Skeptic adopts a continuous strategy \mathcal{S} .

FOR $n = 1, 2, \dots$

 Reality announces $x_n \in X$.

 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 $x_n \in X$.

 Skeptic announces continuous $S_n : [0, 1] \rightarrow \mathbb{R}$.

 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 $x_n \in \mathbf{X}$.

Skeptic announces continuous $S_n : [0, 1] \rightarrow \mathbb{R}$.

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 $S_n(p_n) = 0$.

Why Hilary Putnam thought good probability prediction is impossible...

FOR $n = 1, 2, \dots$

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

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

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

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 Skeptic's move

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

is discontinuous in p . This infinitely abrupt shift—an artificial idealization—is crucial to the counterexample.

Forecaster can defeat any strategy for Skeptic if

- the strategy for Skeptic is continuous in p , or
- Forecaster is allowed to randomize, announcing a probability distribution for p rather than a sharp value for p .

See Working Papers 7 & 8 at www.probabilityandfinance.com.

Probability judgment

To conclude, I will discuss how Bayesian conditioning can be understood as an example of probability judgment.

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

Assumptions

1. $P(A)$ = price of a ticket that pays 1 if A happens.
2. $P(A)x$ = price of a ticket that pays x if A happens.
(Here x can be any real number.)
3. After A happens (we learn A and nothing else),
 $P(B|A)x$ = price of a ticket that pays x if B happens.

Argument

1. Pay $P(A)P(B|A)$ to get $P(B|A)$ if A happens. If A does happen, pay $P(B|A)$ to get 1 if B happens.
2. So $P(A)P(B|A)$ is the cost of getting 1 if $A \& B$ happens.

De Finetti's adopted De Moivre's argument for $P(A \& B) = P(A)P(B|A)$, changing "price" to "an individual's price".

Assumptions

1. $P(A)x$ = price at which I will sell a ticket that pays x if A happens.
2. After A happens (we learn A and nothing else),
 $P(B|A)x$ = price at which I will sell a ticket that pays x if B happens.

Argument

1. You pay me $P(A)P(B|A)$ to get $P(B|A)$ if A happens. If A does happen, you pay me $P(B|A)$ to get 1 if B also happens.
2. So $P(A)P(B|A)$ is what you need to pay me to get 1 if $A \& B$ happens.

De Finetti interpreted De Moivre's prices in a particular way.

There are other ways.

In game-theoretic probability (Shafer and Vovk 2001) we interpret the prices as a prediction.

The prediction: You will not multiply by a large factor the capital you risk at these prices.

The game-theoretic argument for $P(B|A) = \frac{P(A \& B)}{P(A)}$

Context Winning against given prices means multiplying your capital by a large factor buying or selling the tickets priced (and others like them in the long run).

Hypothesis You will not win against $P(A)$ and $P(A \& B)$.

Conclusion You still will not win if after A (and nothing else) is known, $P(A \& B)/P(A)$ is added as a new probability for B .

How to prove it Show that if \mathcal{S} is a strategy against all three probabilities, then there exists a strategy \mathcal{S}' against $P(A)$ and $P(A \& B)$ alone that risks the same risks and payoffs.

Proof: Let M be the amount of B tickets S buys after learning A . To construct \mathcal{S}' from \mathcal{S} , delete these B tickets and add

$$M \text{ tickets on } A \& B \quad \text{and} \quad -M \frac{P(A \& B)}{P(A)} \text{ tickets on } A$$

to S 's purchases in the initial situation.

- The tickets added have zero total initial cost:

$$MP(A \& B) - M \frac{P(A \& B)}{P(A)} P(A) = 0.$$

- The tickets added and the tickets deleted have the same net payoffs:

| | |
|---|------------------------------|
| 0 | if A does not happen; |
| $-M \frac{P(A \& B)}{P(A)}$ | if A happens but not B ; |
| $M \left(1 - \frac{P(A \& B)}{P(A)}\right)$ | if A and B both happen. |

Comments

1. **Game-theoretic advantage over de Finetti:** the condition that we learn only A and nothing else (relevant) has a meaning without a prior protocol (see my 1985 article on conditional probability).
2. **Winning against probabilities** by multiplying the capital risked over the long run: To understand this fully, learn about game-theoretic probability.

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 X with set of possible values \mathcal{X} .
- We learn a mapping $\Gamma : \mathcal{X} \rightarrow 2^\Omega$ with this meaning:
If $X = x$, 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 X and ω does not affect our inability to beat the probabilities for 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.

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

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

- Joe says “Glenn paid.”

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

- New beliefs:

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

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%.

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

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

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

- Bill says “Glenn paid \$ 5.”

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

- New beliefs:

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

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 X with set of possible values \mathcal{X} .
- We learn a mapping $\Gamma : \mathcal{X} \rightarrow 2^\Omega$ with this meaning:

If $X = x$, then $\omega \in \Gamma(x)$.

- $\Gamma(x) = \emptyset$ for some $x \in \mathcal{X}$.
- For $A \subseteq \Omega$, our belief that $\omega \in A$ is now

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

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 X .

Example: The witness caught out

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

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

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

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

- Tom says “Glenn paid \$ 10.”

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

- New beliefs:

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

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

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

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

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

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

Product measure:

$$\mathcal{X}_{\text{Bill} \& \text{ Tom}} =$$

$$\mathbb{P}(\text{Bill precise, Tom precise}) =$$

$$\mathbb{P}(\text{Bill precise, Tom approximate}) =$$

etc.

$$\mathcal{X}_{\text{Bill}} \times \mathcal{X}_{\text{Tom}}$$

$$0.7 \times 0.7 = 0.49$$

$$0.7 \times 0.2 = 0.14$$

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."

$$\Gamma_1(\text{Joe reliable}) = \{\text{paid}\} \quad \Gamma_1(\text{Joe not reliable}) = \{\text{paid}, \text{not paid}\}$$

$$\Gamma_2(\text{Bill reliable}) = \{\text{not paid}\} \quad \Gamma_2(\text{Bill not reliable}) = \{\text{paid}, \text{not paid}\}$$

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

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

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)

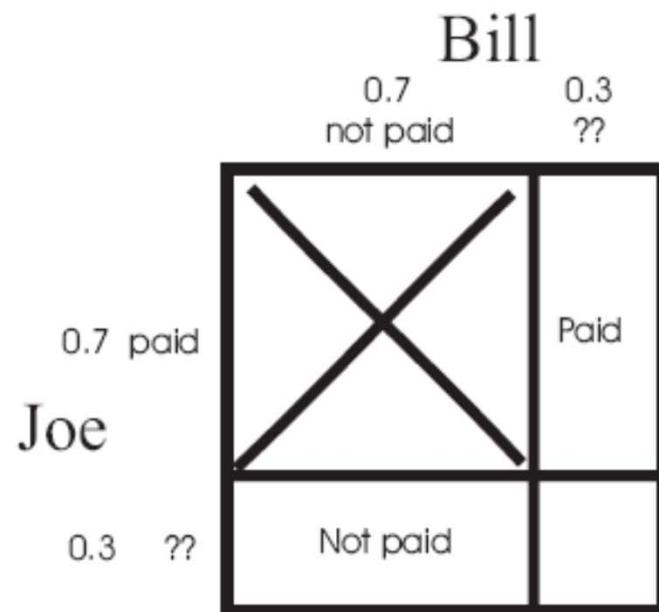
- Variable ω with set of possible values Ω .
- Random variables X_1 and X_2 with sets of possible values \mathcal{X}_1 and \mathcal{X}_2 .
- Form the product measure on $\mathcal{X}_1 \times \mathcal{X}_2$.
- We learn mappings $\Gamma_1 : \mathcal{X}_1 \rightarrow 2^\Omega$ and $\Gamma_2 : \mathcal{X}_2 \rightarrow 2^\Omega$:
If $X_1 = x_1$, then $\omega \in \Gamma_1(x_1)$. If $X_2 = x_2$, then $\omega \in \Gamma_2(x_2)$.
- So if $(X_1, X_2) = (x_1, x_2)$, 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 X_1 and X_2 .

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

Joe: $B_1\{\text{paid}\} = 0.7 \quad B_1\{\text{not paid}\} = 0$
Bill: $B_2\{\text{not paid}\} = 0.7 \quad B_2\{\text{paid}\} = 0$



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

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

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