Tutorial

# Game-theoretic probability
# and
# the history of the philosophy of probability

Glenn Shafer

Tokyo University

March 17, 2006

The game-theoretic framework for probability is simple. But to appreciate it, you need mathematics and philosophy:

**Mathematics** Classical probability theorems can be made game-theoretic.

**Philosophy** The classical French interpretation of probability can be made game-theoretic.

## My talk: Game-Theoretic Probability

Game-theoretic probability competes with measure-theoretic probability.

It is an alternative formalization of classical probability.

**Mathematics** Classical probability theorems become theorems in game theory (someone has a winning strategy).

**Philosophy** Cournot's principle, the classical French interpretation of probability (events of small probability do not happen) becomes game-theoretic (you do not get rich without risking bankruptcy).

Professor Vovk's talk: Defensive Forecasting

Game-theoretic probability produces something radically new, which exists neither in classical nor in measure-theoretic probability.

Even if reality plays against you,

- You can give valid sequential probabilities.

- You can use them to make optimal decisions.

## Part I. The Mathematics of Game-Theoretic Probability

1. Basic framework: Pascal & Ville
2. The strong law of large numbers
3. The weak law of large numbers
4. The central limit theorem
5. The $\sqrt{dt}$ effect

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot
2. Making it the meaning of probability: Lévy & Kolmogorov
3. English indifference & German skepticism
4. Liquidating Cournot's principle: Hitler, Stalin & Doob
5. Making Cournot's principle game-theoretic: Ville

## Part I. The Mathematics of Game-Theoretic Probability

1. Basic framework: Pascal & Ville. Pascal assumed no arbitrage (you cannot make money for sure) in a sequential game. Ville added Cournot's principle (you will not get rich without risking bankruptcy).

2. The strong law of large numbers

3. The weak law of large numbers

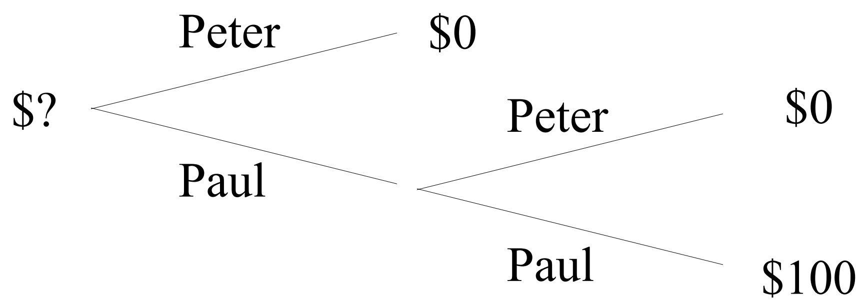
4. The central limit theorem

5. The $\sqrt{dt}$ effect

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

## Pascal: Fair division

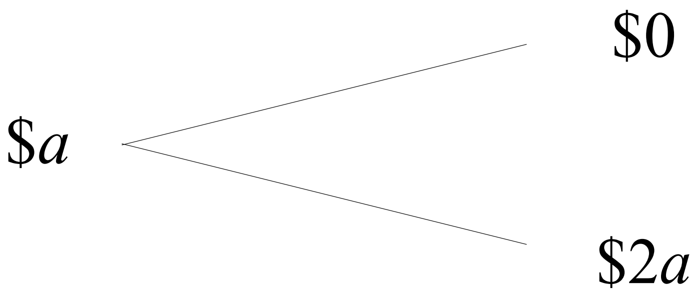
Peter and Paul play for \$100. Paul is behind. Paul needs 2 points to win, and Peter needs only 1.



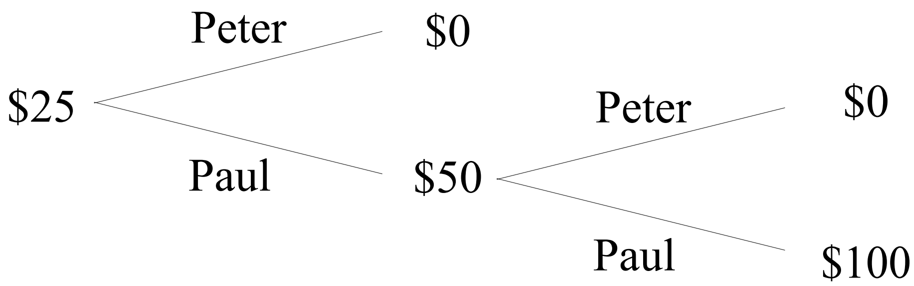

If the game must be broken off, how much of the \$100 should Paul get?

It is fair for players to put up equal stakes if the winner takes all.

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




So Paul should get $25.




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

Binary probability game.

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

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

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

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

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

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

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

No Arbitrage: If Forecaster announces a strategy in advance, the strategy needs to obey the rules of probability in order to keep Skeptic from making money for sure.

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

Synonyms for no arbitrage: coherence, no Dutch book

## The lesson from Pascal

Probability is about fair prices in a sequential game.

Pascal's concept of fairness: no arbitrage.

## The lesson from Ville

Jean Ville developed a second concept of fairness: you will not get rich without risking bankruptcy.

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

In 1939, Ville showed that the laws of probability can be derived from a principle of market efficiency:

> 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.

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

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

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

## Summary

Two aspects of fairness in game-theoretic probability.

**Pascal** No arbitrage: You canot make money for sure.

**Ville** Cournot's principle: You will not get rich without risking bankruptcy.

## Part I. The Mathematics of Game-Theoretic Probability

1. Basic framework: Pascal & Ville

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

3. The weak law of large numbers

4. The central limit theorem

5. The $\sqrt{dt}$ effect

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

$\mathcal{K}_0 = 1$.

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

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

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

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

Skeptic wins if

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

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

Otherwise Reality wins.

Theorem Skeptic has a winning strategy.

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

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

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

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

## The Idea of the Proof

Idea 1 Establish an account for betting on heads. On each round, bet $\epsilon$ 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 $k$th 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 $y_1y_2\ldots$ of moves for Reality.

- An *event* is a set of paths.

- A *situation* is a finite initial sequence of moves for Reality, say $y_1y_2\ldots y_n$.

- $\Box$ is the *initial situation*, a sequence of length zero.

- When $\xi$ is a path, say $\xi = y_1y_2\ldots$, write $\xi^n$ for the situation $y_1y_2\ldots y_n$.

## 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 $\mathcal{P}(y_1 \dots y_{n-1})$ of $y_n$ Skeptic in situation $y_1 \dots y_{n-1}$.

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

- 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 $\alpha$.

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 $\alpha$, 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 $\alpha_1$ and one with initial capital $\alpha_2$. 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).

# Forcing

A strategy $\mathcal{P}$ for Skeptic *forces* an event $E$ if

$$\mathcal{K}^{\mathcal{P}}(t) \geq -1$$

for every situation $t$ and

$$\lim_{n\to\infty} \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty$$

for every path $\xi$ not in $E$.

This means $\mathcal{P}$ is a winning strategy in the game where Skeptic starts with capital 1 and has $E$ as his goal instead of $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i = 0$.

## Weak Forcing

- $\mathcal{P}$ *forces* $E$ if $\mathcal{K}^{\mathcal{P}}(t) \geq -1$ for every situation $t$ and

$$\lim_{n \to \infty} \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty$$

  for every path $\xi$ not in $E$.

- $\mathcal{P}$ *weakly forces* $E$ if $\mathcal{K}^{\mathcal{P}}(t) \geq -1$ for every situation $t$ and

$$\sup_n \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty$$

  for every path $\xi$ not in $E$.

## Equivalence of Forcing and Weak Forcing

The following lemma shows that forcing and weak forcing are practically equivalent.

**Lemma 1** *If Skeptic can weakly force $E$, then he can force $E$.*

*Proof* Suppose $\mathcal{P}$ weakly forces $E$. Choose a number larger than 1, say 2. Starting with initial capital 1, Skeptic play $\mathcal{P}$ until the capital exceeds 2.

Set aside 1 and continue with a rescaled version of $\mathcal{T}$, scaled down to the reduced capital. (Multiply $\mathcal{P}$'s moves on succeeding rounds by the factor by which the capital has been reduced, thus assuring that the capital on succeeding rounds is also multiplied by this factor.)

When the capital again exceeds 2, again set aside 1, and so forth. The money set aside grows without bound.

## Combining Strategies that Force Events

**Lemma 2** *If Skeptic can weakly force each of a sequence $E_1, E_2, \ldots$ of events, then he can weakly force $\bigcap_{k=1}^{\infty} E_k$.*

*Proof* Suppose $\mathcal{P}_k$ weakly forces $E_k$. Then

$$|\mathcal{P}_k(y_1 \ldots y_n)| \leq 1 + \mathcal{K}^{\mathcal{P}_k}(y_1 \ldots y_n) \leq 2^n.$$

Skeptic bets $\mathcal{P}_k(y_1 \ldots y_n)$ on round $n+1$. When he makes this bet, he has capital $1 + \mathcal{K}^{\mathcal{P}_k}(y_1 \ldots y_n)$. The first inequality holds because he cannot bet more than he has (he must avoid risking bankruptcy). The second inequality holds because he cannot, consequently, do more than double his money on each round.

Because for each $y_1 \ldots y_n$ there is a constant $C$ (namely $2^n$) such that $\mathcal{P}_k(y_1 \ldots y_n) \leq C$ for all $k$, a strategy $\mathcal{Q}$ can be defined by

$$\mathcal{Q} := \sum_{k=1}^{\infty} 2^{-k} \mathcal{P}_k.$$

Since $\mathcal{P}_k$ weakly forces $E_k$, $\mathcal{Q}$ also weakly forces $E_k$. So $\mathcal{Q}$ weakly forces $\bigcap_{k=1}^{\infty} E_k$. ∎

## Bounding Reality's Average Move from Above

**Lemma 3** *Suppose $\epsilon > 0$. Then Skeptic can weakly force*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i \leq \epsilon.$$

Proof: Suppose $\epsilon < \frac{1}{2}$. Let $\mathcal{P}$ be the strategy that always buys $\epsilon\alpha$ of $y$, where $\alpha$ is the current capital. Because $y$ is never less than $-1$, this strategy loses at most the fraction $\epsilon$ of the current capital, and hence the capital process is nonnegative. It is given by $1 + \mathcal{K}^{\mathcal{P}}(\Box) = 1$ and

$$1 + \mathcal{K}^{\mathcal{P}}(y_1 \dots y_n) = (1 + \mathcal{K}^{\mathcal{P}}(y_1 \dots y_{n-1}))(1 + \epsilon y_n) = \prod_{i=1}^{n} (1 + \epsilon y_i)\,.$$

Let $\xi = y_1 y_2 \dots$ be a path such that $\sup_n \mathcal{K}^{\mathcal{P}}(y_1 \dots y_n) < \infty$. Then there exists a constant $C_\xi > 0$ such that

$$\prod_{i=1}^{n} (1 + \epsilon y_i) \leq C_\xi \text{ for all } n.$$

Taking logarithms of both sides of $\prod_{i=1}^{n}(1+\epsilon y_i) \le C_\xi$, we find that

$$\sum_{i=1}^{n} \ln(1+\epsilon y_i) \le D_\xi$$

for all $n$ for some $D_\xi$. Since $\ln(1+t) \ge t - t^2$ whenever $t \ge -\frac{1}{2}$, $\xi$ also satisfies

$$\begin{aligned}
\epsilon \sum_{i=1}^{n} y_i - \epsilon^2 \sum_{i=1}^{n} y_i^2 &\le D_\xi, \\
\epsilon \sum_{i=1}^{n} y_i - \epsilon^2 n &\le D_\xi, \\
\epsilon \sum_{i=1}^{n} y_i &\le D_\xi + \epsilon^2 n,
\end{aligned}$$

or

$$\frac{1}{n}\sum_{i=1}^{n} y_i \le \frac{D_\xi}{\epsilon n} + \epsilon$$

for all $n$ and hence satisfies $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y_i \le \epsilon$. Thus $\mathcal{P}$ weakly forces this event.

## Bouding Reality's Average Move from Below

The same argument, with $-\epsilon$ in place of $\epsilon$, establishes the following complementary lemma.

**Lemma 4** *Suppose* $\epsilon > 0$. *Then Skeptic can weakly force*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i \geq -\epsilon.$$

## Proof of the Game-Theoretic Strong Law

According to Lemma 1, it suffices to show he can weakly force

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i = 0.$$

Consider the events

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i \leq \epsilon \quad \text{and} \quad \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i \geq -\epsilon$$

for $\epsilon = 2^{-k}$, where $k$ ranges over all natural numbers. By Lemmas 3 and 4, Skeptic can weakly force each of these events. By Lemma 2, he can therefore weakly force their intersection.

## 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. The Mathematics of Game-Theoretic Probability

1. Basic framework: 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.

4. The central limit theorem

5. The $\sqrt{dt}$ effect

## 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 I. The Mathematics of Game-Theoretic Probability

1. Basic framework: Pascal & Ville

2. The strong law of large numbers

3. The weak law of large numbers

4. The central limit theorem. This is about the price of a payoff that depends on the outcomes of a large number of trials.

5. The $\sqrt{dt}$ effect

## Pricing variables

| |
|---|
| $\mathcal{K}_0 := \alpha$.<br>FOR $n = 1, \dots, N$:<br>  Skeptic announces $M_n \in \mathbb{R}$.<br>  Reality announces $y_n \in \{-1, 1\}$.<br>  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$. |

- A sequence $y_1, \dots, y_N$ of moves by Reality is a *path*.

- The set of all paths is the *sample space*.

- A function on the sample space is a *variable*.

## Upper Price for a Variable $z$:

$$\overline{\mathbb{E}}\, z := \text{ smallest initial stake Skeptic can parlay into } z \text{ or more at the end of the game}$$

$$= \inf\{\mathcal{K}(\square) | \mathcal{K} \text{ is a martingale and } \mathcal{K}(y_1, \ldots, y_N) \geq z(y_1, \ldots, y_N)\}.$$

This is Skeptic's *minimum selling price* for $z$. He can replicate $z$ at this price with no risk of loss.

## Lower Price for a Variable $z$:

$$\underline{\mathbb{E}}\, z := -\overline{\mathbb{E}}[-z].$$

Buying $z$ for $\alpha$ is the same as selling $-z$ for $-\alpha$. So $\underline{\mathbb{E}}\, z$ is Skeptic's *maximum buying price* for $z$.

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

Given a process $\mathcal{P}$, we write $\Delta\mathcal{P}_n$ for its $n$th increment:

$$\Delta\mathcal{P}_n := \mathcal{P}(y_1, \ldots, y_n) - \mathcal{P}(y_1, \ldots, y_{n-1}).$$

In the case of a martingale $\mathcal{K}$,

$$\Delta\mathcal{K}_n = M_n y_n,$$

where $M_n$ is the move specified by the strategy. Notice that $M_n$ is a function of $y_1, \ldots, y_{n-1}$.

# The central limit theorem

We consider only coin-tossing (DeMoivre's theorem). For simplicity, we now score Heads as $1/\sqrt{N}$ and Tails as $-1/\sqrt{N}$.

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

Set $\mathcal{S}_n := \sum_{i=1}^{n} y_i$.

Consider a smooth function $U$.

De Moivre's Theorem For $N$ sufficiently large, both $\overline{\mathbb{E}}\, U(\mathcal{S}_N)$ and $\underline{\mathbb{E}}\, U(\mathcal{S}_N)$ are arbitrarily close to $\int_{-\infty}^{\infty} U(z)\, \mathcal{N}_{0,1}(dz)$.

## How do we prove De Moivre's theorem?

$$\mathcal{S}_n := \sum_{i=1}^{n} y_i.$$

We want to know the price at time 0 of the payoff $U(\mathcal{S}_N)$ at time $N$. Let us also consider its price at time $n$. Intuitively, this should depend on $\mathcal{S}_n$, the value of the sum so far. Assume, optimistically, that the price at time $n$ is given by a function of two variables, $\overline{U}(s, D)$: the price at time $n$ is $\overline{U}(\mathcal{S}_n, \frac{N-n}{N})$.

Successive prices are

$$\overline{U}(0,1), \overline{U}(\mathcal{S}_1, \tfrac{N-1}{N}), \ldots$$
$$\ldots, \overline{U}(\mathcal{S}_{N-1}, \tfrac{1}{N}), \overline{U}(\mathcal{S}_N, 0),$$

These must be the successive values of a martingale.

- $\overline{U}(\mathcal{S}_N, 0)$ must equal $U(\mathcal{S}_N)$.
- $\overline{U}(0, 1)$ is the price that interests us.

A martingale is a process $\mathcal{K}_n$ with increments of the form $\Delta\mathcal{K}_n = M_n y_n$.

Our task: given $U$, choose $\overline{U}(s, D)$ so that

(1) $\overline{U}(\mathcal{S}_n, \frac{N-n}{N})$ is a martingale, and

(2) $\overline{U}(\mathcal{S}_N, 0) = U(\mathcal{S}_N)$.

Consider the increments in $s$, $D$, and $\overline{U}$:

- $\Delta s_n = y_n = \pm\frac{1}{\sqrt{N}}$.
- $\Delta D_n = -\frac{1}{N}$.
- $\Delta\overline{U}_n = \overline{U}(\mathcal{S}_n, \frac{N-n}{N}) - \overline{U}(\mathcal{S}_{n-1}, \frac{N-n+1}{N})$.

Use a Taylor expansion:

$$\Delta\overline{U} \approx \frac{\partial\overline{U}}{\partial s}\Delta s + \frac{\partial\overline{U}}{\partial D}\Delta D + \frac{1}{2}\frac{\partial^2\overline{U}}{\partial s^2}(\Delta s)^2 = \frac{\partial\overline{U}}{\partial s}y - (\frac{\partial\overline{U}}{\partial D} - \frac{1}{2}\frac{\partial^2\overline{U}}{\partial s^2})\frac{1}{N}.$$

$$\boxed{\begin{array}{l} \Delta\mathcal{K}_n = M_n y_n \\ \Delta\overline{U}_n \approx \frac{\partial \overline{U}}{\partial s} y_n - (\frac{\partial \overline{U}}{\partial D} - \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2})\frac{1}{N} \end{array}}$$

We need the second term to go away, which requires

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}$$

Then we obtain the desired martingale by buying $\frac{\partial \overline{U}}{\partial s}$ $y$-tickets on the $n$th round. In other words, we set

$$M_n := \frac{\partial \overline{U}}{\partial s}(\mathcal{S}_{n-1}, \tfrac{N-n+1}{N}).$$

Laplace showed that the solution of the *heat equation*

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}$$

with the initial condition $\overline{U}(s, 0) = U(s)$ is

$$\overline{U}(s, D) = \int_{-\infty}^{\infty} U(z)\, \mathcal{N}_{s,D}(dz).$$

So the initial price $\overline{U}(0, 1)$ is

$$\int_{-\infty}^{\infty} U(z)\, \mathcal{N}_{0,1}(dz).$$

## Part I. The Mathematics of Game-Theoretic Probability

1. Basic framework: Pascal & Ville

2. The strong law of large numbers

3. The weak law of large numbers

4. The central limit theorem.

5. The $\sqrt{dt}$ effect. How to make money if price changes to do scale with the square root of the time interval.

## The $\sqrt{dt}$ effect

Changes in market prices over an interval of time of length $dt$ scale as $\sqrt{dt}$.

When shares are traded 252 days a year, for example, the typical annual price change is $\sqrt{252} = 16$, times as large as the typical daily change.

# Why?

## Stochastic explanation

- Assume price changes are stochastic.
- Successive changes must be uncorrelated; otherwise you could devise a trading strategy with positive expected value.
- Uncorrelatedness of 252 daily changes implies that their sum has standard deviation $\sqrt{252}$ times as large.

## Purely game-theoretic explanation

$\mathcal{K}_0 := 1$.
Market announces $y_0 \in \mathbb{R}$.
FOR $n = 1, 2, \dots, N$:
  Investor announces $s_n \in \mathbb{R}$.
  Market announces $y_n \in \mathbb{R}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n - y_{n-1})$.

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

As it turns out, Investor can make a lot of money in this game unless Market obeys the $\sqrt{dt}$ effect.

Compare the typical daily change

$$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - y_{n-1})^2} \tag{1}$$

to the change over the whole game,

$$\max_{0<n\leq N} |y_n - y_0|. \tag{2}$$

The $\sqrt{dt}$ effect says (2) should have order of magnitude $\sqrt{N}$ times that of (1):

$$\sum_{n=1}^{N}(y_n - y_{n-1})^2 \sim \max_{0<n\leq N}(y_n - y_0)^2.$$

Average a momentum strategy (hold $Cy_{n-1}$ shares) and a contrarian strategy (hold $-Cy_{n-1}$ shares). The resulting strategy makes money if

$$\sum_{n=1}^{N}(y_n - y_{n-1})^2 \sim \max_{0<n\leq N}(y_n - y_0)^2$$

is violated.

1. If Investor can count on $\sum(y_n - y_{n-1})^2 \leq E$ and $\max(y_n - y_0)^2 \geq D$, he can choose $C$ so that the momentum strategy turns \$1 into \$$D/E$ or more for sure.

2. If Investor can count on $\sum(y_n - y_{n-1})^2 \geq E$ and $\max(y_n - y_0)^2 \leq D$, he can choose $C$ so that the contrarian strategy turns \$1 into \$$E/D$ or more for sure.

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot

2. Making it the meaning of probability: Lévy & Kolmogorov

3. English indifference & German skepticism

4. Liquidating Cournot's principle: Hitler, Stalin & Doob

5. Making Cournot's principle game-theoretic: Ville

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot. Pascal was concerned only with fairness. Bernoulli wanted to connect probabilities with what happens in the world.

2. Making it the meaning of probability: Lévy & Kolmogorov

3. English indifference & German skepticism

4. Liquidating Cournot's principle: Hitler, Stalin & Doob

5. Making Cournot's principle game-theoretic: Ville

Bernoulli introduced mathematical probability is his celebrated *Ars Conjectandi* (1713).

Jakob Bernoulli
1654–1705

"Something is *morally certain* if its probability is so close to certainty that the shortfall is imperceptible."

"Something is *morally impossible* if its probability is no more than the amount by which moral certainty falls short of complete certainty."

How Bernoulli connected probability with the world:

"*Because it is only rarely possible to obtain full certainty, necessity and custom demand that what is merely morally certain be taken as certain.* It would therefore be useful if fixed limits were set for moral certainty by the authority of the magistracy—if it were determined, that is to say, whether 99/100 certainty is sufficient or 999/1000 is required..."

In other words, an event with very small probability will not happen.

Antoine Cournot
1801–1877

Maurice Fréchet, 1878–1973, proposed the name *Cournot's principle.*

Cournot discussed both *moral impossibility* (very small probability) and *physical impossibility* (infinitely small probability).

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.

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot

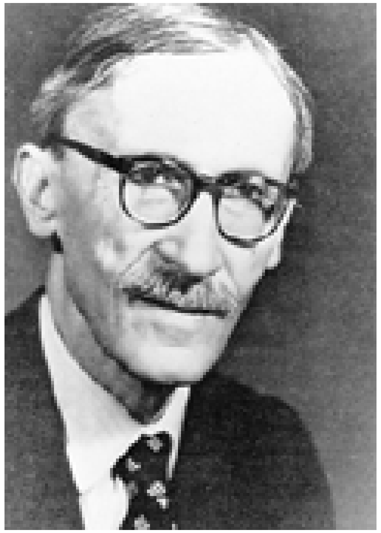
2. Making it the meaning of probability: Lévy & Kolmogorov.
   Lévy articulated the idea best.

3. English indifference & German skepticism

4. Liquidating Cournot's principle: Hitler, Stalin & Doob

5. Making Cournot's principle game-theoretic: Ville

Paul Lévy was emphatic: *Cournot's principle is the only connection between probability and the empirical world.* This became a consensus among the classical French probabilists.



Paul Lévy
1886–1971

In 1925, Lévy explained that probability is based on two principles:

- The principle of equally likely events, which is the foundation for mathematics.
- The principle of the very unlikely event, which is the basis of applications.

Andrei Kolmogorov
1903–1987
The Soviet Euler

In his *Grundbegriffe* (1933), Kolmogorov gave two principles for connecting probability with the empirical world:

**Principle A:** Over many trials, the frequency with which $E$ happens will approximate P($E$).

**Principle B:** On a single trial, if P($E$) very small, we can be practically certain $E$ will not happen.

According to the weak law of large numbers, B implies A.

Richard von Mises
1883-1953

Von Mises invented the theory of collectives

Kolmogorov acknowledged the influence of Richard von Mises's frequentism. Von Mises said a sequence of trials is random if we not know how to select a subsequence of trials that will be different.

Kolmogorov connected this with Cournot's principle. If an event does not usually happen (because it has small probability), and there is nothing that marks next trial as different, then we can assume the event will not happen on the next trial.

Émile Borel
1871–1956

Inventor of measure theory

Minister of the French navy in 1925

By 1910, Borel was already the uncontested leader of classical French probability. But only in the 1940s was he as clear as Lévy about Cournot's principle being the only link between probability and the world.

Borel's way of saying it: 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: $p < 10^{-50}$.

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot

2. Making it the meaning of probability: Lévy & Kolmogorov

3. English indifference & German skepticism. The Germans were too Kantian; the English were too practical.

4. Liquidating Cournot's principle: Hitler, Stalin & Doob.

5. Making Cournot's principle game-theoretic: Ville

## The Germans had no use for Cournot's principle.

Hans Reichenbach
1891–1953

Whereas the French and the Russian mathematicians did their own philosophy in the late 18th and early 19th centuries, the Germans had already established a modern division of labor.

For the German philosophers, the guide was Emmanuel Kant, and the probabilistic truths about the world were synthetic. In this optic, Cournot's principle made no sense.

The English had no use for Cournot's principle.

Ronald Fisher
1890–1962

The British statisticians saw little substance in French theorizing. Why start with something purely notional (equally likely cases) and then try to relate it to the world? Start with what you see in the world!

For Fisher, a probability was a relative frequency in a large (infinitely large!) population. End of story.

Not all the British were frequentists. Some (Jevons, Edgeworth, Jeffreys) were subjectivists.

What they had in common was their idea that probability should model something in the world (frequency or belief) directly.

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot

2. Making it the meaning of probability: Lévy & Kolmogorov

3. English indifference & German skepticism

4. Liquidating Cournot's principle: Hitler Stalin & Doob. Why did it disappear?

5. Making Cournot's principle game-theoretic: Ville

## Why did Cournot's priniciple disappear in the second half of the 20th century?

Two monsters:

- Adolf Hitler (1889–1945)

- Joseph Stalin (1879–1953)

A great mathematician:

- Joseph Doob (born 1910)

## The great destroyer

Adolf Hitler
1889–1945

World War II and the Holocaust destroyed the primacy of the Parisian school of probability.

Driven out of Vienna by Hitler, Reichenbach and Carnap settled in the USA. There they set the framework for postwar philosophy of probability, without Cournot's principle.

## The great silencer


Joseph Stalin
1879–1953

Connectomg probability theory with the real world (statistics) was dangerous under Stalin.

So Kolmogorov stated his philosophy seldom and tersely. Western readers often concluded that he had no philosophy.

Probability is measure, and there is nothing more to say.

## The champion of measure theory

Joseph Doob, 1910–2004, receiving the National Medal of Science from President Carter in 1979.

Picking up where Kolmogorov left off, Doob showed how continuous random processes (e.g., Brownian motion) can be put in the measure-theoretic framework.

He borrowed the idea of a martingale from Ville and made it measure-theoretic.

## Martingales

One of Doob's great contributions was to fit martingales into the measure-theoretic framework.

Intuitively, a martingale is the path followed by a gambler's wealth as he makes successive bets.

Doob showed that if the gambler does not risk bankruptcy, then the martingale becomes infinitely large with probability zero.

Intuitively,

$$\boxed{\text{event of probability zero will not happen}} \Leftrightarrow \boxed{\text{martingale will not become infinitely large}}$$

## Doob's problem

The philosophical foundation for probability espoused by the French and by Kolmogorov (Bernoulli's theorem + Cournot's principle) breaks down for stochastic processes.

Bernoulli's theorem does not apply because we are not repeating the same random experiment over and over.

Doob, the practical American, resolved the problem by belittling philosophy altogether.

## Neyman's solution

After Doob, those who preferred an objective interpretation of probability were less enamored with "probability=frequency".

Often they instead located the meaning of probabilities in their role in generating outcomes.

As Jerzy Neyman explained in a famous article in 1960,

- Laws are needed to produce deterministic phenomena.
- Probabilities are needed to produce indeterministic phenomena.

Indeterministic phenomena exist. Therefore objective probabilities exist.

## Part II. The History of Cournot's Principle

1. Inventing Cournot's principle: Bernoulli & Cournot

2. Making it the meaning of probability: Lévy & Kolmogorov

3. English indifference & German skepticism

4. Liquidating Cournot's principle: Hitler, Stalin & Doob.

5. Making Cournot's principle game-theoretic: Ville. Although Ville's work was neglected, it is the basis of game-theoretic probability.

## Ville's Theorem

Consider binary $Y_1, Y_2, \dots$ with joint probability distribution P.

### Binary Probability Protocol

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

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

Ville showed that Skeptic's getting rich in this protocol is equivalent to an event of small probability happening, in the following sense:

1. When Skeptic follows a measurable strategy (a rule that gives $s_n$ as a function of $y_1, \ldots, y_{n-1}$),
$$\mathsf{P}\{\sup_n \mathcal{K}_n \geq \frac{1}{\epsilon}\} \leq \epsilon \tag{3}$$
for every $\epsilon > 0$. (This is because the capital process $\mathcal{K}_0, \mathcal{K}_1, \ldots$ is a non-negative martingale; Equation (3) is sometimes called *Doob's inequality*.)

2. If $A$ is a measurable subset of $\{0,1\}^\infty$ with $\mathsf{P}(A) \leq \epsilon$, then Skeptic has a measurable strategy that guarantees
$$\liminf_{n\to\infty} \mathcal{K}_n \geq \frac{1}{\epsilon}$$
whenever $(y_1, y_2, \ldots) \in A$.

Loosely: Skeptic's being able to multiply his capital by a factor of $1/\epsilon$ or more is equivalent to the happening of an event with probability $\epsilon$ or less.

Ville made the generality of his ideas clear. They apply whenever prices are regular conditional expected values for a known joint probability distribution P for a sequence of random variables $Y_1, Y_2, \ldots$:

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
    Skeptic announces $s_n : \mathbb{R} \to \mathbb{R}$ such that
        $\mathsf{E}(s_n(Y_n) | Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1})$ exists.
    Reality announces $y_n \in \mathbb{R}$.
        $\mathcal{K}_n := \mathcal{K}_{n-1} + s_n(y_n) - \mathsf{E}(s_n(Y_n) | Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1})$.

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

Cournot's principle in game-theoretic terms:

**Finitary.** Instead of saying that an event of small probability singled out in advance will not happen, we say that a strategy chosen by Skeptic, if it avoids risk of bankruptcy, will not multiply his capital by a large factor.

**Infinitary.** Instead of saying that an event of zero probability singled out in advance will not happen, we say that a strategy chosen by Skeptic, if it avoids risk of bankruptcy, will not make him infinitely rich.

Like Kolmogorov, Ville was inspired by von Mises.

Von Mises considered a sequence $y_1, y_2, \ldots$ of 0s and 1s random if no subsequence with a different frequency of 1s can be picked out by a gambler to whom the $y$s are presented sequentially. This would keep the gambler from getting rich by deciding when to bet.

Ville showed that von Mises's condition is insufficient. It does not rule out the gambler's getting rich by varying the direction and amount to bet.

The best source of information on game-theoretic probability in English is www.probabilityandfinance.com.

- Excerpts from our 2001 book.

- Reviews and responses.

- Working papers on $\sqrt{dt}$, history of Cournot's principle, and defensive forecasting.

We need to add references to Japanese work!