A betting interpretation for probabilities and Dempster-Shafer degrees of belief

Glenn Shafer

April 2, 2010

Brest

Jeyzy Neyman's statistical philosophy:

inductive behavior

As a statistician, I am in the business of telling clients things with 95% confidence.

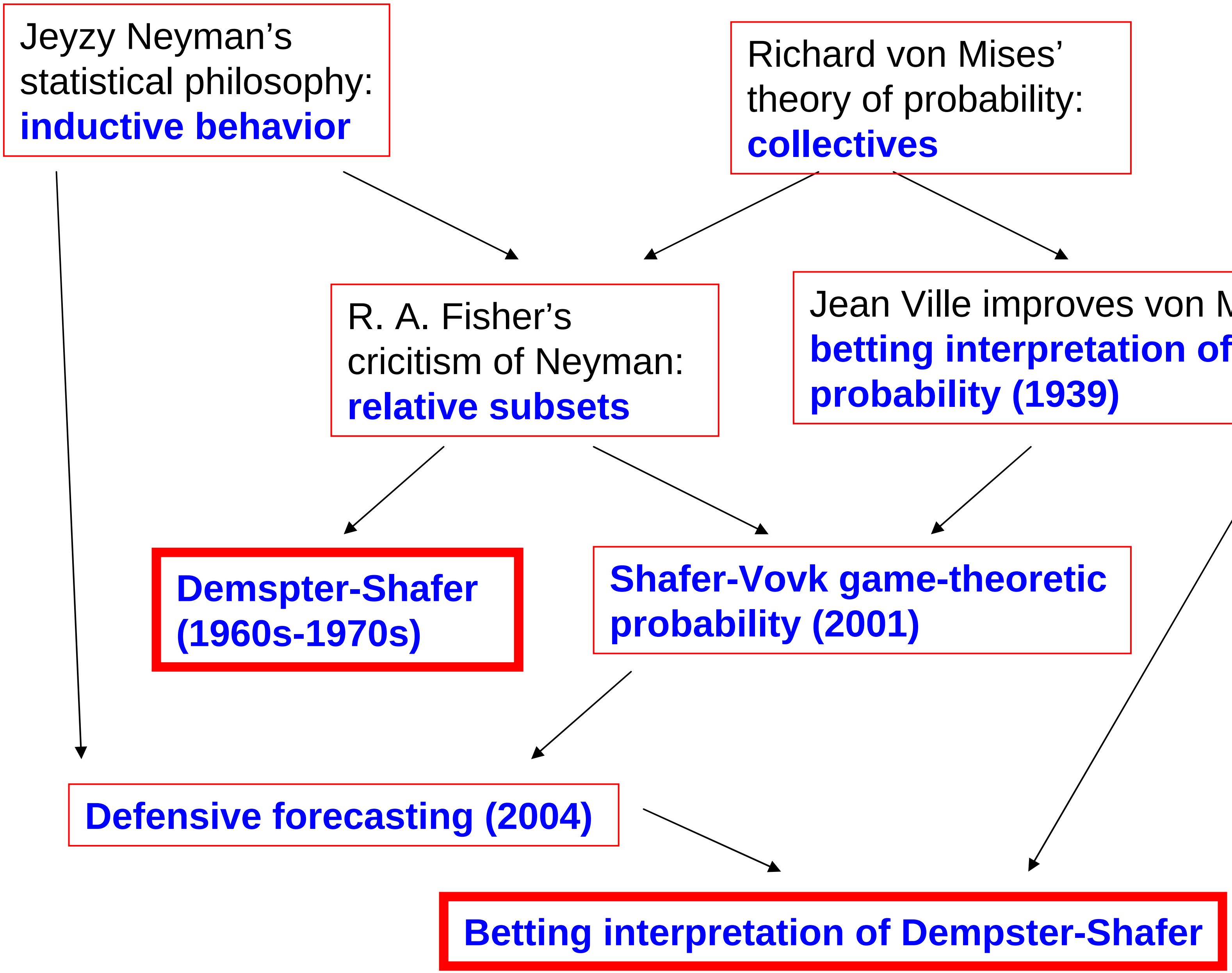
My goals:

be informative

be right 95% of the time

Question: Why isn't this good enough for a theory of evidence?

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


Jeyzy Neyman's statistical philosophy: inductive behavior
Richard von Mises' theory of probability: collectives
R. A. Fisher's cricitism of Neyman: relative subsets
Jean Ville improves von Mises: betting interpretation of probability (1939)
Demspter-Shafer (1960s-1970s)
Shafer-Vovk game-theoretic probability (2001)
Defensive forecasting (2004)
Betting interpretation of Dempster-Shafer

Jeyzy Neyman's statistical philosophy: **inductive behavior**

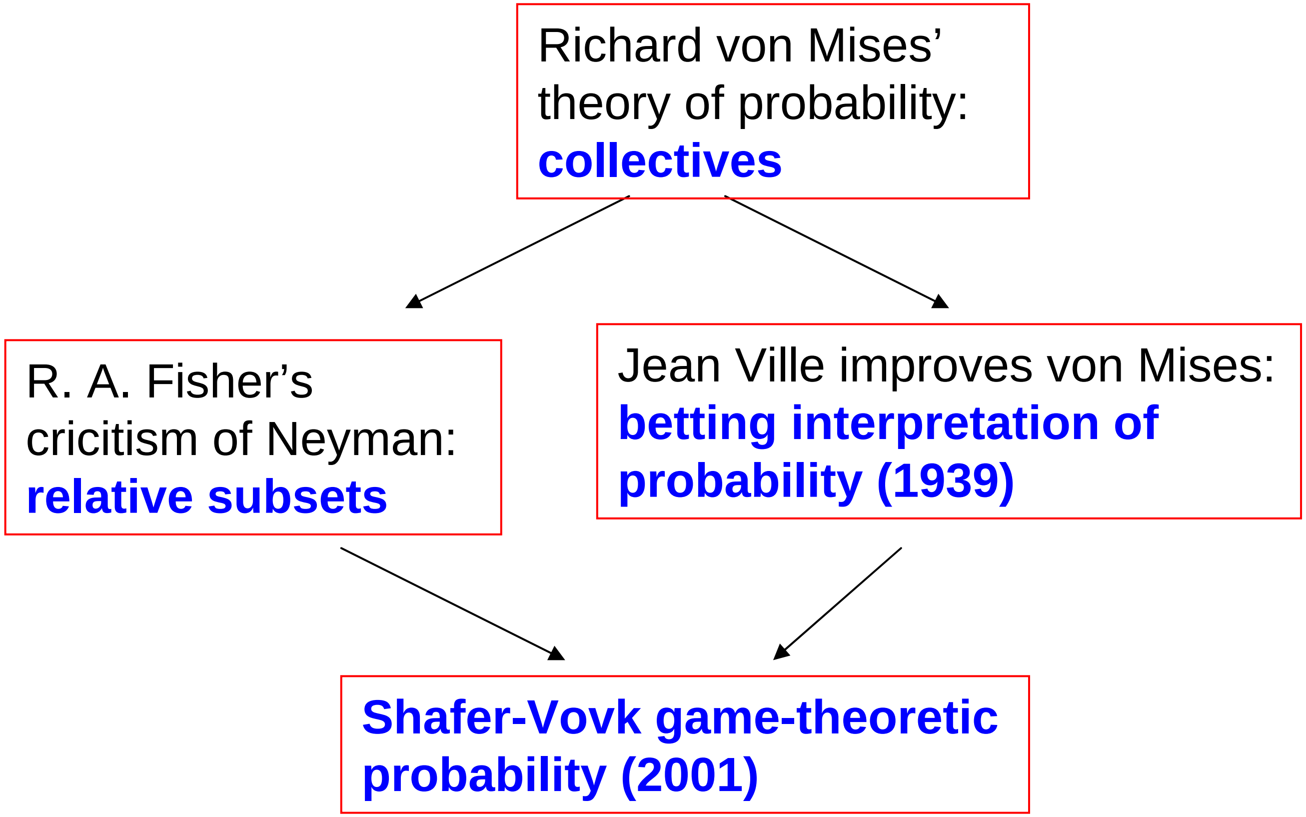
R. A. Fisher's cricitism of Neyman: **relative subsets**

Demspter-Shafer (1960s-1970s)

Ronald A. Fisher
1890–1962

Jerzy Neyman
1894–1981

Dempster-Shafer methods for combining multiple observations from a parametric model do not solve the relevant subsets problem, either.



Ville's betting interpretation of probability:

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

Jeyzy Neyman's
statistical philosophy:
inductive behavior

Ville's betting interpretation of probability:

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

Shafer-Vovk game-theoretic probability (2001)

Defensive forecasting (2004)

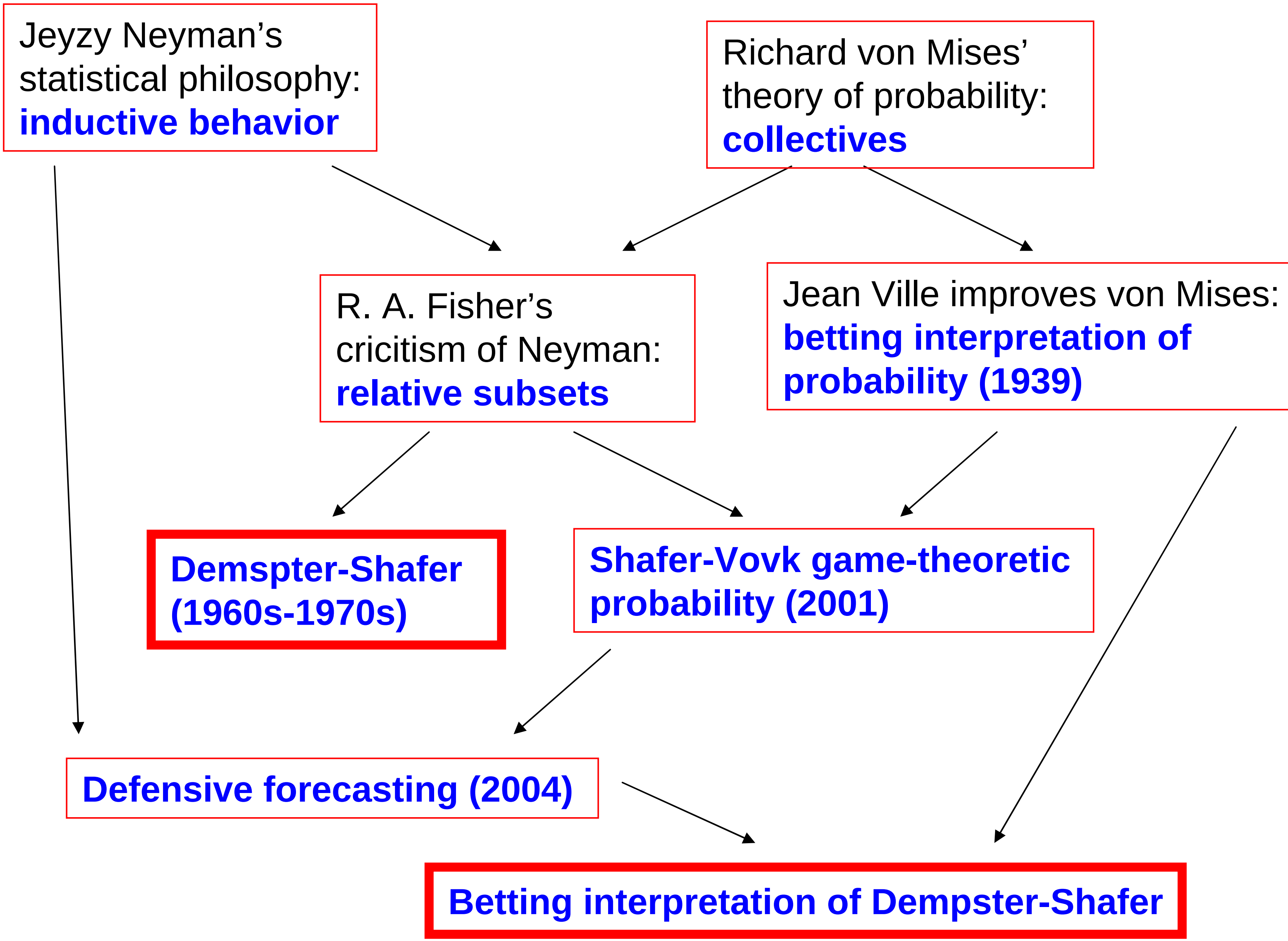
It turns out that you can make forecasts that pass Ville's tests
for a specified sequence of forecasting tasks.

Jean Ville improves von Mises: **betting interpretation of probability (1939)**

How can Ville’s interpretation handle the case of two defensive forecasters who tell the court different or even contradictory things?

Defensive forecasting (2004)

Betting interpretation of Dempster-Shafer


Jeyzy Neyman's statistical philosophy: inductive behavior
Richard von Mises' theory of probability: collectives
R. A. Fisher's cricitism of Neyman: relative subsets
Jean Ville improves von Mises: betting interpretation of probability (1939)
Demspter-Shafer (1960s-1970s)
Shafer-Vovk game-theoretic probability (2001)
Defensive forecasting (2004)
Betting interpretation of Dempster-Shafer

Two ways of interpreting degrees of belief in terms of betting:

De Finetti: Offer to bet at the odds defined by the degrees of belief.

Ville: Judge that a strategy for taking advantage of such betting offers will not multiply the capital it risks by a large factor.

Both can justify updating ordinary probabilities by conditioning.

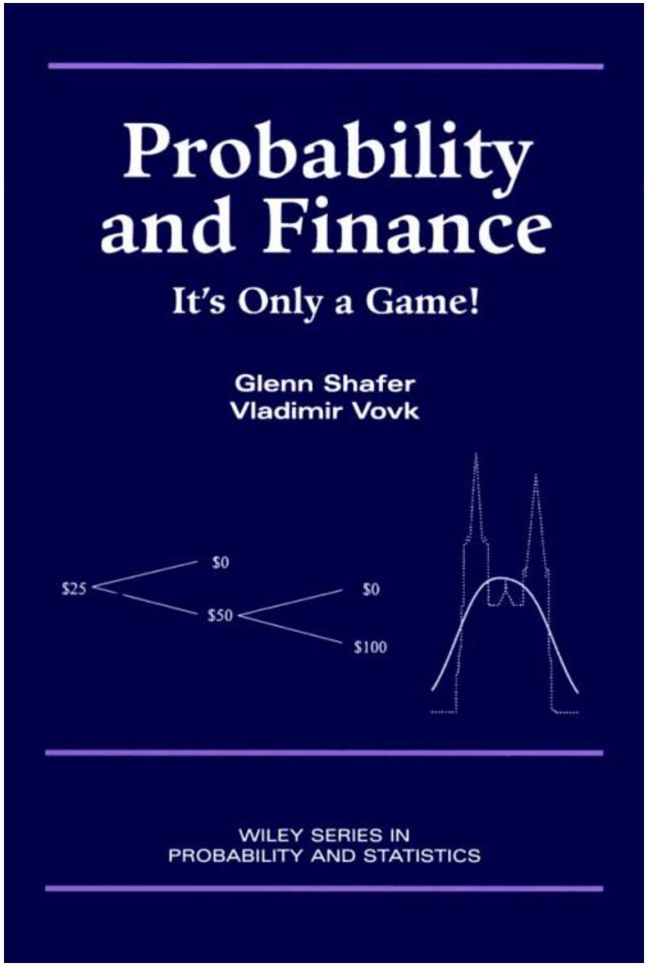
Only Ville can justify Dempster's rule of combination.

http://arxiv.org/abs/1001.1653

Ville is natural in game-theoretic probability, where P(A) is the cost of a ticket that pays $1 if A happens.

Game-Theoretic Probability

www.probabilityandfinance.com



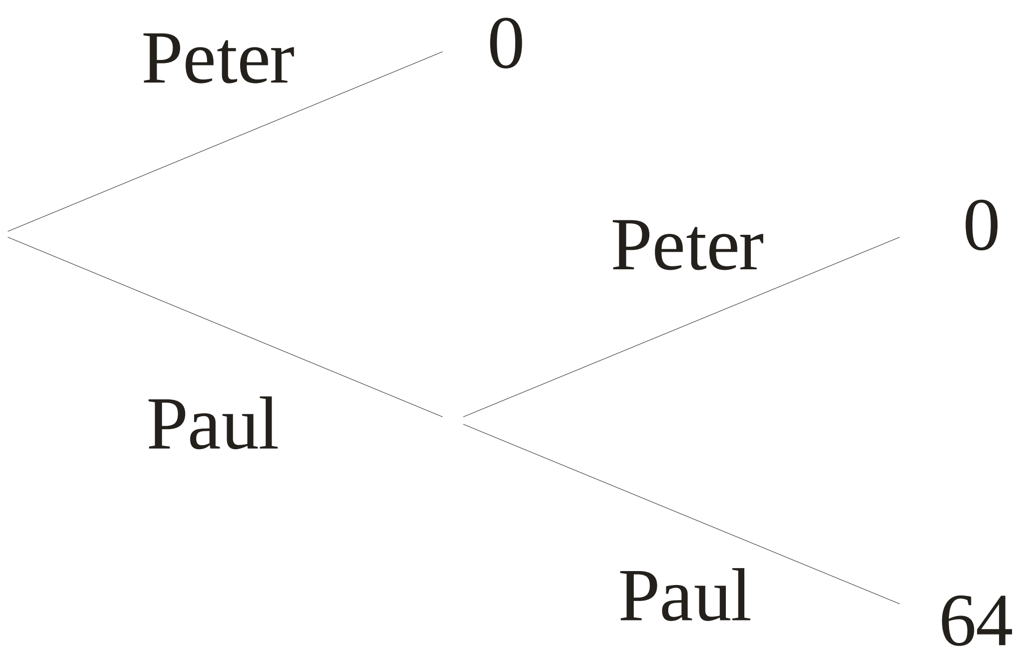
Glenn Shafer & Vladimir Vovk
Wiley, 2001

Second edition planned for 2012.

Use game theory instead of measure theory as a mathematical framework for probability.

Classical theorems are proven by betting strategies that multiply a player's stake by a large factor if the theorem's prediction fails.

Pascal’s question to Fermat

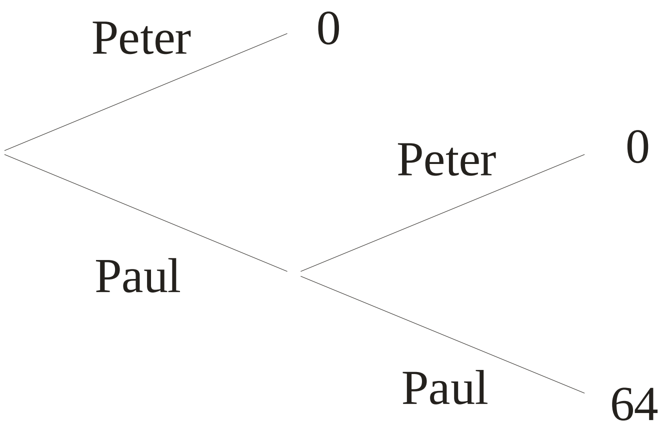


Paul needs 2 points to win.
Peter needs only one.

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:

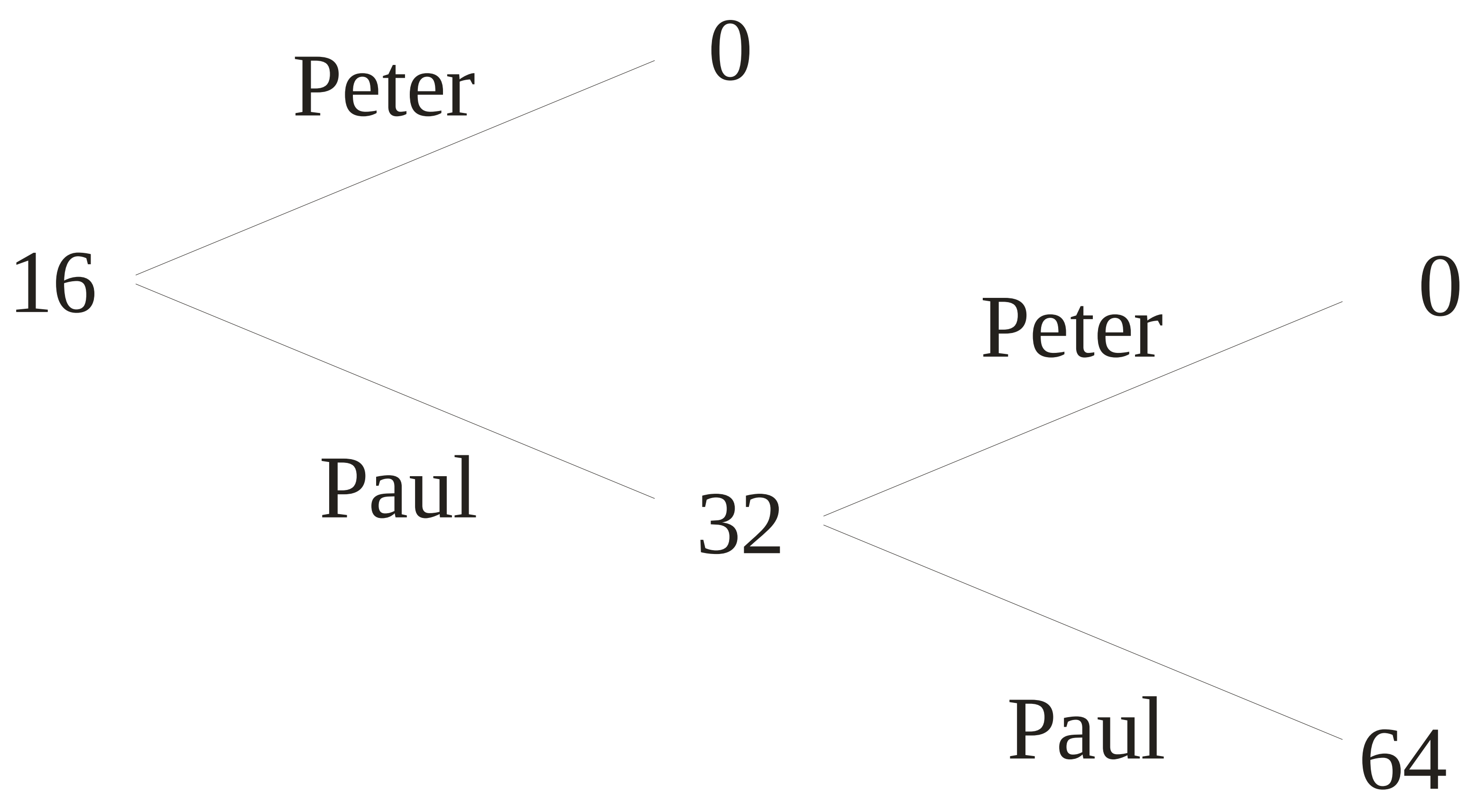
1. Peter wins first, Peter wins second
2. Peter wins first, Paul wins second
3. Paul wins first, Peter wins second
4. Paul wins first, Paul wins second

Paul wins only in outcome 4. So his share should be ¼, or 16 pistoles.

Pascal didn’t like the argument.

Pierre Fermat, 1601-1665

Pascal's answer (game theory)



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

Probability of A = total of probabilities for elementary events that favor A.

Game-theoretic probability begins with a game:
- One player offers prices for uncertain payoffs.
- Another player decides what to buy.

Probability of A = initial stake you need in order to get 1 if A happens.

Interpretation of game-theoretic probability

Mathematical definition of probability:
P(A) = cost of getting $1 if A happens

Version 1. An event of very small probability will not happen. (Cournot's principle)
Version 2. You will not multiply your capital by a large factor without risking bankruptcy. (Efficient market hypothesis / Ville's principle)

Borel, Kolmogorov, and others advocated version 1 between 1900 and 1950.

Jean Ville stated version 2 in the 1930s.

Measure-theoretic probability		Game-theoretic probability	
			
Emile Borel 1871-1956	Andrei Kolmogorov 1903-1987	Jean André Ville 1910-1989 (Borel's student)	Volodya Vovk 1960- (Kolmogorov's student)

You do the mathematics of probability by finding the measures of sets.

You do the mathematics of probability by finding betting strategies.

Theorems

We prove a claim (e.g., law of large numbers) by constructing a strategy that multiplies the capital risked by a large factor if the claim fails.

Statistics

A statistical test is a strategy for trying to multiply the capital risked.

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

Example of a game-theoretic probability theorem.

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

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

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

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

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

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

Skeptic wins if

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

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

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

Theorem Skeptic has a winning strategy.

Ville's 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.

As an empirical theory, game-theoretic probability makes predictions: A will not happen if there is a strategy that multiplies your capital without risking bankruptcy when A happens.

Defensive forecasting:

Amazingly, predictions that pass all statistical tests are possible (defensive forecasting).

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 Hilary Putnam thought good probability prediction is impossible. . .

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

Reality can make Forecaster uncalibrated by setting

$$y_n := \begin{cases} 1 & \text{if } p_n < 0.5 \\ 0 & \text{if } p_n \geq 0.5 \end{cases}$$

Skeptic can then make steady money with

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

But 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.

Nuances:

1. When a probabilistic theory successfully predicts a long sequence of future events (as quantum mechanics does), it tells us something about phenomena.

2. When a probabilistic theory predicts only one step at a time (basing each successive prediction on what happened previously), it has practical value but tells us nothing about phenomena. Defensive forecasting pass statistical tests *regardless of how events come out*.

3. When we talk about the probability of an isolated event, which different people can place in different sequences, we are weighing arguments. This is the place of evidence theory.

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

Assumptions

1. $\mathrm{P}(A) =$ price of a ticket that pays 1 if A happens.
2. $\mathrm{P}(A)x =$ 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), $\mathrm{P}(B|A)x =$ price of a ticket that pays x if B happens.

Argument

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

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

Assumptions

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

Argument

1. You pay me $\mathsf{P}(A)\mathsf{P}(B|A)$ to get $\mathsf{P}(B|A)$ if A happens. If A does happen, you pay me $\mathsf{P}(B|A)$ to get 1 if B also happens.
2. So $\mathsf{P}(A)\mathsf{P}(B|A)$ 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 $\mathrm{P}(B|A) = \frac{\mathrm{P}(A\&B)}{\mathrm{P}(A)}$

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 $\mathrm{P}(A)$ and $\mathrm{P}(A\&B)$.

Conclusion You still will not win if after A (and nothing else) is known, $\mathrm{P}(A\&B)/\mathrm{P}(A)$ 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 $\mathrm{P}(A)$ and $\mathrm{P}(A\&B)$ alone that risks the same risks and payoffs.

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

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

to $\mathcal{S}$'s purchases in the initial situation.

- The tickets added have zero total initial cost:

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

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

$$\begin{array}{ll} 0 & \text{if } A \text{ does not happen;} \\ -M\dfrac{\mathsf{P}(A\&B)}{\mathsf{P}(A)} & \text{if } A \text{ happens but not } B\text{;} \\ M\left(1-\dfrac{\mathsf{P}(A\&B)}{\mathsf{P}(A)}\right) & \text{if } A \text{ and } B \text{ both happen.} \end{array}$$

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

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

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

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

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

Example: The sometimes reliable witness

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

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

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

- Joe says "Glenn paid."

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

- New beliefs:

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

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

Example: The more or less precise witness

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

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

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

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

- Bill says "Glenn paid \$ 5."

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

- New beliefs:

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

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

Conditioning

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

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

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

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

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

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

Example: The witness caught out

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

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

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

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

- Tom says "Glenn paid \$ 10."

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

- New beliefs:

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

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

Independence

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

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

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

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

Product measure:

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

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

Nothing novel here. Dempsterian independence = Cournotian independence.

Example: Independent contradictory witnesses

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

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

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

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

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

Dempster's rule (independence + conditioning)

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

  If $\mathbf{X}_1 = x_1$, then $\omega \in \Gamma_1(x_1)$. If $\mathbf{X}_2 = x_2$, then $\omega \in \Gamma_2(x_2)$.
- So if $(\mathbf{X}_1, \mathbf{X}_2) = (x_1, x_2)$, then $\omega \in \Gamma_1(x_1) \cap \Gamma_2(x_2)$.
- Conditioning on what is not ruled out,

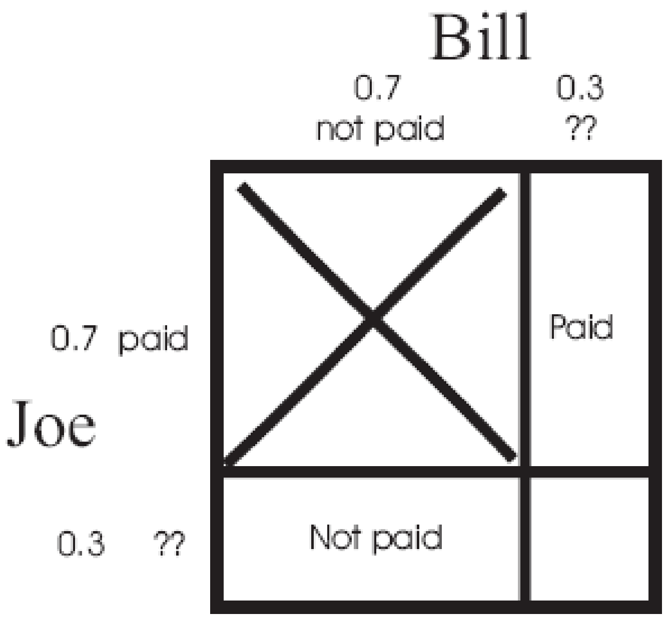
$$\mathbb{B}(A) = \frac{\mathbb{P}\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2) \subseteq A\}}{\mathbb{P}\{(x_1, x_2) | \emptyset \neq \Gamma_1(x_1) \cap \Gamma_2(x_2)\}}$$

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

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

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

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



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

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

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

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

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

Two advertisements

1. *Electronic Journal for History of Probability and Statistics*

2. Upcoming workshop on game-theoretic probability, Royal Holloway, June 21-23

Electronic Journal
for
History of Probability and Statistics

www.jehps.net

June 2009 issue: History of Martingales

Who knew this journal exists?

GTP 2010

Third Workshop on Game-Theoretic Probability and Related Topics

21 - 23 June 2010

Royal Holloway, University of London

Egham, Surrey, UK

http://www.clrc.rhul.ac.uk/GTP2010/