The Language of Betting as a Strategy for Scientific Communication

Glenn Shafer

gshafer@business.rutgers.edu

Rutgers Experimental Mathematics Seminar

February 14, 2019

Hill 705, 5:00 PM to 5:48 PM

The language of betting
as a strategy for scientific communication

Point 1. The conventional vocabulary for statistical testing is too complicated for scientific communication. We can communicate statistical results better using the language of betting.

Point 2. We can communicate even better using fully defined betting games.

Point 3. We can also avoid the fantasy of many worlds.

Betting language can make statistical conclusions appear less objective, and this can play into the hands of those who think science is their enemy. But confusion about statistics also weakens science.

The language of betting can

- clarify what statistical studies can and cannot accomplish, and
- clarify the games scientists must and do play – honest games that are essential to the advancement of knowledge.

This is in the spirit of Andrew Gelman and John Carlin's conclusion that the only solution to the crisis about p-values is "to move toward a greater acceptance of uncertainty and embracing of variation".

Point 1

The conventional vocabulary for statistical testing is too complicated for scientific communication.

We can communicate statistical results better using the language of betting.

Conventional language for testing P against Q

Conventional concept	Conventional explanation
likelihood ratio	$\frac{Q(\text{observation})}{P(\text{observation})}$
significance level: $\alpha = P(\text{test rejects } P)$	Probability, given P is true, that test will err by rejecting it by chance
power $= Q(\text{test rejects } P)$	When power is small, test can reject only by chance.
p-value $= P(T \geq t)$. T is test statistic; t is its observed value.	Probability, if P is true, of getting a result as extreme as the one observed.

Too complicated for scientific communication:

- Most teachers of statistics and researchers who use p-values cannot correctly answer questions about p-values.
- Power is ignored in most applications.

Consider a typical medical research study, for example designed to test the efficacy of a drug, in which a null hypothesis H_0 ('no effect') is tested against an alternative hypothesis H_1 ('some effect'). Suppose that the study results pass a test of statistical significance (that is P-value <0.05) in favor of H_1. What has been shown?

1. H_0 is false.
2. H_0 is probably false.
3. H_1 is true.
4. H_1 is probably true.
5. Both (1) and (3)
6. Both (2) and (4)
7. None of the above. √

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say, 20 subjects in each sample). Furthermore, suppose you use a simple independent means t-test and your result is significant ($t = 2.7$, $df = 18$, $p = .01$). Please mark each of the statements below as "true" or "false." "False" means that the statement does not follow logically from the above premises. Also note that several or none of the statements may be correct.

(1) You have absolutely disproved the null hypothesis (i.e., there is no difference between the population means). False
(2) You have found the probability of the null hypothesis being true. False
(3) You have absolutely proved your experimental hypothesis (that there is a difference between the population means). False
(4) You can deduce the probability of the experimental hypothesis being true. False
(5) You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision. False
(6) You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions. False

- Blakeley B. McShane & David Gal (2017). Statistical significance and the dichotomization of evidence, *Journal of the American Statistical Association* **112**(519):885-895.
- Gerd Gigerenzer (2018). Statistical rituals: The replication delusion and how we got there, *Advances in Methods and Practices in Psychological Science* **1**(2):198-218.

Conventional language for testing P against Q

Conventional concept	Conventional explanation	Betting interpretation
likelihood ratio	$\frac{Q(\text{observation})}{P(\text{observation})}$	Betting score
significance level: $\alpha = P(\text{test rejects } P)$	Probability, given P is true, that test will err by rejecting it by chance	Inverse of betting score for a winning all-or-nothing bet
power $= Q(\text{test rejects } P)$	When power is small, test can reject only by chance.	When power is small, betting score can be large only by chance.
p-value $= P(T \geq t)$. T is test statistic; t is its observed value.	Probability, if P is true, of getting a result as extreme as the one observed.	Inverse of betting score for an all-or-nothing bet that cheats

Definition of *betting score*.

- You claim phenomenon Y is described by probability distribution P.
- You back this up by offering me any payoff $f(Y)$ for its expected value under P.
- I buy a nonnegative f with expected value \$1.
  Because f is nonnegative, I am risking only this initial \$1.
- Given the outcome $Y = y$, my payoff $f(y)$ is my *betting score* — the factor by which I multiplied the money I risked.
- A large betting score $f(y)$ is the best evidence I could have against P.
  I bet against P and won.
- But the possibility that I was merely lucky remains in view.
  There is no better way to communicate the remaining uncertainty.

Notation introduced by Markov in 1900:

- Y is an unknown quantity.
- y is a particular value for Y.

The amount I risk is so small that I do not care about losing it.

No decision theory here. No utility. No Bayesian reasoning.

Conventional concept	Conventional explanation	Betting interpretation
likelihood ratio	$\frac{Q(\text{observation})}{P(\text{observation})}$	Betting score

likelihood ratio = betting score

- I buy a nonnegative f with expected value 1 with respect to P.
- So $\sum_y f(y)P(y) = 1$.
- So Q is a probability distribution, where $Q(y) := f(y)P(y)$.
- So the betting score $f(y)$ is equal to the likelihood ratio $\frac{Q(y)}{P(y)}$.

Your bet defines an alternative hypothesis Q.

likelihood ratio = betting score

- I buy a nonnegative f with expected value 1 with respect to P.
- So $\sum_y f(y)P(y) = 1$.
- So Q is a probability distribution, where $Q(y) := f(y)P(y)$.
- So the betting score $f(y)$ is equal to the likelihood ratio $\frac{Q(y)}{P(y)}$.

Your bet defines an alternative hypothesis Q.

By Gibb's inequality, $Q(Y)/P(Y)$ is optimal for testing P against Q, in the sense that

$$\mathbf{E}_Q\left(\ln \frac{Q(Y)}{P(Y)}\right) \geq \mathbf{E}_Q\left(\ln \frac{R(Y)}{P(Y)}\right)$$

for any other probability distribution R.

The logarithm is pertinent because scores from successive tests multiply. Logarithmic loss is used in information theory and machine learning for similar reasons.

Conventional concept	Conventional explanation	Betting interpretation
likelihood ratio	$\frac{Q(\text{observation})}{P(\text{observation})}$	Betting score
significance level: $\alpha = P(\text{test rejects } P)$	Probability, given P is true, that test will err by rejecting it by chance	Inverse of betting score for a winning all-or-nothing bet

The significance level α in betting language

- Following the Neyman-Pearson theory, I choose E such that $P(E) = \alpha$ and reject if E happens.
- To explain this in betting language, I bet α that E will happen: I pay $\$\alpha$ and get back \$1 if E happens and \$0 if it does not.
- Or I pay \$1 and get back $\$(1/\alpha)$ if E happens and \$0 if it does not:

$$f(y) = \begin{cases} \frac{1}{\alpha} & \text{if } y \in E \\ 0 & \text{if } y \notin E. \end{cases}$$

- This all-or-nothing bet f is usually not optimal, because the probability distribution $f \times P$ is usually not the most plausible alternative hypothesis.

Conventional concept	Conventional explanation	Betting interpretation
significance level: $\alpha = P(\text{test rejects } P)$	Probability, given P is true, that test will err by rejecting it by chance	Inverse of betting score for a winning all-or-nothing bet
power $= Q(\text{test rejects } P)$	When power is small, test can reject only by chance.	When power is small, betting score can be large only by chance.

Power is usually ignored in social science!

A test that is not powerful: Test whether a coin is fair by flipping it 100 times.

Let Y be the number of heads. Its variance is 5. You reject at 5% if $|Y - 50|$ is greater than two standard deviations—i.e., $Y < 40$ or $Y > 60$.

If $Y = 61$, you reject. But this rejection is just luck if the alternative gives heads probability 0.52, because then the power is only about 6%.

Thinking about power is too hard for practitioners and goes against their desire to publish.

Betting language makes the possibility of "just luck" harder to ignore.

The likelihood ratio is

$$\frac{52^{61}48^{39}}{50^{100}} = 2.2$$

You multiply your money by 2.2, not by 20.

Conventional concept	Conventional explanation	Betting interpretation
p-value $= P(T \geq t)$. T is test statistic; t is its observed value.	Probability, if P is true, of getting a result as extreme as the one observed.	Inverse of betting score for an all-or-nothing bet that cheats

A p-value always cheats.
You cannot bet on $T(Y) \geq T(y)$ before you know y.

Betting on a p-value

The p-value from a test statistic $T(Y)$ is

$$p(y) = P(T(Y) \geq T(y)).$$

To bet on the p-value being small, buy a payoff $f(p(Y))$ with expected value 1 or less under P.

My favorite, easy to remember and calculate, is

$$f(p) = \begin{cases} \frac{2}{\sqrt{p}} & \text{if } p \leq \frac{1}{16} \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbf{E}(f(p(Y))) \leq 1$ because $P(p(Y) \leq p) \leq p$.

p-value	$\frac{1}{\text{p-value}}$	$f(\text{p-value})$
0.10	10	0
0.05	20	8.9
0.01	100	20
0.005	200	28
0.001	1000	63

The strategy of betting as a strategy for scientific communication

Point 1. The conventional vocabulary for statistical testing is too complicated for scientific communication. We can communicate statistical results better using the language of betting.

Point 2. We can communicate even better using fully defined betting games.

Point 3. We can also avoid the fantasy of many worlds.

Point 2

We can communicate even better using fully defined betting games.

Probability theory using betting games

Protocol for tossing fair coin:

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots, 100$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $y_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \frac{1}{2})$.

- Perfect-information: players see each others' moves as they are made.
- $\mathcal{K}$ is Skeptic's capital.
- $y = 1$ means heads; $y = 0$ means tails.
- When $M > 0$, Skeptic is betting on heads.
- When $M < 0$, Skeptic is betting on tails.

Define a law-of-large-numbers game by giving a rule for who wins:

Skeptic wins if all $\mathcal{K}_n > 0$ and either
$\mathcal{K}_{100} \geq 20$ or
$\left|\overline{y}_{100} - \frac{1}{2}\right| \leq 0.1$.
Otherwise Reality wins.

Skeptic has a winning strategy in this game!

How a probability distribution represents a phenomenon: An event of small probability, such as $\left|\overline{y}_{100} - \frac{1}{2}\right| \leq 0.1$, will not happen.

How a protocol represents a phenomenon: An event that allows Skeptic to multiply capital he risks by a large factor, such as $\left|\overline{y}_{100} - \frac{1}{2}\right| \leq 0.1$, will not happen.

Statistics using betting games is a little more complicated.

We can communicate even better using fully defined betting games.

1. The game may be partly hidden from the statistician.
   - R. A. Fisher assumed only partial knowledge of the probabilities describing a phenomenon. The statistician knows only that the true distribution is in a known class $(P_\theta)_{\theta \in \Theta}$.
   - Similarly, game-theoretic statistics assumes that the statistician sees only some of the moves in a betting game.
2. Betting offers may fall short of a probability distribution.
   - A probability distribution for Y prices every payoff $f(Y)$.
   - Some betting games give fewer prices.

Protocol where betting offers fall short of a probability distribution.

$\mathcal{K}_0 := 1.$
FOR $n = 1, 2, \ldots, 100$:
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $\epsilon_n \in [-1, 1]$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \epsilon_n$.

Think of Reality's moves as errors of measurement.

In 1821, Gauss assumed that:

- errors are bounded between certain limits, and
- errors that are equal but of opposite signs are equally likely.

- On the nth round, Skeptic can buy ϵ_n in any amount (positive, negative, or zero) at the price 0.
- The values of previous ϵ do not matter.

Protocol where betting offers fall short of a probability distribution.

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

FOR $n = 1, 2, \ldots, 100$:

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

  Reality announces $\epsilon_n \in [-1, 1]$.

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

Make this protocol a game by giving a rule for who wins.

Skeptic wins if $\mathcal{K}_1, \ldots, \mathcal{K}_{100} \geq 0$ and either $K_{100} \geq 20$ or $|\overline{\epsilon}_{100}| \leq 0.272$.

This follows from the game-theoretic form of Hoeffding's inequality; see Section 3.3 of *Game-Theoretic Foundations for Probability and Finance*.

Skeptic has a winning strategy in this game.

So statistician can bet 19 to 1 that $|\overline{\epsilon}_{100}| \leq 0.272$.

$\mathcal{K}_0 := 1.$
FOR $n = 1, 2, \ldots, 100$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $\epsilon_n \in [-1, 1]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \epsilon_n$.

Skeptic wins if $\mathcal{K}_1, \ldots, \mathcal{K}_{100} \geq 0$ and either $K_{100} \geq 20$ or $|\overline{\epsilon}_{100}| \leq 0.272$.

Now add α and $\epsilon_1, \ldots, \epsilon_{100}$ to the protocol.

It remains a perfect-information protocol: Both Reality and Skeptic see α and the ϵ_n.

$\mathcal{K}_0 := 1.$
Reality announces $\alpha \in \mathbb{R}^K$.
FOR $n = 1, 2, \ldots, 100$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $\epsilon_n \in [-1, 1]$ and sets $y_n := \alpha + \epsilon_n$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \epsilon_n$.

The statistician stands outside the protocol. She sees the x_n and the y_n, but she does see α or the ϵ_n.

Skeptic's 19 to 1 bet that $|\overline{\epsilon}_{100}| \leq 0.272$ is now a 19 to 1 bet that $|\overline{y}_{100} - \alpha| \leq 0.272$.

The statistician can share Skeptic's *confidence* in the 19 to 1 bet that α is in the interval $\overline{y}_{100} \pm 0.272$.

The argument generalizes to least squares estimation.

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

Reality announces $\beta \in \mathbb{R}^K$.

FOR $n = 1, 2, \dots, 100$:

  Reality announces $x_n \in \mathbb{R}^K$.

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

  Reality announces $\epsilon_n \in [-1, 1]$ and sets $y_n := \langle \beta, x_n \rangle + \epsilon_n$.

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

dot product

See *Game-Theoretic Foundations for Probability and Finance*:

- Section 10.4 discusses consistency for least square estimates for bounded errors, in the spirit of Lai and Wei (1982).
- Chapter 4 shows how the absolute bound on errors can be replaced by a quadratic hedge.

The games scientists play:

- p-hacking: screening ideas, screening drugs
- multiple testing
- meta-analysis
- the crisis of replication

These issues are not new.

- Fourier published a table of significance levels in 1821.
- Cournot discussed the pitfalls of multiple testing in 1843.

The betting language puts them on the table at the outset.

What game is your laboratory or research group playing? When you reject H_0 at the 5% level, did you risk just one dollar to win 20? Or did you already lose money on other tests or other experiments or other variables before you found a winner? You can claim credit only for <u>the factor by which you multiplied all the money you risked</u>.

What game is the scientific community playing? A meta-analysis must ask whether the second experiment was undertaken only because the first was promising.

The language of betting
as a strategy for scientific communication

Point 1. The conventional vocabulary for statistical testing is too complicated for scientific communication. We can communicate statistical results better using the language of betting.

Point 2. We can communicate even better using fully defined betting games.

Point 3. We can also avoid the fantasy of many worlds.

Point 3

We can also avoid the fantasy of many worlds.

Protocol for tossing fair coin:

$\mathcal{K}_0 = 1.$
FOR $n = 1, 2, \ldots, 100$:
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0, 1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \frac{1}{2})$.

Define a law-of-large-numbers game by giving a rule for who wins:

Skeptic wins if all $\mathcal{K}_n > 0$ and either
$\mathcal{K}_{100} \geq 20$ or
$|\overline{y}_{100} - \frac{1}{2}| \leq 0.1$.
Otherwise Reality wins.

Any probability distribution can be interpreted this way.

The approximation of probability by frequency is only one theorem.

Basic principle is that Skeptic will not multiply the capital he risks by a large factor.

Probabilities may change on every round.

You still have theorems about frequencies.

Protocol for probability forecasting:

$\mathcal{K}_0 = 1.$
FOR $n = 1, 2, \ldots, 100$:
  Reality announces signal$_n$.
  Forecaster or theory announces $p \in [01]$.
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $y_n \in \{0, 1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - p_n)$.

Define a law-of-large-numbers game by giving a rule for who wins:

Skeptic wins if all $\mathcal{K}_n > 0$ and either
$\mathcal{K}_{100} \geq 20$ or
$|\overline{y}_{100} - \overline{p}_{100}| \leq 0.1$.
Otherwise Reality wins.

A physicist using probability in statistical mechanics, quantum mechanics, or cosmology is in the same position as a statistician using probability in medicine or social science.

1. She does not see all the moves in the game.
2. Probability theory does not force her to suppose that the first move is repeated endlessly.

Game-theoretic probability as a strategy for scientific communication

1. The conventional vocabulary for statistical testing (likelihood, significance level, power, p-value, etc.) is too complicated for scientific communication. It is easier to communicate statistical results in terms of betting. A likelihood ratio, for example, is the amount we multiply the capital we risk when we bet against one probabilistic theory using an alternative.
2. We can communicate even better by using fully defined betting games. Betting offers describe a phenomenon if a player cannot use them to multiply the capital he risks by a large factor. Just as R. A. Fisher's theory of statistics begins by supposing that the statistician has only partial knowledge of the probabilities describing a phenomenon, game-theoretic statistics begins by supposing that the statistician sees only some of the moves in the betting game.
3. The offers in a betting game need not include odds on every event or prices for every payoff. This saves game-theoretic probability from the many-world fantasies that we find in some probabilistic treatments of statistical mechanics, quantum mechanics, and cosmology.

References

- **On the difficulty of communicating statistical results:**
  --Blakeley B. McShane & David Gal (2017). Statistical significance and the dichotomization of evidence, *Journal of the American Statistical Association* **112**(519):885-895.
  --Andrew Gelman and John Carlin (2007). Some natural solution to the p-value communication problem—and why they won't work, *Journal of the American Statistical Association* **112**(519):899-901.
- **On Fisher:** David J. Hand (2015). From evidence to understanding: a commentary on Fisher (1922) 'On the mathematical foundations of theoretical statistics', *Phil. Trans. R. Soc. A* **373**.
- **On game-theoretic probability:** Working papers at www.probabilityandfinance.com and *Game-Theoretic Foundations for Probability and Finance*, by Glenn Shafer and Vladimir Vovk (Wiley, May 2019).

Game-theoretic probability and finance come of age

Glenn Shafer and Vladimir Vovk's *Probability and Finance*, published in 2001, showed that perfect-information games can be used to define mathematical probability. Based on fifteen years of further research, *Game-Theoretic Foundations for Probability and Finance* presents a mature view of the foundational role game theory can play. Its account of probability theory opens the way to new methods of prediction and testing and makes many statistical methods more transparent and widely usable. Its contributions to finance theory include purely game-theoretic accounts of Ito's stochastic calculus, the capital asset pricing model, the equity premium, and portfolio theory.

Game-Theoretic Foundations for Probability and Finance is a book of research. It is also a teaching resource. Each chapter is supplemented with carefully designed exercises and notes relating the new theory to its historical context.

Praise from early readers

"Ever since Kolmogorov's *Grundbegriffe*, the standard mathematical treatment of probability theory has been measure-theoretic. In this ground-breaking work, Shafer and Vovk give a game-theoretic foundation instead. While being just as rigorous, the game-theoretic approach allows for vast and useful generalizations of classical measure-theoretic results, while also giving rise to new, radical ideas for prediction, statistics and mathematical finance without stochastic assumptions. The authors set out their theory in great detail, resulting in what is definitely one of the most important books on the foundations of probability to have appeared in the last few decades."
—**Peter Grünwald**, CWI and University of Leiden

"Shafer and Vovk have thoroughly re-written their 2001 book on the game-theoretic foundations for probability and for finance. They have included an account of the tremendous growth that has occurred since, in the game-theoretic and pathwise approaches to stochastic analysis and in their applications to continuous-time finance. This new book will undoubtedly spur a better understanding of the foundations of these very important fields, and we should all be grateful to its authors."
—**Ioannis Karatzas**, Columbia University

Glenn Shafer is University Professor at Rutgers University.

Vladimir Vovk is Professor in the Department of Computer Science at Royal Holloway, University of London. They are the authors of *Probability and Finance: It's Only a Game*, published by Wiley and co-authors of *Algorithmic Learning in a Random World*. Shafer's other previous books include *A Mathematical Theory of Evidence* and *The Art of Causal Conjecture*.

Cover Design: Wiley
Cover Image: © Web Gallery of Art/Wikimedia Commons

www.wiley.com

WILEY

ISBN 978-0-470-90305-6

Shafer
Vovk

Game-Theoretic Foundations for Probability and Finance

WILEY

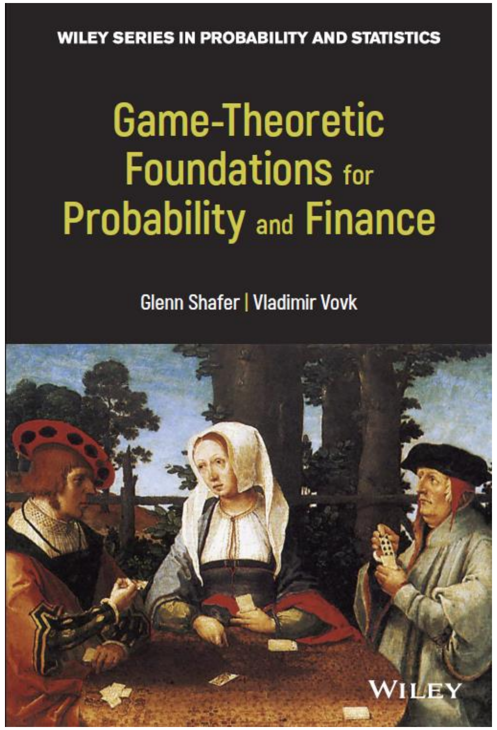
WILEY SERIES IN PROBABILITY AND STATISTICS

Game-Theoretic Foundations for Probability and Finance

Glenn Shafer | Vladimir Vovk

WILEY

May 2019



The Card Players
c. 1520
Lukas van Leyden

Players might be
- Charles V
- Margaret of Austria
- Cardinal Woolsey

Ever since Kolmogorov's *Grundbegriffe*, the standard mathematical treatment of probability theory has been measure-theoretic. In this ground-breaking work, Shafer and Vovk give a game-theoretic foundation instead. While being just as rigorous, the game-theoretic approach allows for vast and useful generalizations of classical measure-theoretic results, while also giving rise to new, radical ideas for prediction, statistics and mathematical finance without stochastic assumptions. The authors set out their theory in great detail, resulting in what is definitely one of the most important books on the foundations of probability to have appeared in the last few decades.

– Peter Grünwald, CWI and University of Leiden

Shafer and Vovk have thoroughly re-written their 2001 book on the game-theoretic foundations for probability and for finance. They have included an account of the tremendous growth that has occurred since, in the game-theoretic and pathwise approaches to stochastic analysis and in their applications to continuous-time finance. This new book will undoubtedly spur a better understanding of the foundations of these very important fields, and we should all be grateful to its authors.

– Ioannis Karatzas, Columbia University