Let's replace p-values with betting outcomes!

Glenn Shafer
gshafer@business.rutgers.edu

PROGIC 2019
Frankfurt, July 9, 2019

When we generalize Neyman-Pearson using betting,

- we see that betting outcomes are likelihood ratios,
- we obtain a new and better concept of power, and
- we better understand the meaning of probability.

Testing by Betting

Hypothesis: P describes random variable Y.

Question: How do we use $Y = y$ to test P?

Conventional answer:

- Choose *significance level* α, say 0.05.
- Choose E such that $P(E) = 0.05$.
- Reject P if $y \in E$.

Hypothesis: P describes random variable Y.

Question: How do we use $Y = y$ to test P?

Conventional answer:

- Choose *significance level* α, say 0.05.
- Choose E such that $P(E) = 0.05$.
- Reject P if $y \in E$.

Betting interpretation: Bet on E.

- Pay \$1.
- If E does not happen, get back \$0.
- If E happens, get back \$20.
  - Then brag that you discredited P.
  - You multiplied your money by a large factor.
  - What better evidence against P could you have?

Question: How do we measure the strength of evidence against P?

Conventional answer:

- Use a test statistic to define a test for each $\alpha \in (0, 1)$.
- The *p-value* is the smallest α for which the test rejects.
- The smaller the p-value, the more evidence against P.

Too complicated!

Question: How do we measure the strength of evidence against P?

Conventional answer:

- Use a test statistic to define a test for each $\alpha \in (0, 1)$.
- The *p-value* is the smallest α for which the test rejects.
- The smaller the p-value, the more evidence against P.

Betting alternative: Instead of an all-or-nothing bet (a bet that pays either \$0 or \$20, say), make a bet on Y that can pay many different amounts.

- Such a bet is a function $S(Y)$.
- Choose S such that $E_P(S) = 1$.
- Pay \$1.
- Get back \$$S(y)$. So $S(y)$ is the factor by which you multiplied your money.
- Call $S(y)$ your *betting score*.
- The larger $S(y)$, the more evidence against P.

- Choose S such that $E_P(S) = 1$.
- Pay \$1. Get back \$$S(y)$.
- Your *betting score* $S(y)$ is the factor by which you multiply your money.

- If $E_P(S) \neq 1$, the betting score is $\frac{S(y)}{E_P(S)}$.

- The betting score does not change when we multiply S by a positive constant.
- You can bet so little that both $E_P(S)$ and $S(y)$ are negligible.
- No decision theory here.
- No need to play with real money.
- It's only a game!

Betting score

= factor by which I multiply money risked.

Large betting score

= best evidence I can have against P.

But maybe I was merely lucky.

Betting language

= best way to communicate uncertainty.

Likelihood Ratios

A betting score, as just defined, is the same thing as a likelihood ratio.

- A bet S is a function of Y satisfying $S \geq 0$ and $\sum_y S(y)P(y) = 1$.
- So SP is also a probability distribution. Call it the alternative Q.
- But $Q(y) = S(y)P(y)$ implies $S(y) = Q(y)/P(y)$.
- A bet against P defines an alternative Q and the betting score $S(y)$ is the likelihood ratio $Q(y)/P(y)$.
- Conversely, if you start with an alternative Q, then Q/P is a bet.

- My bet S defines the alternative hypothesis $Q = SP$, even if I did not think about Q when choosing S. (Perhaps I did not know the theory. Perhaps Q is difficult to calculate.)

- On the other hand, if I begin with an alternative Q, then I can make the bet Q/P.
  Proof that Q/P is a bet: $E_P(Q/P) = 1$, because

$$\sum_y \frac{Q(y)}{P(y)} P(y) = \sum_y Q(y) = 1.$$

But is liking Q any reason to choose Q/P as my bet?

Multiple Testing

You say P describes Y.

I want to bet against you.

I think Q describes Y.

Should I use Q/P as my bet?

$S = Q/P$ maximizes $\mathbf{E}_Q(\ln S)$.

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

Kullback-Leibler divergence
Gibbs's inequality

Why maximize $\mathbf{E}_Q(\ln S)$? Why not $\mathbf{E}_Q(S)$? Or $Q(S \geq 20)$?

Neyman-Pearson lemma

When S is the product of successive factors, $\mathbf{E}(\ln S)$ measures the rate of growth (Kelly, 1956). This has been used in gambling theory, information theory, finance theory, and machine learning. Here it opens the way to a theory of multiple testing and meta-analysis.

Successive tests of P

- P purports to describe $Y_1, Y_2, \ldots$.
- I test P by buying $S_1(Y_1)$ for \$1. Betting score $S_1(y_1)$ is mediocre — not much larger than 1.
- I continue testing. Score $S_2(Y_2)$ again mediocre.

Two ways of filling out the story

- I made the second bet by taking another \$1 out of my wallet. So I risked \$2. Final betting score is the mediocre

$$\frac{S_1(y_1) + S_2(y_2)}{2}.$$

- I made the second bet risking the winnings from the first. Final betting score is

$$S_1(y_1)S_2(y_2).$$

The second way is more powerful. So aim for large $S_1(y_1)S_2(y_2)$ rather than large $S_1(y_1) + S_2(y_2)$.

Replace power with *implied target*.

The *implied target* of the test $S = Q/P$ is $\exp(E_Q(\ln S))$.

$$\mathbf{E}_Q(\ln S) = \sum_y Q(y) \ln S(y) = \sum_y P(y) S(y) \ln S(y) = \mathbf{E}_P(S \ln S)$$

Use the implied target to evaluate the test in advance.

Even if I do not take *Q* seriously, my critics will.

Why should the editor invest in my test if it is unlikely to produce a high betting score even when it is optimal?

Elements of a study that tests a probability distribution by betting

	name	notation
Proposed study		
initially unknown outcome	phenomenon	Y
probability distribution for Y	null hypothesis	P
nonnegative function of Y with expected value 1 under P	bet	S
$S \times P$	implied alternative	Q
$\exp\left(\mathbf{E}_Q(\ln S)\right)$	implied target	S^*
Results		
actual value of Y	outcome	y
factor by which money risked has been multiplied	betting score	$S(y)$

Two Examples

Example 1

1. P says Y is normal, mean 0, standard deviation 10.
2. Q says Y is normal, mean 1, standard deviation 10.
3. Statistician A uses the Neyman-Pearson bet with
666 $\alpha = 0.0015$, which rejects P when $y > 29.68$. Power=6%

Implied targets

4. Statistician B uses likelihood ratio

1.10

$$\frac{q(y)}{p(y)} = \exp\left(\frac{2y-1}{200}\right).$$

5. We observe $y = 30$.
6. A multiplies money by $1/0.0015 \approx 666$.
7. B multiplies money by $\exp(59/200) \approx 1.34$.

Example 2

1. P says Y is uniform on $[0,1]$; $p(y) = 1$ for $y \in [0,1]$.
2. Q says Y has density $q(y) = 121y^{120}$ for $y \in [0,1]$.
3. Statistician A uses the Neyman-Pearson bet with $\alpha = 0.05$, which rejects P when $y \geq 0.95$.

20 Power 99.8%

Implied targets

4. Statistician B uses likelihood ratio

45

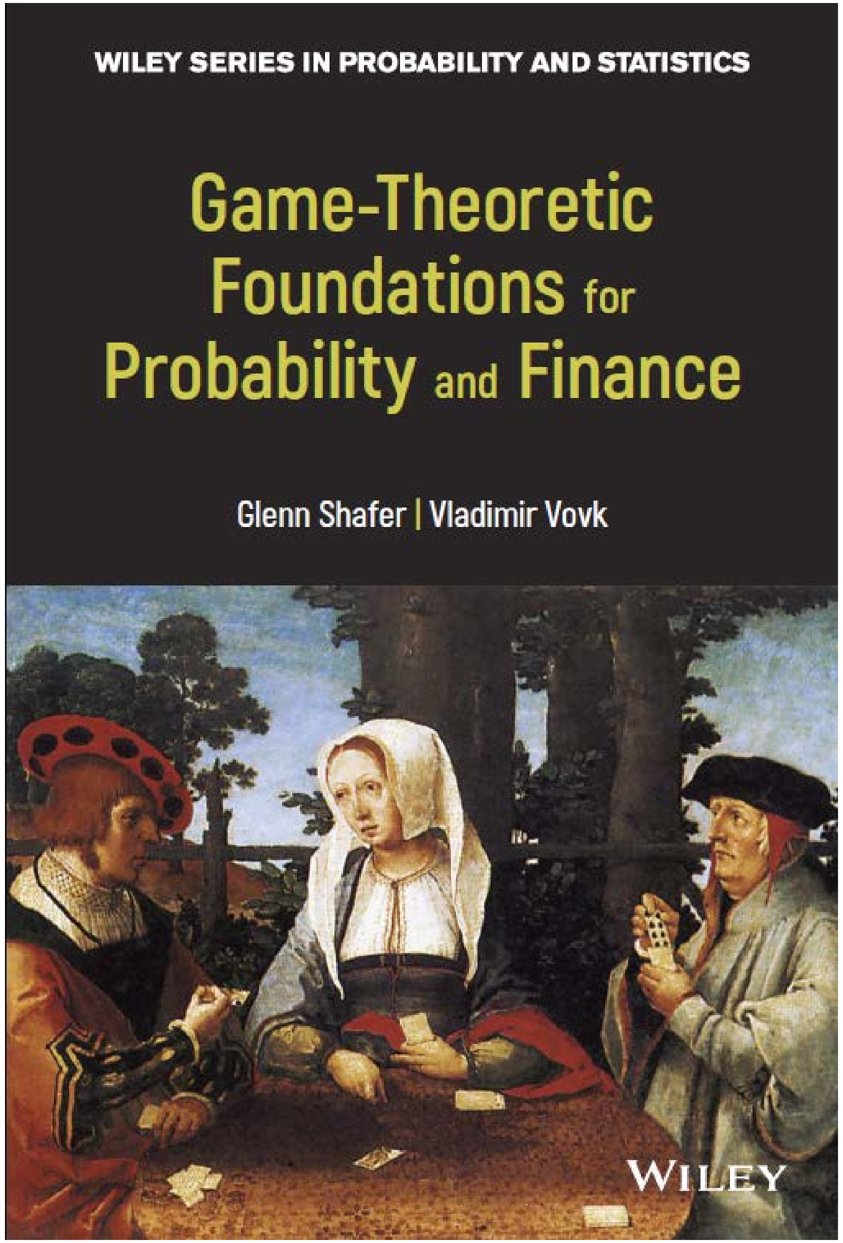
$$\frac{q(y)}{p(y)} = 121y^{120}$$

5. We observe 0.95.
6. A multiplies money by $1/0.05 = 20$.
7. B multiplies money by $121(0.95)^{120} \approx 0.25$.

References

This talk is based on my paper, "The Language of Betting as a Strategy for Statistical and Scientific Communication."

http://probabilityandfinance.com/articles/54.pdf



Our book locates the meaning of a probability model in its resistance to betting tests.

This interpretation extends to imprecise probability models.

Optional Stopping

Optional Continuation

With this interpretation of probability models, **optional stopping comes free**.

Bet as you please.

If the model makes sequential predictions, you can improvise as you go along.

- You need not adopt a strategy in advance.
- You can stop whenever you want.
- Then you can decide to start again.

But don’t cheat:

- Don’t pretend you made a bet that you did not make.
- Don’t pretend you stopped if you actually continued and lost the money.

Apply this thinking to meta-analysis:

One team of scientists obtains a betting score $S_1(y_1)$. Another team decides that the result is promising but not conclusive. So they do a larger test (more subjects, higher implied target), obtaining a betting score $S_2(y_2)$.

The overall betting score is $S_1(y_1)S_2(y_2)$. But the two teams did not have a joint strategy at the outset of the story.

Summary

Probability is about betting, even when it is used to describe phenomena.

In the quest for objectivity, we have created a confusing language (p-value, etc.) that pushes betting into the background.

The language of betting can better communicate

- the meaning of probability,
- the strength of statistical evidence.