One way testing by betting can improve data analysis:

Statistical testing with optional continuation

Glenn Shafer
Rutgers University

Department of Civil & Environmental Engineering
University of Maryland
September 13, 2023

Is it legitimate to deliberate on the basis of initial data about whether and how to collect and analyze further data?

Standard statistical testing says NO.

Game-theoretic probability says YES.

Game-Theoretic Foundations for Probability and Finance,
Glenn Shafer and Vladimir Vovk, Wiley, 2019

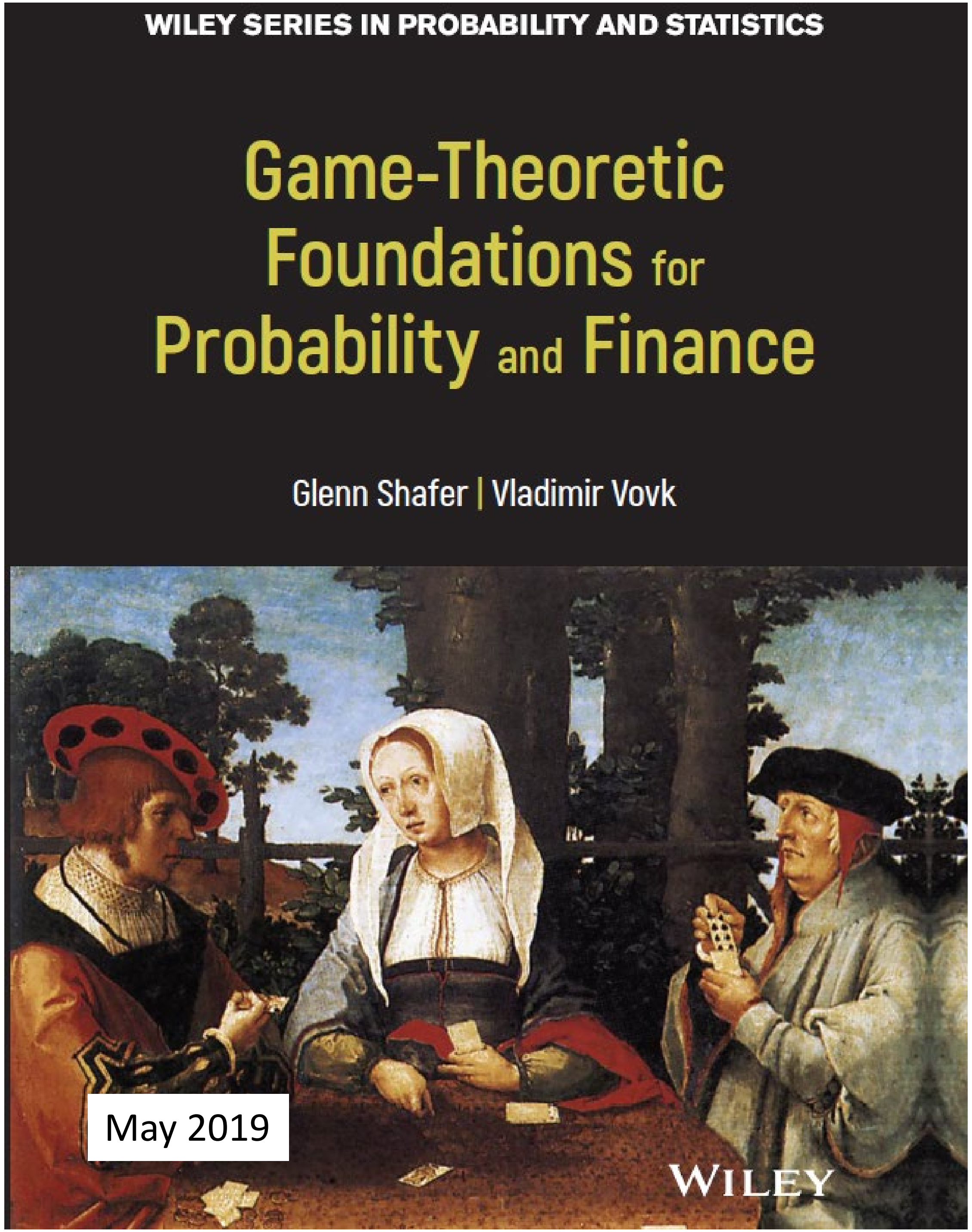
Two ways game-theoretic probability can improve data analysis,
Glenn Shafer, 2023

1. How testing by betting works

2. How standard statistical testing works

3. Why standard statistical testing does not handle optional continuation as well

1. How testing by betting works



Base mathematical probability on testing by betting.

Theorems in measure theory become theorems in game theory.

Related working papers at www.probabilityandfinance.com

Royal Statistical Society paper, [Testing by betting: a strategy for statistical and scientific communication]()

Basic idea of game-theoretic probability

Game with three players:

Forecaster gives probability.

Skeptic selects bet.

Reality decides outcome.

Maybe a *Bayesian*.

Forecaster gives probability.

Skeptic selects bet.

Reality decides outcome.

A *frequentist*, or some sort of *objectivist*, because he is testing whether the probabilities are consistent with Reality.

Forecaster gives probability.

Skeptic selects bet.

Reality decides outcome.

Skeptic is testing whether the probability is consistent with Reality.

The more money Skeptic makes, the more the probability (and Forecaster) are discredited.

Many rounds

Forecaster gives probabilities.

Skeptic selects bets.

Reality decides outcomes.

The more money Skeptic makes,
the more Forecaster is discredited.

Forecaster gives probabilities.

Skeptic selects bets.

Reality decides outcomes.

Fundamental principle for testing by betting

Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the final capital is large.

Fundamental principle

Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the final capital is large.

We assume unit capital for just for convenience.

Wordier fundamental principle

Successive bets against a forecaster that begin with positive capital and never risk more discredit the forecaster to the extent that the final capital is a large multiple of the initial capital.

Fundamental principle

Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the final capital is large.

Optional continuation

- You can decide how to bet as you go along.
- You can stop (or continue) whenever you want.
- But: You are cheating if you pretend you stopped earlier and take the earlier capital as your final capital.

Do you prefer mathematical symbols to words?

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

Forecaster and Skeptic might be the same person (perhaps a statistician).

Forecaster and Reality might be the same person (perhaps a market).

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

We assume *perfect information*:

the players move in order and see each other's moves.

This is a standard concept in game theory.

Additional information, possibly private, is not ruled out.

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

We can make explicit additional information that all the players see.

PARAMETER: $N \in \mathbb{N}$, set $\mathcal{X}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

Somebody announces $x_n \in \mathcal{X}$,

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

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

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

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

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

A large value of $\mathcal{K}_n/\mathcal{K}_0$ will not discredit Forecaster unless Skeptic always chooses M_n so that $\mathcal{K}_n \geq 0$ no matter how Reality chooses y_n.

PARAMETER: $N \in \mathbb{N}$

$\mathcal{K}_0 = 1$.

For $n = 1, \ldots, N$:

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

Skeptic announces $M_n \in [-\mathcal{K}_{n-1}/(1 - p_n), \mathcal{K}_{n-1}/p_n]$.

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

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

Here “begin with unit capital and never risk more” is built into the protocol.

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

Forecaster and Skeptic are free players, not required to follow any strategy.

Now we impose a strategy on Forecaster: always announce p.

PARAMETER: $N \in \mathbb{N}$ and $p \in [0, 1]$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

This is a protocol for testing a probability p with repeated trials.

Protocol for testing a probability p with repeated trials.

$$
\begin{aligned}
&\textsc{Parameter:}\ N \in \mathbb{N} \text{ and } p \in [0,1] \\
&\quad \text{Skeptic announces } \mathcal{K}_0 \in \mathbb{R}. \\
&\quad \text{For } n = 1, 2, \dots, N: \\
&\qquad \text{Skeptic announces } M_n \in \mathbb{R}. \\
&\qquad \text{Reality announces } y_n \in \{0,1\}. \\
&\qquad \mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - p).
\end{aligned}
$$

We can also use a stochastic process as Forecaster's strategy.

$$
\begin{aligned}
&\textsc{Parameters:}\ N \in \mathbb{N}, \text{ stochastic process } \mathbf{P} \text{ on } \{0,1\}^N \\
&\quad \mathcal{K}_0 = 1. \\
&\quad \text{For } n = 1, \dots, N: \\
&\qquad \text{Skeptic announces } M_n \in \mathbb{R}. \\
&\qquad \text{Reality announces } y_n \in \{0,1\}. \\
&\qquad \mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mathbf{P}(y_n = 1 \mid y_1, \dots, y_{n-1})).
\end{aligned}
$$

PARAMETER: $N \in \mathbb{N}$

Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots, N$:

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

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

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

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

Here forecaster gives just one probability on each round.

PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

$\mathcal{K}_0 = 1$.

For $n = 1, \ldots, N$:

Forecaster announces probability distribution P_n on $\mathcal{Y}$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

Generalization where Forecaster gives a probability distribution on each round.

PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

$\mathcal{K}_0 = 1$.

For $n = 1, \dots, N$:

Forecaster announces probability distribution P_n on $\mathcal{Y}$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

A version more adapted to experimentation.

PARAMETER: $N \in \mathbb{N}$

$\mathcal{K}_0 = 1$.

For $n = 1, \dots, N$:

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

A little more complicated…

PARAMETERS: $N \in \mathbb{N}$, set $\mathcal{X}$

$\mathcal{K}_0 = 1.$

For $n = 1, \ldots, N$:

Somebody announces $x_n \in \mathcal{X}$.

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

PARAMETERS: $N \in \mathbb{N}$, set $\mathcal{X}$

$\mathcal{K}_0 = 1$.

For $n = 1, \dots, N$:

Somebody announces $x_n \in \mathcal{X}$.

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

Optional continuation is built into the picture.

No one is required to choose a strategy at the outset.

To emphasize optional continuation, we could replace N with ∞.

PARAMETERS: $N \in \mathbb{N}$, set $\mathcal{X}$

$\mathcal{K}_0 = 1$.

For $n = 1, \dots, N$:

Somebody announces $x_n \in \mathcal{X}$.

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

Optional continuation includes optional stopping.

Skeptic can stop by choosing constants.

Experimenter can stop by choosing singletons.

Quantum mechanics as an example (from Section 10.6 of Shafer & Vovk's 2019 book)

PARAMETER: t_0 (initial time)

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

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

Observer announces an observable A_n and a time $t_n > t_{n-1}$.

Quantum Mechanics announces $P_n \in \mathcal{P}(\mathbb{R})$.

Skeptic announces $f_n \in [0, \infty]^{\mathbb{R}}$ such that $P_n(f_n) = 1$.

At time t_n, Reality announces the measurement $y_n \in \mathbb{R}$.

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

1. How testing by betting works

2. How standard statistical testing works

3. Why standard statistical testing does not handle optional continuation as well

2. How standard statistical testing works

How do we discredit a probability distribution P using observations?

Classical answer:

- Select an event E with small probability $P(E)$.
- P is discredited if E happens.

Call this *Cournot's principle*.

Cournot's Principle

Specify in advance an event E with probability α. If E happens, then we may take α as a measure of evidence against P.

Discredit measured by

how small α is, i.e.

how large $1/\alpha$ is.

Call $1/\alpha$ the *test score*:

$$\text{test score} = \begin{cases} 1/\alpha & \text{if } E \text{ happens} \\ 0 & \text{if } E \text{ does not happen.} \end{cases}$$

Fundamental principle for testing by betting

Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the final capital is large.

Cournot's principle

The happening of an event chosen in advance discredits the probability distribution to the extent that the event's probability is small.

How are these two principles related?

It's complicated.

Single bet

Cournot's Principle

Specify in advance an event E with probability α. If E happens, then we may take α as a measure of evidence against P.

Discredit measured by

how small α is, i.e.

how large $1/\alpha$ is.

Call $1/\alpha$ the *test score*:

$$\text{test score} = \begin{cases} 1/\alpha & \text{if } E \text{ happens} \\ 0 & \text{if } E \text{ does not happen.} \end{cases}$$

Cournot is equivalent to testing by betting with a single bet.

Betting against a stochastic process

PARAMETERS: $N \in \mathbb{N}$, stochastic process $\mathbf{P}$ on $\{0,1\}^N$

$\mathcal{K}_0 = 1$.

For $n = 1, \dots, N$:

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

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

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mathbf{P}(y_n = 1 \mid y_1, \dots, y_{n-1}))$.

A strategy for Skeptic that does not risk more than the initial capital makes $\mathcal{K}_N$ a nonnegative random variable.

Markov's inequality: $\mathbf{P}\left(\mathcal{K}_N \geq \frac{1}{\alpha}\right) \leq \alpha$ for every $\alpha \geq 0$.

Testing a stochastic process by betting using a strategy chosen in advance is consistent with Cournot.

But conservative in a certain sense.

Infinite horizon

PARAMETER: Stochastic process $\mathbf{P}$ on $\{0,1\}^\infty$

$\mathcal{K}_0 = 1$.

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

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

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

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mathbf{P}(y_n = 1 \mid y_1, \ldots, y_{n-1}))$.

A strategy for Skeptic that does not risk more than the initial capital makes $\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2, \ldots$ a nonnegative martingale.

Ville's inequality: $\mathbf{P}\left(\sup_n \mathcal{K}_n \geq \frac{1}{\alpha}\right) \leq \alpha$ for every $\alpha \geq 0$.

Even if you cheat by pretending you stopped earlier, testing by betting using a strategy chosen in advance is still consistent with Cournot and more conservative.

1. How testing by betting works

2. How standard statistical testing works

3. Why standard statistical testing does not handle optional continuation as well

3. Why standard statistical testing does not handle optional continuation as well

PARAMETER: Stochastic process $\mathbf{P}$ on $\{0,1\}^\infty$

$\mathcal{K}_0 = 1.$

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

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

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

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - \mathbf{P}(y_n = 1 \,|\, y_1, \dots, y_{n-1}))$.

Fundamental principle for testing by betting

Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the final capital is large.

The fundamental principle authorizes more than Cournot. Skeptic can be a free player, not tied down to a strategy chosen in advance.

True optional stopping: Skeptic can change his mind about when to stop.

More: Skeptic can change his mind about how to bet.

No stochastic process here

PARAMETERS: $N \in \mathbb{N}$, set $\mathcal{X}$

$\mathcal{K}_0 = 1$.

For $n = 1, \ldots, N$:

Somebody announces $x_n \in \mathcal{X}$.

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces random variable V_n with finite $\mathbf{E}_{P_n}(V_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(y_n) - \mathbf{E}_{P_n}(V_n)$.

The fundamental principle for testing by betting authorizes true optional continuation.

The statistician, acting as Experimenter and Skeptic, can deliberate at each step about the next experiment and the next bet.

Quote from R. A. Fisher:

> The present use of the term sequential is intended to be of a broader import than the formal use of the word as associated with the systematic procedure known as sequential analysis. The experimenter does not regard his material as wholly passive but instead looks to what may be learnt from it with a view to the improvement and extension of the enquiry. This willingness to learn from it how to proceed is the essential quality of sequential procedures. Wald introduced the sequential test, but the sequential idea is much older. For example, what is the policy of a research unit? It is that in time we may learn to do better and follow up our more promising results. The essence of sequential experimentation is a series of experiments each of which depends on what has gone before. For example, in a sample survey scheme, as explained by Yates, a pilot survey is intended to supply a basis for efficiently planning the subsequent stages of a survey.

“Sequential experimentation”, *Biometrics* 8:183—187, 1952 arXiv:2308.14959

Two ways game-theoretic probability can improve data analysis

Glenn Shafer

Abstract: When testing a statistical hypothesis, is it legitimate to deliberate on the basis of initial data about whether and how to collect further data? Game-theoretic probability's fundamental principle for testing by betting says yes, provided that you are testing by betting and do not risk more capital than initially committed. Standard statistical theory uses Cournot's principle, which does not allow such optional continuation. Cournot's principle can be extended to allow optional continuation when testing is carried out by multiplying likelihood ratios, but the extension lacks the simplicity and generality of testing by betting. Game-theoretic probability can also help us with descriptive data analysis. To obtain a purely and honestly descriptive analysis using competing probability distributions, we have them bet against each other using the Kelly principle. The place of confidence intervals is then taken by a sets of distributions that do relatively well in the competition. In the simplest implementation, these sets coincide with R. A. Fisher's likelihood intervals.