University of Kansas School of Business

October 20, 2017

# Three principles for using probability

1. The fiducial principle
2. Poisson's principle
3. Cournot's principle

**Glenn Shafer**
**Rutgers University**
**www.glennshafer.com**

The paper is inspired by the Bayesian, Frequentist, Fiducial movement, which will hold its 5th annual meeting at the University of Michigan next spring. See https://sph.umich.edu/biostat/events/bff-conference.html.

I have elaborated on the paper's ideas in http://www.probabilityandfinance.com/articles/50.pdf

Non-Bayesian statistical methods are often called *frequentist*.

I call them *Bernoullian*, in honor of Jacob Bernoulli.

Name previously used by

- Francis Edgeworth
- Richard von Mises
- Arthur Dempster
- Ian Hacking

# What was R. A. Fisher's fiducial argument?

We measure a quantity $\theta$ with error. Write $x$ for the result and $u$ for its error:

$$x = \theta + u. \qquad \text{(structural equation)}$$

$$\theta = x - u \qquad \text{(fiducial equation)}$$

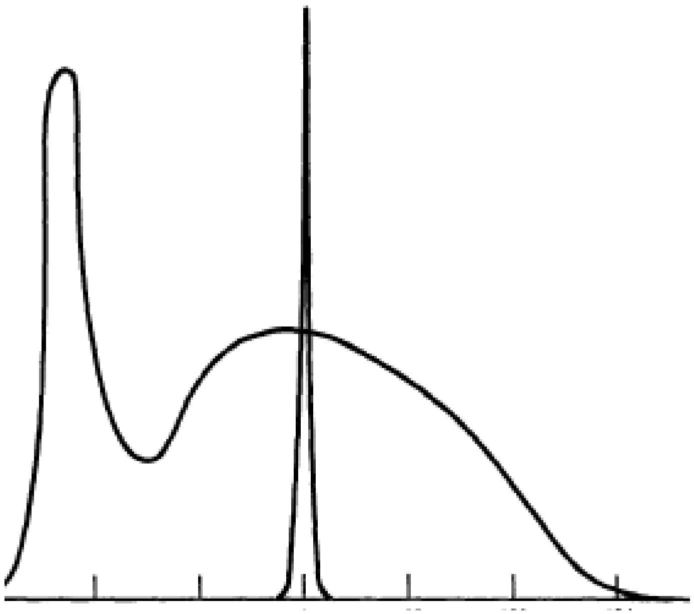
$$u = x - \theta. \qquad \text{(pivot equation)}$$

- We think $u$ is normal with mean zero and variance one.
- If we observe $x = 2.3$, say, then $\theta = 2.3 - u$.
- If we continue to trust our probabilities for $u$, then $\theta$ is normal with mean 2.3 and variance one. This is $\theta$'s *fiducial probability distribution.*

## Why is the fiducial argument coming back?

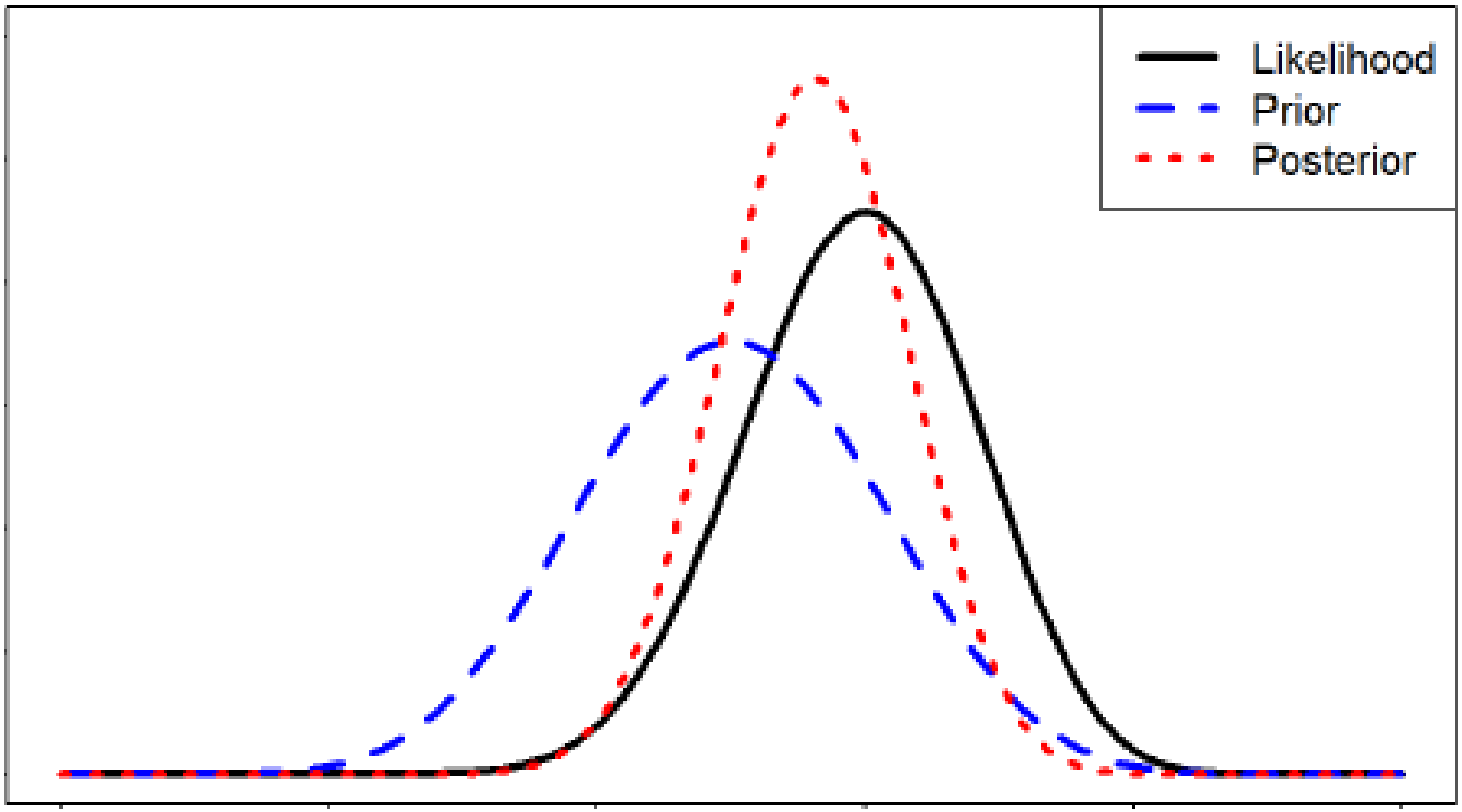
Because it is hard to find Bayesian methods with good Bernoullian properties when there are many, many parameters.

**Principle of stable estimation:** The prior doesn't matter much if you have a large sample (and only one parameter).

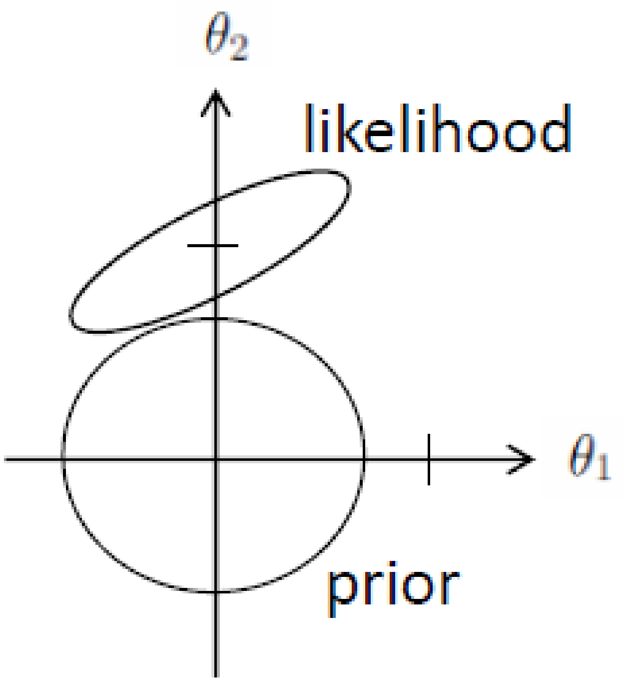


The bimodal curve is the prior density. The spiked curve is the likelihood function, which can be expected, when there many observations, to have the form of a normal density. A normal density is effectively zero outside an interval extending a few standard deviations from its peak. So the posterior density, which is proportional to the product of the prior density and the likelihood, will also be zero outside this interval. When their are many observations, the interval is very narrow. On the reasonable assumption that the prior density is approximately constant over this narrow interval, it will make little difference.

**Principle of compromise:** Even if your sample is small, the posterior is a compromise between the prior and the likelihood (if there is only one parameter).



These principles fail when $\theta$ is many-dimensional, say $\theta = (\theta_1, \ldots, \theta_n)$, and you are interested in a particular $\theta_k$ or some other feature $h(\theta)$.




The circle is a contour for the prior density. The tilted ellipse is a contour for the likelihood function. Both suggest that 0 is the most likely value for $\theta_1$.

The posterior density, being proportional to product of the prior and the likelihood, is greatest in the region where the two contours come closest, suggesting a negative value for $\theta_1$.

Far from being exceptional, this failure to compromise arises for some feature $h(\theta_1, \theta_2)$ whenever the prior density and likelihood function are tilted with respect to each other.

**New models in many fields, from medicine to macroeconomics, have far more parameters than observations.**

Bayesian analyses can go wrong in ways we do not understand.

Some statisticians are abandoning Bayes in favor of fiducial arguments that can be shown to have targeted Bernoullian properties.

Paul Romer. The trouble with macroeconomics. `https://paulromer.net/wp-content/uploads/2016/09/WP-Trouble.pdf`, September 14, 2016. To appear in *The American Economist*. 15

1. **The fiducial principle**

2. **Poisson’s principle**

3. **Cournot’s principle**

# 1. The fiducial principle

All use of mathematical probability is fiducial.

It requires a decision to trust particular probabilities in a particular situation…

…even though these probabilities are initially hypothetical, theoretical, subjective, or derived from other situations, which can never be identical to the situation at hand in all respects.

Fiducial principle: All probability is fiducial.

Probabilities come from theory, from conjecture, or from experience of frequencies.

There is always other information. The fiducial move is to judge that this other information can be neglected.

# Who was fiducial?

- Bernoulli (1713). The estimation of probability from frequency is fiducial.
- Bayes (1763). Fallacious 5$^{th}$ proposition is a fiducial argument.
- Laplace (1770s). His principle of inverse probability (also called Bayes's rule) was fiducial.
- Fisher (1930). Invented fiducial probability by inverting a continuous cumulative distribution function.
- Dempster (1967). Extended Fisher's fiducial argument to the discrete case, obtaining upper and lower probabilities.

# De Moivre's argument for $$\mathbf{P}(A\&B) = \mathbf{P}(A) \times \mathbf{P}(B|\text{ after } A),$$ where $A$ and $B$ are successive events

Meaning of the probabilities:

- $\mathbf{P}(A)$ = price of a ticket that pays 1 if $A$ happens.
- $\mathbf{P}(A)x$ = price of a ticket that pays $x$ if $A$ happens.
- $\mathbf{P}(B|\text{ after } A)$ = price after $A$ happens of a ticket that pays $x$ if $B$ happens.

Argument

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

# Bayes's fiducial argument

Suppose $B$ is *subsequent* to $A$. ($A$ is decided first.)

Bayes adopted De Moivre's game-theoretic proof that the probability of $B$ after $A$ has happened, $P(B|A)$, should satisfy $P(B|A) = P(A \cap B)/P(A)$.

He also wanted to prove that if you learn $B$ without knowing about $A$, you should change $P(A)$ to $P(A \cap B)/P(B)$.

The "proof" imagines a sequence $(A_1, B_1), (A_2, B_2), \ldots$ and posits that your $A$ is the one with the first $B$ that happens.

This sneaks in a FIDUCIAL judgement: Being told $B$ happened is independent of (irrelevant to) whether $A$ happened.

Dempster's rule of combination makes this same fiducial judgement in a more general framework.

# Bayes's second argument

Realizing that his 5th proposition might not persuade, Bayes added his billiard-table argument.

In the 9th edition of the *Encyclopedia Britannica*, Morgan Crofton simplified the billiard table to a line segment.

Here the fiducial judgement is the independence of the random draw of $p$ from whether subsequent draws fall to the left or right of $p$.

Art Dempster's D-S argument uses the same fiducial judgement of independence.

- The fiducial judgement is always a judgement.
- In a particular case, you may have reason not to make it.
- It may be a serious error.

Cournot on Bayes (1843)

Bayes's rule ... has no utility aside from fixing bets under a certain hypothesis about what the arbiter knows and does not know.

It leads to unfair bets if the arbiter knows more than we suppose about the real conditions of the random trial.

# Whitehead's 1938 puzzle of the two aces

As you watch, I do the following.

- Prepare a deck with four cards: A♠, A♣, 2♠, 2♣
- Shuffle.
- Deal myself two cards.
- Look at them without your seeing. them

$E :=$ {I have two aces} $\quad \mathbf{P}(E) = \frac{1}{6}$

$F_1 :=$ {I have at least one ace} $\quad \mathbf{P}(F_1) = \frac{5}{6}$

$F_2 :=$ {I have A♠} $\quad \mathbf{P}(F_2) = \frac{1}{2}$

Now I smile and say, "I have an ace." Bayes says that you should increase your probability for $E$ to

$$\mathbf{P}(E|F_1) = \frac{\mathbf{P}(E)}{\mathbf{P}(F_1)} = \frac{1}{5}.$$

Now I smile again and say, "I have the ace of spades." Bayes says that you should increase your probability for $E$ further to

$$\mathbf{P}(E|F_2) = \frac{\mathbf{P}(E)}{\mathbf{P}(F_2)} = \frac{1}{3}.$$

Why should my identifying a suit change your probability?

You do not know why I gave you the information.

You need a larger model that includes probabilities for my behavior.

John E. Freund, Puzzle or paradox, *American Statistician* 19(4):29-44, 1965.
N. T. Gridgeman, Letter to the editor, *American Statistician* 21(3):38, 1967.
Glenn Shafer, A subjective interpretation of conditional probability. *Journal of Philosophical Logic* **12** 453-466. 1983

# Some references

- Bayes. 5th proposition is fiducial argument.

  My *Annals of Statistics* 1982, 10(4):1075-1089.
  http://projecteuclid.org/download/pdf_1/euclid.aos/1176345974

- Laplace. Inverse probability was fiducial mistake.

  Steve Stigler, *The History of Statistics*, 1986, Chapter 3.

- Fisher. Invented fiducial probability.

  Sandy Zabell, *Statistical Science*, 1992, 7(3):369-387.
  https://projecteuclid.org/download/pdf_1/euclid.ss/1177011233

# 2. Poisson's principle

Even varying probabilities allow probabilistic prediction.

The law of large numbers does not require iid trials.

Predictions may concern averages or other statistics rather than frequencies.

An interpretation of probability that emphasizes prediction should not, therefore, be called *frequentist*.

# Poisson's principle: Even varying probabilities allow probabilistic prediction.

Law of large numbers does not require iid trials.

**Poisson 1837**

Things of every nature are subject to a universal law that we may call the law of large numbers. … if you observe a very considerable number of events of the same nature, depending on causes that vary irregularly, … you will find a nearly constant ratio …

**Chebyshev 1846**

For $n$ sufficiently large,

$$\left| \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} p_i}{n} \right| \leq \epsilon$$

with probability at least $1 - \alpha$.

**Bernstein/Lévy 1920s/1930s**

For $n$ sufficiently large,

$$\left| \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} \mathrm{E}(x_i | x_1, \ldots, x_{i-1})}{n} \right| \leq \epsilon$$

with probability at least $1 - \alpha$.

After 180 years, Poisson's principle still not central in statistics.

## Fisher's picture still central. On the mathematical foundations of theoretical statistics (1922)

"… a quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole…"

- "This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample."

- "The law of distribution of this hypothetical population is specified by relatively few parameters, …"

# 3. Cournot's principle

Probability acquires objective content only by its predictions.

To predict using probability, you single out an event that has high probability and predict it will happen.

Or, equivalently, you single out an event that has small probability and predict it will not happen.

In 1944, Trygve Haavelmo refounded econometrics by explaining that you can model all your data as one observation.

In 1960, Jerzy Neyman proclaimed that stochastic processes are the future of statistics.

But time series, martingales, and stochastic processes remain peripheral to philosophical discussion of “frequentist” statistics.

Haavelmo 1944:

…it is *not* necessary that the observations should be independent and that they should all follow the same one-dimensional probability law.

It is sufficient to assume that the *whole set* of, say *n*, observations may be considered as *one* observation of *n* variables or a \`sample point‘ following an *n*-dimensional *joint* probability law, the \`existence' of which may be purely hypothetical.

Then, one can test hypotheses regarding this joint probability law, and draw inferences as to its possible form, by means of *one* sample point (in *n* dimensions).

## Fiducial principle: In use, all probability is fiducial.

To use a probability, which begins as a subjective or purely theoretical betting rate, we must judge that other information is irrelevant.

## Poisson's principle: Even varying probabilities allow probabilistic prediction.

The law of large numbers, does not require iid trials.

## Cournot's principle: Probability acquires objective content only by its predictions.

To predict using probability,

- single out event with small probability,
- predict it will not happen.

Bernoulli's Theorem (1713): In a large number of independent trials of an event with probability p,

$$\text{Probability(relative frequency} \approx p) \approx 1.$$

To make sense of the second probability: Interpret a probability close to one, singled out in advance, not as a frequency but as practical certainty.

- Bernoulli brought this notion of practical certainty into mathematical probability.
- Antoine Augustin Cournot (1801-1877) added that it is the only to connect probability with phenomena.

# Cournot's principle: Probability acquires objective content only by its predictions.

To predict using probability,

- single out event with small probability,
- predict it will not happen.

Cournot 1843: The physically impossible event is therefore the one that has infinitely small probability, and only this remark gives substance— objective and phenomenal value—to the theory of mathematical probability

Haavelmo 1944: The class of scientific statements that can be expressed in probability terms is enormous. In fact, this class contains all the `laws' that have, so far, been formulated. For such `laws' say no more and no less than this: The probability is almost 1 that a certain event will occur.

# Unified theory of probability

- Bernoulli estimation
- Cournotian testing
- Bayes
- Dempster-Shafer belief function

1. Construct probability (betting) model from relatively quantifiable evidence.

2. If possible, calculate probabilities close to one and use them to test model.

3. Given additional evidence, judge that it does not change your willingness to make certain bets (this “conditioning” is a fiducial judgement of irrelevance).

4. Interpret these bets as additional predictions.

# Liberation

When we no longer have a random sample from a hypothetical population, Fisher's notion of a parametric model no longer so natural.

Instead of being puzzled that we cannot get probabilities for imagined parameters, reexamine the evidence used to construct the parametric model. Can it be modeled more directly by limited bets?

This leads to imprecise and game-theoretic probability.

Details and references in working papers at www.probabilityandfinance.com:

48. Cournot in English

http://www.probabilityandfinance.com/articles/48.pdf

49. Game-theoretic significance testing

http://www.probabilityandfinance.com/articles/49.pdf

50. Bayesian, fiducial, frequentist

http://www.probabilityandfinance.com/articles/50.pdf

# The game-theoretic framework for probability

1. Pascal's definition of probability: Given that certain bets are offered, the probability of an event is the amount you must risk in order to get one monetary unit if the event happens.

2. Game-theoretic version of Cournot's principle: you will not multiply the amount you risk by a very large factor.

See my 2001 book with Vovk and the subsequent papers posted at www.probabilityandfinance.com.

# Mathematical Results

1. Classical limit theorems (law of large numbers, central limit theorem, etc.) follow from applying Cournot's principle to simple strategies.

2. Abstract measure-theoretic probability can be obtained from assumptions similar to the those used in the theory of imprecise probabilities.

# p-values

Simple gambling strategies can be used to adjust p-values to account for the fact that they fall short of tests with fixed significance levels.

Rule of thumb: $3\sqrt{p}$.

# Defensive Forecasting:

Represent statistical tests by betting strategies.

By playing against these strategies, a forecaster who is given feedback can pass the tests.

I conclude that adaptive or non-stationary forecasting is possible but that it teaches us little about the world.

# Stochastic Calculus as an Emergent Phenomenon

The assumption that a speculator will not multiply capital risked by a large factor relative to a market index using certain simple strategies implies that that the paths of security prices will look like (possibly time-distorted) geometric Brownian motion.

Moroever, the market index will appreciate in proportion to its accumulated variance.

This resolves the equity premium puzzle and explains why equity performance is related to apparent risk without making assumptions about investors' probabilities and utilities.