Fourth Bayesian, Fiducial, and Frequentist Conference

Department of Statistics, Harvard University, May 1, 2017

Discussion of Dempster by Shafer

(Glenn Shafer at Rutgers, www.glennshafer.com)

Dempster-Shafer is fiducial…

…and so are you.

Fiducial principle: To use a probability, we must make the judgement that other information is irrelevant.

In a nutshell: In use, all probability is fiducial.

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 is irrelevant.

(Allow me to deny that I have already integrated all the evidence and can find the resulting probabilities by examining my own dispositions to act.)

Who was fiducial?

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

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

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'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.

Art can probably go along with my way of stating the fiducial principle.

But I have more sympathy than Art does with frequentism, which I call Cournotian testing and Bernoullian estimation.

We need two more principles to bring frequentism into the Bayes/Fiducial/D-S tent.

Three principles to unify BFF

1. Fiducial principle.

2. Poisson's principle.

3. Cournot's principle.

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.

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, …”

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,

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

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