What does “frequentist” mean?

Glenn Shafer

Workshop on Fusion Learning, BFF inferences and Statistical Foundations

April 11, 2016

Department of Statistics

Rutgers University – New Brunswick

Term invented by philosopher Ernest Nagel in 1936.
First used in statistics by Maurice Kendall in 1949.
Embraced by Jerzy Neyman and other non-Bayesians in the 1960s.
Philosophers and statisticians use it differently.

FREQUENTISM

=

frequency interpretation of probability

Part 1. Philosophers

- Probability = relative frequency
- Distinct from propensity interpretation
- Naïve

Part 2. Statisticians

- **SOME** probabilities are relative frequencies.
- Probability models must be testable by observations.
- Test using Cournot's principle.

Part 3. Generalizing to game-theoretic probability

- The game-theoretic Cournot principle
- Implications for convergence of frequencies
- Classical Cournot principle as special case

Invention of the words

Frequentist and *frequentism* not widely used until 1960s.

So far as I know, *frequentist* was
- first used by philosopher Ernest Nagel 1936,
- then by statistician Maurice Kendall 1949.

The philosopher Donald Williams used *frequentism* in 1945.

Ernest Nagel, The Meaning of Probability, *Journal of the American Statistical Association* 31(193):10-30, 1936.
(See also Ernest Nagel, *Principles of the Theory of Probability*, University of Chicago Press, 1939.)
Maurice G. Kendall, On the Reconciliation of Theories of Probability, *Biometrika* 36(1/2):101-116, 1949.
Donald Williams, The Challenging Situation in the Philosophy of Probability, *Philosophy and Phenomenological Research* 6(1):67-86.

FREQUENTISM

Part I. What it means to philosophers

- Probability = relative frequency
- Distinct from propensity interpretation
- Naïve

Ernest Nagel
1901-1985

Emigrated to US from Slovakia at age 10.
Spent career at Columbia.

Coined *frequentist*.

Oxford English Dictionary

Frequentist: One who believes that the probability of an event should be defined as the limit of its relative frequency in a large number of trials.

This definition was given by Nagel in 1936 and has been used by philosophers since then.

Nagel anticipated by John Venn (1834-1923).

…we may define the probability … of the event happening that particular way as the numerical fraction which represents the proportion between the two different classes in the long run.

1886 edition of *The Logic of Chance* (page 163)

Nagel listed three interpretations of probability:

1. Belief (e.g., De Morgan)
2. Logical (e.g., Keynes)
3. Frequency

Savage (1954):

1. Personalistic
2. Necessary
3. Objectivistic

In mathematical statistics in 1954, “objectivistic” views were dominant, but “frequentist” was not yet widely used.

Nagel on the frequency interpretation:

…already implicit in Aristotle, but has become prominent only within the last century as a consequence of applying the probability calculus to statistics and physics. Its central idea is that by the probability of a proposition or an "event" is meant the relative frequency of the "event" in an indefinite class of events.

…on this interpretation every statement involving probabilities is a material proposition whose truth or falsity is to be discovered by examining objective relative frequencies.

In the *Stanford Encyclopedia of Philosophy*, Alan Hájek lists six interpretations of probability:

1. Classical (Bernoulli, Laplace)
2. Logical (Johnson, Keynes, Jeffreys, Carnap)
3. Subjective (Ramsey, de Finetti,
4. Frequency (Venn, Reichenbach, von Mises)
5. Propensity (Peirce, Popper)
6. Best-system (David Lewis)

In 2005, Maria Carla Galavotti listed the first 5.

In 2009, Hájek gave 15 arguments against frequentism.

Philosophical Introduction to Probability, by Maria Carla Galavotti, 2005.

"Mises Redux" -- Redux: Fifteen Arguments against Finite Frequentism, by Alan Hájek, Erkenntnis 45(2/3):209-227, 1996

Philosophers consider frequentism hopelessly naïve.

Many statisticians call themselves frequentists.

Hypothesis: The two groups don't really understand the word in the same way.

FREQUENTISM

Part 2. What it means to statisticians

- **SOME** probabilities are relative frequencies.
- Probability models must be testable by observations.
- Test using Cournot's principle.

Maurice G. Kendall
1907-1983
British statistician

Apparently first statistician to use *frequentist*.

In 1949, he sought reconciliation with subjectivists.

On the Reconciliation of Theories of Probability,
"*Biometrika*, **36**:104, 1949.

Statisticians' problem with frequentism as defined by Nagel:

Bernoulli's Theorem: In a large number of independent trials of an event with probability p, Probability(relative frequency ≈ p) ≈ 1.

There are two probabilities in the theorem. Are they both frequencies?

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

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

<u>Borel's Strong Law of Large Numbers (1909):</u> In an infinite sequence of independent trials of an event with probability p,

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

Each statement involves two probabilities.

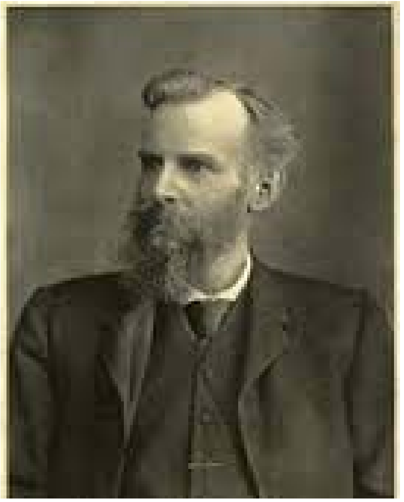
1. If <u>probability=relative frequency</u> by definition, why do you need the theorem?

2. How do you make sense of the second probability? (Infinite regress?)

John Venn cleared the deck by rejecting Bernoulli’s theorem.

…if there is any value in the foregoing criticism, the basis on which the mathematics rest is faulty, owing to there being really nothing which we can with propriety call an objective probability.

The Logic of Chance, 1886, page 91



John Venn
1834-1923

Venn's attitude was viable in his time, when mathematical probability had fallen into low repute.

- Even the French (Bertrand, Poincaré) had lost interest in the classical limit theorems (Bernoulli, De Moivre, Poisson).
- Statisticians were exploiting floods of data without probability theory. (*Kollektivmasslehre*)
- The British mostly treated probability as a source of toy problems (geometric probability) or syllogisms.

But objective probability was more than frequency

- in classical probability as revived on the continent (Chebyshev, Markoc, Cantelli, Lévy, Kolmogorov…), and
- in mathematical statistics after it absorbed continental probability (Cramer, Neyman…)

<u>Bernoulli's Theorem (1713):</u> 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 <u>only</u> to connect probability with phenomena.

Antoine Augustin Cournot
1801-1877

. . . 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. [1843, p. 78]

(Kant had distinguished between the noumenon, or thing-in-itself, and the phenomenon, or object of experience.)

From Cournot's principle to market efficiency, by Glenn Shafer. Working Paper #15, 2006, at www.probabilityandfinance.com.

Another way of expressing Cournot’s principle:

content of a probability model

=

its testable predictions

Commonplace among statisticians and probabilists.

Even when they accept the label “frequentist”.

Founders of modern probability and statistics adopted Cournot's principle as the basis of their frequentism.

- Oskar Anderson (founder of LMU statistics)
- Jacques Hadamard (the last universal mathematician)
- Paul Lévy, Emile Borel, Maurice Fréchet (who revived French probability in the 20^{th} century)
- Andrei Kolmogorov (founder of modern probability)
- Jerzy Neyman (founder of classical frequentist statistics)

Alexandr Chuprov
1874-1926

Oskar Anderson
1887-1960

Chuprov's 1911 Russian book on the philosophy of statistics emphasized Cournot's principle as the "bridge" between Bernoulli's mathematics and the conclusion that frequencies will match probabilities.

Anderson, Chuprov's student at Saint Petersburg, had a long career promoting his ideas. His German textbook on mathematical statistics appeared in 1935.

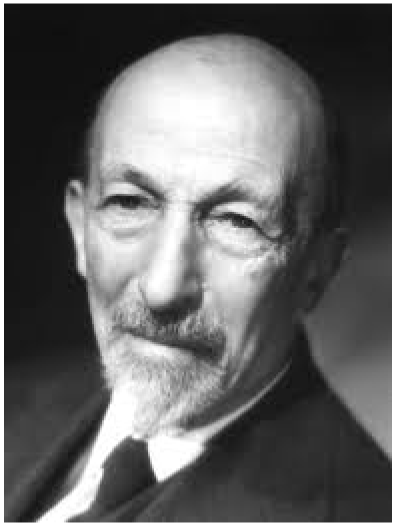
Hans Richter
1912-1978

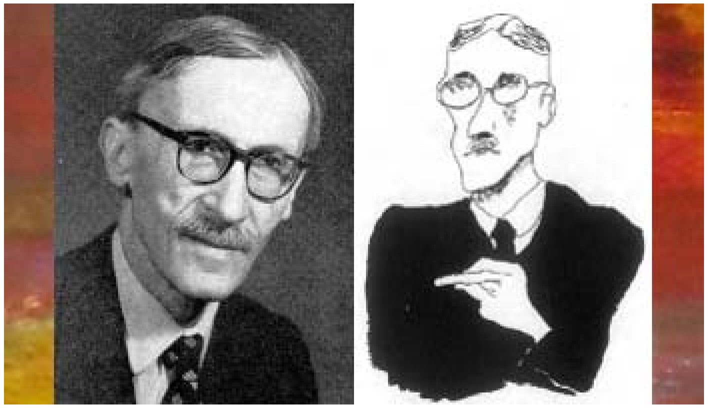
Richter emphasized the *Cournotsche Prinzip* in his 1956 textbook and his contributions to *Dialectica*.

1955 wurde er ordentlicher Professor an der Ludwig-Maximilians-Universität München mit dem neuen Lehrstuhl für Mathematische Statistik und Wirtschaftsmathematik.

Gemeinsam mit Leopold Schmetterer gründete Richter die *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* (seit 1986 *Probability Theory and Related Fields*).

Jacques Hadamard
1865-1963





Paul Lévy
1886-1971

Hadamard's two fundamental notions of probability

1. Equally probable events. The (subjective) basis for the mathematics of probability.

2. Event of very small probability. Only way to provide an objective value to initially subjective probabilities.

1. Evénements également probables
2. Evénement très peu probable

How Lévy said it in 1937

We can only discuss the objective value of the notion of probability when we know the theory's verifiable consequences.

They all flow from this principle: a sufficiently small probability can be neglected.

In other words: an event sufficiently unlikely can be considered practically impossible.

Théorie de l'addition des variables aléatoire (p. 3): Nous ne pouvons discuter la valeur objective de la notion de probabilité que quand nous saurons quelles sont les consequences vérifiables de la théorie. Elles découlent toutes de ce principe: une probabilité suffisamment petite peut être négligéé; en autre termes : *un événement suffisamment peu probable peut être pratiquement considéré comme impossible.*

In his celebrated *Grundbegriffe* (1933), Kolmogorov explained the relation of probability theory to the world of experience.

Andrei Kolmogorov
1903-1987

Under certain conditions … we may assume that an event *A* that does or does not occur under conditions *S* is assigned a real number P(*A*) with the following properties:

A. One can be practically certain that if the system of conditions *S* is repeated a large number of times, n, and the event A occurs m times, then the ratio m/n will differ only slightly from P(*A*).

B. If P(*A*) is very small, then one can be practically certain that the event *A* will not occur on a single realization of the conditions *S*.

The origins and legacy of Kolmogorov's *Grundbegriffe*, by Glenn Shafer. Working Paper #4, 2013, at www.probabilityandfinance.com.

Kolmogorov's students stated Cournot's principle concisely in an encyclopedia article by in the 1980s:

Only probabilities close to zero or one have empirical meaning.

Prokhorov and Sevastyanov. Probability theory. *Soviet Mathematical Encyclopaedia* (English translation) 7:307-313, 1987-1994.

Yuri Vasilyevich Prokhorov (1929-2013)
Boris Aleksandrovich Sevastyanov (1923-2013)

Von Mises, the prototypical frequentist, did not rely on Cournot's principle.

Instead, he sought a mathematical model for randomness, analogous to the Euclidean model for the physical world.

Richard von Mises
1883-1953

Simplest version of von Mises' model

A sequence of 0s and 1s with limiting frequency p of 1s in any subsequence chosen recursively.

In 1934, Wald showed existence of sequences with limiting frequency p for a given countable set of such subsequences.

Abraham Wald
1902-1950

Emile Borel
1871-1956

Maurice Fréchet
1878-1973

Jean Ville
1910-1989
Student of Borel and Fréchet

In 1936, Ville showed that Wald's limiting frequencies are not enough to make a sequence safe for betting.

- Sequence must also obey other probability-one properties.
- For example, it must converge at the law-of-iterated-logarithm rate.

Beginning in the 1940s, Borel called the principle that an event of low probability will not happen the only law of chance.

In the 1950s, Fréchet coined the name Cournot's principle.

Jerzy Neyman (1894-1981)

In the 1930s, Neyman launched "classical" mathematical statistics—i.e., mathematical statistics based on the classical theory of probability.

Whenever we succeed in arranging the technique of a random experiment, say E, such that the relative frequencies of its different results in a long series sufficiently approach, in our opinion, the probabilities calculated from a fundamental probability set (A), we shall say that the set (A) adequately represents the method of carrying out the experiment E. The theory developed below is entirely independent of whether the law of big numbers holds good or not. But the applications of the theory do depend on the assumption that it is valid. (pp. 339-340)

Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability, *Philosophical Transactions of the Royal Society* 236:333-380, 1937.

Jerzy Neyman 1960

In 1960, Neyman declared that the study of stochastic processes, rather than repeated experiments, had become central to science. Yet he still talked about frequencies!

The fourth period in the history of indeterminism, currently in full swing, the period of "dynamic indeterminism," is characterized by the search for evolutionary chance mechanisms capable of explaining the various frequencies observed in the development of the phenomena studied. The chance mechanism of carcinogenesis and the chance mechanism behind the varying properties of the comets in the Solar System exemplify the subjects of dynamic indeterministic studies.

Indeterminism in science and new demands on statisticians, *Journal of the American Statistical Association* 55(292):625-639, 1960.

Jerzy Neyman 1977

In the 1970s, Neyman embraced the word "frequentist" while emphasizing testing.

…the Mendel law specifies only a chance mechanism of inheritance and it is not contended that each n_i must be equal to its expectation. The question is about an intelligible methodology for deciding whether the observed numbers n_i, differing from np_i, contradict the stochastic model of Mendel.

Frequentist Probability and Frequentist Statistics, *Synthese* 36(1):97-131, 1977.

Jerzy Neyman 1977

The idea of frequentist models of natural phenomena seems to be due to Emile Borel. In fact, in his book [13], first published in 1909, Borel identified the construction of stochastic models with the general problem of mathematical statistics:

Le problème général de la statistique mathématique est le suivant. Déterminer un système de tirages effectués dans urnes de composition fixe de telle manière que les résultats d'une série de tirages, interprétés à l'aide de coefficients fixes convenablement choisis, puissent avec une très grande vraisemblance conduire à un tableau identique au tableau des observations.

Where is the philosophical literature on Cournot's principle?

- Discussed by philosophers in French in 1950s.
- Is there anything in English?
- Not mentioned in discussions of lottery paradox.

Is Cournot's principle too continental?

Foreign to British empiricism and American realism?

Tries to relate a *théorie* or *Konzept* to phenomena rather than reasoning directly about the facts.

FREQUENTISM

Part 3. Generalizing to game-theoretic probability

- The game-theoretic Cournot principle
- Implications for convergence of frequencies
- Classical Cournot principle as special case

Aleksandr Khinchin
1894-1959
Russian mathematician

Proved the law of the iterated logarithm in 1922.

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

Game-theoretic Cournot principle

Called "fundamental interpretative hypothesis" in Shafer/Vovk 2001

Meaning of forecasts

=

Skeptic will not multiply capital by large factor.

Forecaster is discredited if Skeptic multiplies capital by large factor.

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

Game-theoretic Cournot principle

Meaning of forecasts

=

Skeptic will not multiply capital by large factor.

Recall that Forecaster and Skeptic may use auxiliary information.

Parameter: Information set $\mathcal{Z}$
  Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
  FOR $n = 1, 2, \ldots$:
    Reality announces $z_n \in \mathcal{Z}$.
    Forecaster announces $p_n \in [0, 1]$.
    Skeptic announces $L_n \in \mathbb{R}$.
    Reality announces $y_n \in \{0, 1\}$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + L_n(y_n - p_n)$.

Parameter: Information set $\mathcal{Z}$
Skeptic announces $\mathcal{K}_0 \in \mathbb{R}$.
FOR $n = 1, 2, \ldots$:
Reality announces $z_n \in \mathcal{Z}$.
Forecaster announces $p_n \in [0, 1]$.
Skeptic announces $L_n \in \mathbb{R}$.
Reality announces $y_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + L_n(y_n - p_n)$.

Game-theoretic Cournot principle

Meaning of forecasts
=
Skeptic will not multiply capital by large factor.

Skeptic has strategies that multiply his capital by a large factor if ...

- The overall long-run average of the y_n does not approximate the overall long-run average of the p_n.
- The overall long-run average of the y_n does not approximate the overall long-run average of the p_n in some subset of rounds identified by the auxiliary information z_n.
- The overall long-run average of the y_n does not tend towards the overall long-run average of the p_n at the rate specified by the law of the iterated logarithm.

Classical Cournot principle

Meaning of probability model

=

Small-probability event selected in advance will not happen.

Model is discredited if event happens.

Game-theoretic Cournot principle

Meaning of forecasts

=

Skeptic will not multiply capital by large factor.

Forecaster is discredited if Skeptic multiplies capital by large factor.

Classical Cournot principle	Game-theoretic Cournot principle
Meaning of probability model	Meaning of forecasts
=	=
Small-probability event selected in advance will not happen.	Skeptic will not multiply capital by large factor.

The classic principle is a special case of the game-theoretic principle.

Classical Cournot principle	Game-theoretic Cournot principle
Meaning of probability model	Meaning of forecasts
=	=
Event of small probability 1/K selected in advance will not happen.	Skeptic will not multiply capital risked by large factor.

Classic principle as special case:

1. Assume each forecast is probability distribution for Reality's next move.
2. Fix a strategy for Forecaster, thus defining a classical probability model for Reality's moves.
3. Fix a strategy for Skeptic (including a stopping time).
4. Fix a factor K by which Skeptic aims to multiply capital.

Under these assumptions, the two principles are equivalent by Markov's inequality.

Points 1 and 2 on previous slide:

1. Assume forecast on each round is probability distribution for Reality's next move.
2. Fix a strategy for Forecaster, thus defining a classical probability model for Reality's moves.

To define a probability distribution for Reality's moves $y_1, y_2, \ldots$, it suffices to give

- probabilities for y_1 and
- probabilities for y_n conditional on $y_1, \ldots, y_{n-1}$.

This is precisely what a strategy for Forecaster supplies if he is required to give a probability distribution for Reality's next move on each round.

Points 3 and 4:

3. Fix a strategy for Skeptic (including a stopping time).
4. Fix a factor K by which Skeptic aims to multiply capital.

- A *variable* X is a function of Forecaster's and Reality's moves.
- $\mathbb{E}(X)$ = stake Skeptic needs to obtain at least X.
- **Markov's inequality:**
  For nonnegative random variable X, $\mathbf{P}\{X \geq K\mathbf{E}(X)\} \leq 1/K$.

- Assume Skeptic begins with capital 1.
- Because Forecaster's moves define a probability distribution for Reality's moves, and Skeptic is following a strategy that does not risk going into the red, Skeptic's payoff X is a nonnegative random variable.
- $\mathbb{E}(X) = 1$, by the game-theoretic definition of expected value.
- The event we reject the model is the event $\{X \geq K\}$.
- By Markov's inequality, $\mathbb{P}\{X \geq K\} \leq 1/K$.

Classical Cournot principle	Game-theoretic Cournot principle
Meaning of probability model	Meaning of forecasts
=	=
Event of small probability 1/K selected in advance will not happen.	Skeptic will not multiply capital risked by large factor.

Classic principle as special case of game-theoretic principle:

1. Assume forecast on each round is probability distribution for Reality's next move.
2. Fix a strategy for Forecaster, thus defining a classical probability model for Reality's moves.
3. Fix a strategy for Skeptic (including a stopping time).
4. Fix a factor K by which Skeptic aims to multiply capital.

In this special case, the two principles are equivalent, because

$$\mathbb{P}\{\text{Skeptic multiplies capital by } K\} \leq 1/K.$$

Classical Cournot principle	Game-theoretic Cournot principle
Meaning of probability model = Event of small probability 1/K selected in advance will not happen.	Meaning of forecasts = Skeptic will not multiply capital risked by large factor.

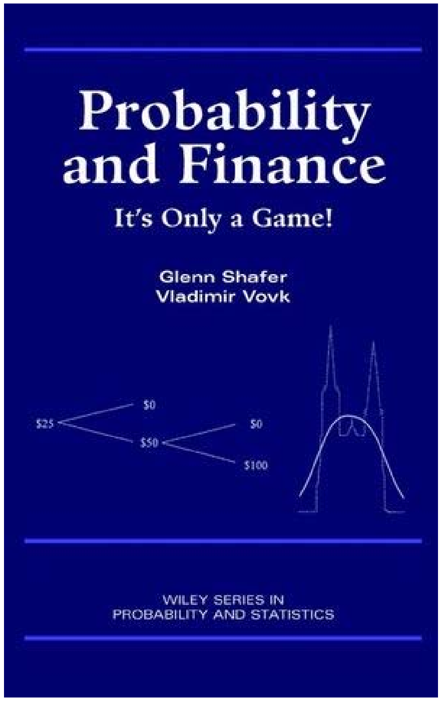
The scope of the generalization:

1. Forecast on each round may fall short of a complete probability distribution for Reality's next move.
2. Forecaster need not follow a strategy.
3. Skeptic need not follow a strategy.
4. Skeptic need not set a goal for multiplying his capital.

But Skeptic must not pretend to have stopped earlier!

More on game-theoretic probability:

- my 2001 book with Vovk
- www.probabilityandfinance.com



Vladimir Vovk
Born 1960
Chervonograd, Ukraine

Part 3. Generalizing to game-theoretic probability