# A History of Evidence in Medical Decisions: From the Diagnostic Sign to Bayesian Inference

*Dennis J. Mazur, MD, PhD*

*Bayesian inference in medical decision making is a concept that has a long history with 3 essential developments: 1) the recognition of the need for data (demonstrable scientific evidence), 2) the development of probability, and 3) the development of inverse probability. Beginning with the demonstrative evidence of the physician's sign, continuing through the development of probability theory based on considerations of games of chance, and ending with the work of Jakob Bernoulli, Laplace, and others, we will examine how Bayesian inference developed.* ***Key words:*** *evidence; games of chance; history of evidence; inference; probability.* ***(Med Decis Making 2012;32:227–231)***

There are many senses of the term *evidence* in the history of medicine other than scientifically gathered (collected) evidence. Historically, evidence was first based on testimony and opinion by those respected for medical expertise.

## SIGN AS EVIDENCE IN MEDICINE

Ian Hacking singles out the physician's sign as the first example of demonstrable scientific evidence.[1] Once the physician's sign was recognized as a form of evidence, numbers could be attached to these signs. Signs recorded by physicians in the same patient (or set of patients) over time supported a claim by a physician that what is going on in a patient at a certain time is related to what is going on in that patient at a later time. In contrast to testimony and opinion, this new form of evidence is described by Hacking as that form of evidence where one thing points to another.[1, pp 36-37] Hacking gives an example of a thing (a precipitate) as evidence for another thing (the state of a person's insides).[1] Hacking argues that when humans began to examine things pointing beyond themselves to other things, we are examining a new type of evidence to support the truth or falsehood of a scientific proposition independent of testimony or opinion. This form of evidence can also be used to challenge the data themselves in supporting the truth or falsehood of a proposition.

In this treatment of the evidence contained in the physician's diagnostic sign, Hacking is examining evidence as a phenomenon of the ''low sciences'' of the Renaissance, which included alchemy, astrology, geology, and medicine.[1] The low sciences were contrasted with the ''high sciences'' of the Renaissance, namely optics, astronomy, and mechanics, which are characterized by signs with measurable precisions. Whereas the low sciences treated evidence in terms of opinion and testimony, the high sciences according to Hacking could scorn opinion as a form of evidence and instead begin to rely on demonstration. Hacking argues that only ''diagnosis'' places the physician on a different plane of evidence than that of other practitioners of the low sciences.

## FORWARD PROBABILITY IN THE HIGH SCIENCES

The physician's diagnostic sign, confirmed over time and enumerated in terms of successful v. unsuccessful observations, provides a logic of numbers. However, a logic of numbers alone could not be

Received 06 May 2011 from Department of Veterans Administration Medical Center, Oregon Health and Science University, Portland, OR (DJM). Revision accepted for publication 03 November 2011.

Address correspondence to Dennis J. Mazur, Medical Service (P3-MED), Oregon Health and Science University, 3710 S.W. U.S. Veterans Hospital Road, Portland, OR 97201; telephone: (503) 220-8262; e-mail: dennis.mazur@med.va.gov.

DOI: 10.1177/0272989X11434739

used to explain the phenomena of the high sciences, that is, explain the phenomenon of optics, astronomy, or mechanics. Instead, the high sciences needed a new form of mathematics more common to the gambler than to the scientist, that is, a logic of probability. Hacking argues that this logic of probability was based on games of chance such as the flips of a fair coin recorded over time and on an individual's willingness to pay a sum to play such a game of chance.

Bernstein notes that the Chevalier de Méré was known to bet repeatedly on games of chance with just a narrow margin in his favor.[2] Bernstein alludes to Pascal's recognition that de Méré knew that the probability of throwing a 6 with one die rises above 50% with 4 throws to 51.77469136% and that de Méré was a successful gambler winning small amounts over a large number of throws of a die.[2] However, at one point the Chevalier changed his game and bet he would roll a total of 12, or a double 6 (also known as the *sonnez*[3]) on 24 rolls of 2 dice. Soon he realized that the new game was a losing proposition, and Leibniz notes that the Chevalier asked Blaise Pascal why his new game was not as profitable as his original. Pascal, intrigued by de Méré's question, passed the question to Pierre de Fermat, resulting in an exchange of letters. Pascal worked through the problem and found that the probability of winning the new game was only 49.1% compared with 51.8% for the old game. Shafer credits Pascal and Fermat with founding mathematical probability based on their work on fair odds in games of chance and solving the problem of points (the problem of equitably dividing the stakes when a fair game is halted before either player has enough points to win).[4] Shafer considers Christiaan Huygens's *De Ratiociniis in Aleae Ludo (On the Calculations in Games of Chance)* in 1657, possibly the first written text on probability, with popularizing the main ideas of Pascal and Fermat.

But probability derivations based on considerations of games of chance using fair flips of fair coins alone could not found the basis of the use of data in the sciences based on prior evidence. Games of chance could yield insights on expected utility[5] but could not be used to infer odds of future occurrences based on past probabilities where there were no assumptions about countable outcomes—known and specifiable before an event occurs. For example, in dice and cards, all countable outcomes are specifiable before any game with fair pieces (dice or cards) is played fairly.

Shafer notes that the term *probability* was not used in discussions related to games of chance in the letters and works of Pascal, Fermat, and Huygens.[6] Their focus instead was on the ''the number between zero and one'' that was the proportion of the stakes due a player in a game of chance. Pascal, Fermat, and Huygens were interested in predicting from a set of past observations of fair events (what Huygens referred to as *fair lays*) what will happen in the future. However, if the events being considered are not countable and are not fair events, then one can no longer base decision making on their approach to the computation of chances. If this is forward probability, one may ask what is backward (inverse, reverse) probability and where do the origins of that concept lie?

## BACKWARD PROBABILITY

Backward (inverse, reverse) probability has a history in the 1800s that involves 3 key figures: Jakob Bernoulli, Thomas Bayes, and Pierre Simon Laplace. It is a time when the issues of original ideas (who was first to develop a concept, who was first to apply the concept, and attribution of ideas to originators) are less than straightforward.

### Jakob Bernoulli (1654–1705)

Shafer credits Jakob Bernoulli's theorem and his argument that it is morally certain that the frequency of an event in a large number of trials will approximate its probability as a notion closer to what we consider today as probability. For Bernoulli, ''the more observations that are taken, the less the danger will be of deviating from the truth.''[7(p257)] (See Hald's discussion of Jakob Bernoulli's theorem.[7]) Shafer sees the work of Bernoulli and De Moivre as a precursor to the modern theory of confidence intervals.[6]

Jakob Bernoulli included Huygen's book on the computation of chances in fair lays within his book, *Ars Conjectandi* (1713).[8] Within *Ars Conjectandi*, Jakob Bernoulli makes points that can be interpreted as an early position that the computation of chances can be ''inverted'' and then ''suggests'' that this process can be applied to make a guess. For Bernoulli, this guess would be about the probability of an event occurring given a set of observations about previous outcomes that need not be considered as necessarily countable or fair events. Dale notes that in *Ars Conjectandi* there may be an intent to apply probability theory in an inverse manner. However, Dale can find no application of such a result in *Ars Conjectandi*.[9] Stigler[10, p.10] argues that the ''chief conceptual step'' taken in the 18th century toward the application of probability to quantitative inference involved ''the inversion of the probabilities'' analyses of Jacob Bernoulli and de Moivre. Stigler also argues that what was needed at this time was a step away from consideration of games of chance involving coins and dice where the benefits of inverse inference were not obvious.

### Thomas Bayes (born in 1701 or 1702–1761)

Bayes defined probability as ''the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening.''[11] In the Bayesian view, a probability can be assigned to a hypothesis before that hypothesis is tested. Bayesian probabilities can be about either the objective state of knowledge in the world (objective Bayesian view) or a personal belief about the world (subjective Bayesian probabilities).

Historically, what we refer to as ''Bayesian inference'' has been difficult to trace back to the work of Reverend Bayes. Stigler notes that Bayes, an ordained nonconformist minister in Turnbridge Wells, England, never had any of his works published during his lifetime. Bayes's *An Essay Toward Solving a Problem in the Doctrine of Chances* was published posthumously.[10] Stigler notes that within this work it is hard to identify the prior works from which Bayes developed his theory. The paper itself was first read to the Royal Society 2 years after Bayes's death by Richard Price, a Presbyterian minister. Stigler points out that the paper's introduction, although written by Bayes, was added to and possibly amended by Price in such a fashion that it is impossible to separate the contributions of Bayes from those of Price in this introduction. As to why Bayes pursued the work, Stigler argues that several theories can be proposed. The first theory argues that the work itself was an attempt on the part of Bayes to provide a mathematical proof of the existence of a First Cause. A second theory, Stigler argues, is found in Price's contribution to the introduction to Bayes's essay. In his contribution, Price places the context of Bayes's writing in relation to the work of Thomas Simpson in his 1740 book, *The Nature and Laws of Chance*,[12] Simpson's letter of 1755 republished in 1757 (regarding Simpson's concept of the continuous error distribution), and Abraham De Moivre's *The Doctrine of Chance* (1718).[13]

De Moivre and Simpson were itinerant lecturers in the London coffee houses known as *penny universities* because patrons were charged an entrance fee of 1 penny to hear the lecture. In probability, Simpson is known for his work on proving that the arithmetic mean of a number of observations is of more analytical value than a single one.[14] De Moivre first published *The Doctrine of Chance: A Method of Calculating the Probabilities of Events in Play* in 1718.[13]

Again looking at Bayes's work, McGrayne describes Bayes's essay as one on the probability of causes, moving from observations about the real world back to their most probable cause.[15] Yet others share Stigler's hesitation to attribute the work to Bayes. Jaynes and Bretthorst argue that what is usually called ''Bayes's theorem,'' that is, the product rule of probability, is something that Bayes himself never wrote.[16] They also note that the product rule of probability had been recognized by others including Jakob Bernoulli and de Moivre before Bayes. However, Bayes may well lay claim to the early explicit attention to prior probabilities.

Fienberg[17] argues that the phrase ''Bayesian inference'' came about during the middle of the 20th century to describe what had up to that period of time been referred to as inverse probability. Inverse probability is the probability distribution of an unobserved variable. The probability is an ''inverse'' because it involves inferring backward from data (effects) to the parameter (causes).

### Pierre-Simon Laplace (1749–1827)

Laplace defined probability as ''the theory of chance [that] consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.''[18, pp. 6-7] In a Laplacean view, probabilities can be assigned in the presence of symmetrically balanced evidence. And the classical Laplacean definition of probability is confined to finite sample spaces but can be enhanced via entropy arguments to apply to infinite spaces.

McGrayne argues that ''Laplace had owned Bayes's rule in all but name since 1781.''[15(p32)] And Laplace applied it in demography, judicial reform, and astronomy. A considerable part of Laplace's contributions to probability theory are based on inverse probabilities. By this method, an a posteriori probability of a certain hypothesis could be calculated from the results of random experiments, usually under the tacit assumption of an a priori equiprobability of all possible hypotheses (see Laplace's definition of probability above). Focusing on the early work of Laplace on the application of probability to inference through Laplace's *Memoir on the Probability of the Causes of Events*[19] and his *Memoir on Probabilities*,[20] Stigler argues that both works can be taken together as ''the most influential eighteenth-century work on the use of probability in inference.''[10(p100)] In the case of Thomas Simpson and Laplace, it was their focus on errors that led to the recognition that one distribution, that is, the distribution of errors, could provide the necessary element for both ''forward'' and ''inverse'' probability statements. But from what real-world examples did Laplace's conceptions of inverse probability originate? Stigler argues that Laplace was able to conceive of inverse probability by focusing not on observations but on errors of observation. But where does the focus on the errors of observation arise?

## INVERSE PROBABILITY IN THE HIGH SCIENCES

McGrayne[15] argues that after reading de Moivre's *Doctrine of Chances* (1756), Laplace became convinced that probability might help him deal with uncertainties in the solar system. Writing in 1906 in his fourth edition of *Probability and Theory of Errors*, Woodward[21] notes that nothing is more satisfactory in the history of science than the success with which the unique method of ''least squares'' has been applied to the problems presented by the earth and the other members of the solar system. Howie[22] argues that astronomical observations, rather than games of chance, were the grounds of derivation of concepts related to inverse inference. Howie[22] notes that Laplace was able to ''invert'' a famous problem in astronomy to find that some cause, rather than chance alone, had resulted in the orbits of the planets being so nearly coplanar. Howie argues that the concept of inversion is easier to grasp in an astronomical framework because an observation of a star has 2 components: the star's true position and the error component of measuring that star's true position. Thus, when given a set of past observations with their error components, one can produce directly a probability distribution for the unknown position of a star. According to Cowles,[23] Laplace assumed that measurements of stars within our galaxy will not always produce the same results, but the assumption is made that a true value exists. In this context, all variations in measurements away from that true value are errors.

Shafer[6] argues that Laplace discovered inverse probabilities in the course of his work on the theory of errors. Shafer[6] argues that Laplace realized that probabilities for errors, once the observations are fixed, translate into probabilities for the unknown true values being estimated. Laplace called these unknown true values *posterior probabilities*, and he justified using them by adopting the principle that after an observation, the probabilities of its possible causes are proportional to the probabilities given the observation of the causes. Shafer credits Laplace's *Théorie Analytique des Probabilities* (1812, 1814, and 1820)[24] for developing techniques for evaluating posterior probabilities. Daston[25] notes that Laplace's preference for a ''physical or statistical'' (objective) interpretation of probabilities (as contrasted with a subjective interpretation) had a lasting effect on the subsequent development of inverse probabilities as a form of evidence based on past observations in the sciences.

Shafer notes that Laplace's views dominated probability for a generation, but in time they gave way to a different understanding of probability. The error distributions at the base of Laplace's approach seemed to the empiricists of the late 19th and early 20th centuries to have a frequency interpretation, and in this view frequency is argued to be the real foundation of probability.

## THE CONCEPTUAL FRAMEWORK OF PROBABILITY TODAY

The modern conceptual framework of probability as the weighing of evidence and opinion is based on 2 distinct views, that is, the frequentist (objective, physical) view and the subjective (Bayesian) view. *Frequency probabilities* are probabilities of countable events. Pascal, Fermat, and Huygens were interested in countable events in games of chance. Today, physicians and medical decision makers are interested in countable events related to patients in *N*-of-1 clinical trials or human study participants in research-based clinical trials. Countable instances here include the number of individuals who survive after treatment versus the numbers of individuals who die after that same treatment. If all individuals survive a particular therapy, then interest will be on the number of individuals still alive at 1 year after treatment, at 2 years after treatment, and so on. All of these events are assumed to occur at persistent, relative rates if trials are conducted with large enough numbers of participants and if the trials are run long enough. *Subjective probabilities* allow reasoning about uncertain statements. Subjective probability interpretations allow prediction of future events even when the circumstances are essentially uncountable.

Both the frequentist and the subjective views affect inverse probability in distinct ways. The frequentist interpretation only permits use of the inverse probability equation in the constrained circumstance where objects or events are countable (as are balls in an urn); the subjective interpretation allows prediction of future events even when the objects or events are uncountable (such as changes in the weather). Thus, inverse probability is not confined to one interpretation of probability.

## CONCLUSION

Bayesian concepts in medical decision making today have benefited from a long history of development in medicine, mathematics, and astronomy. The concept of evidence in medicine began in the demonstrable diagnostic sign of the physician, but soon it began to be considered in reference to the theory and mathematics of inverse probability, affording a system for inferring unobserved causes from probabilities of observed events. It is doubtful that Bayes recognized the import of his thinking that led to the notion of Bayesian inference as we understand the concept today in medical decision making.

## REFERENCES

1. Hacking I. The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference. New York: Cambridge University Press; 1991.

2. Bernstein PL. Against the Gods: The Remarkable Story of Risk. New York: Wiley; 1996.

3. Edgeworth FY. Probability. In Chisholm H, ed. The Encyclopedia Britannica: A Dictionary of Arts, Sciences. Vol 22, 11th ed. Cambridge, UK: Cambridge University Press; 1911. p 376–403.

4. Shafer G. The early development of mathematical probability. In Grattan-Guinness I, ed. The Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences. New York: Routledge; 1993. p 1293–1302.

5. Bernoulli D. (1738). Exposition of a New Theory on the Measurement of Risk. Translated from the original Latin to English by Louise Sommer. Econometrica. 1954;22:23–36.

6. Shafer G. The early development of mathematical probability. Available from: URL: http://www.glennshafer.com/assets/downloads/articles/article50.pdf

7. Hald A. A History of Probability and Statistics and Their Applications Before 1750. Hoboken, NJ: John Wiley; 2003.

8. Bernoulli J. Ars Contendandi. Opus Posthumum. Bruxelles, Belguim: Culture e Civilization; 1968.

9. Dale AI. A History of Inverse Probability From Thomas Bayes to Karl Pearson. (2nd ed). New York: Springer-Verlag, 1999, p. 7.

10. Stigler SM. The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Harvard University Press; 1986.

11. An essay towards solving a problem in the doctrine of chances by the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton MA and FRS. Read December 23, 1763. First publication. Philos Trans R Soc Lond. 1764;53:370–418. http://www.stat.ucla.edu/history/essay.pdf, p. 4.

12. Simpson T. The Nature and Laws of Chance. London; London: Printed by Edward Cave, St. John's Gate. 1740.

13. de Moivre A. The Doctrine of Chances: Or, a Method of Calculating the Probabilities of Events in Play. 3rd ed. London: Printed for A. Millar, in the Strand; 1756.

14. O'Connor JJ, Robertson EF. Thomas Simpson. MacTutor History of Mathematics. JOC/EFR, 1996. Available from: URL: http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Simpson.html

15. McGrayne SB. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian submarines, & Emerged Triumphant from Two Centuries of Controversy. New Haven, CT: Yale University Press; 2011.

16. Jaynes ET, Bretthorst GL. Probability Theory: The Logic of Science. New York: Cambridge University Press; 2007.

17. Fienberg SE. When did Bayesian inference become Bayesian? Bayesian Anal. 2006;1:1–40.

18. Laplace PS. A Philosophical Essay on Probabilities. Translated from the 6th French edition by Frederick Wilson Truscott and Frederick Lincoln Emory, New York: Dover Publications; 1814; English edition 1951.

19. Laplace PS. Memoir on the Probability of the Causes of Events. Translated from the original French to English by S.M. Stigler. Stat Sci. 1986;1:359–63.

20. Laplace PS. Mémoire sur les Probabilités. Mémoires de l'Académie Royale des Sciences de Paris, 1778, pp. 227-332.

21. Woodward RS. Probability and Theory of Errors. 4th ed. New York: Wiley & Sons; 1906.

22. Howie D. Interpreting Probability: Controversies and Developments in the Early Twentieth Century. New York: Cambridge University Press; 2002.

23. Cowles M. Statistics in Psychology: An Historical Perspective. Mahwah, NJ: Lawrence Erlbaum; 2001.

24. Laplace PS. Théorie Analytique des Probabilités; par M. le comte de Laplace… [With:] Supplement [Premier-Deuxième-Troisième]. Paris: Mme Ve Courcier; 1812 [–1820].

25. Daston L. Classical Probability in the Enlightenment. Princeton, NJ: Princeton University Press; 1988.