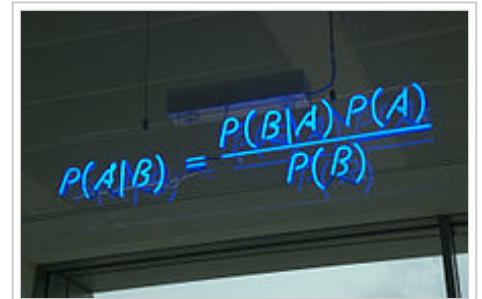


Bayes' theorem

From Wikipedia, the free encyclopedia

In probability theory and statistics, **Bayes' theorem** (alternatively **Bayes' law** or Bayes' rule) describes the probability of an event, based on conditions that might be related to the event. For example, suppose one is interested in whether a woman has cancer, and knows that she is 65. If cancer is related to age, information about her age can be used to more accurately assess the probability of her having cancer using Bayes' Theorem.

When applied, the probabilities involved in Bayes' theorem may have different probability interpretations. In one of these interpretations, the theorem is used directly as part of a particular approach to statistical inference. With the Bayesian probability interpretation the theorem expresses how a subjective degree of belief should rationally change to account for evidence: this is Bayesian inference, which is fundamental to Bayesian statistics. However, Bayes' theorem has applications in a wide range of calculations involving probabilities, not just in Bayesian inference.



A blue neon sign, showing the simple statement of Bayes' theorem

Bayes' theorem is named after Rev. Thomas Bayes (/ˈbeɪz/; 1701–1761), who first provided an equation that allows new evidence to update beliefs. It was further developed by Pierre-Simon Laplace, who first published the modern formulation in his 1812 *Théorie analytique des probabilités*. Sir Harold Jeffreys put Bayes' algorithm and Laplace's formulation on an axiomatic basis. Jeffreys wrote that Bayes' theorem "is to the theory of probability what the Pythagorean theorem is to geometry".^[1]

Contents

- - 1 Statement of theorem
- - 2 Examples
 - - 2.1 Cancer at age 65
 - - 2.2 Drug testing
 - - 2.3 A more complex example
- - 3 Interpretations
 - - 3.1 Bayesian interpretation
 - - 3.2 Frequentist interpretation
 - - 3.2.1 Example
 - - 4 Forms
 -

4.1 Events

- - 4.1.1 Simple form
 - - 4.1.2 Alternative form
 - - 4.1.3 Extended form

4.2 Random variables

- - 4.2.1 Simple form
 - - 4.2.2 Extended form

4.3 Bayes' rule

5 Derivation

- - 5.1 For events
 - - 5.2 For random variables

6 History

7 See also

8 Notes

9 Further reading

10 External links

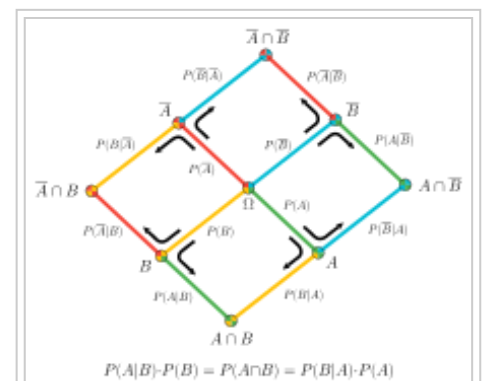
Statement of theorem

Bayes' theorem is stated mathematically as the following equation:^[2]

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)},$$

where A and B are events.

- $P(A)$ and $P(B)$ are the probabilities of A and B without regard to each other.
- $P(A | B)$, a conditional probability, is the probability of observing event A given that B is true.
- $P(B | A)$ is the probability of observing event B given that A is true.



Visualization of Bayes' theorem by the superposition of two decision trees

Examples

Cancer at age 65

Suppose we want to know a person's probability of having cancer, but we know nothing about him or her. Despite not knowing anything about that person, a probability can be assigned based on the general prevalence of cancer. For the sake of this example, suppose it is 1%. This is known as the base rate or prior probability of having cancer. "Prior" refers to the time before being informed about the particular case at hand.

Next, suppose we find out that person is 65 years old. If we assume that cancer and age are related, this new piece of information can be used to better assess that person's risk of having cancer. More precisely, we'd like to know the probability that a person has cancer when it is known that he or she is 65 years old. This quantity is known as the current probability, where "current" refers to upon finding out information about the particular case at hand.

In order to apply knowledge of that person's age in conjunction with Bayes' Theorem, two additional pieces of information are needed. Note, however, that the additional information is not specific to that person. The needed information is as follows:

1. The probability of being 65 years old. Suppose it is 0.2%
2. The probability that a person *with cancer* is 65 years old. Suppose it is 0.5%. Note that this is greater than the previous value. This reflects that people with cancer are disproportionately 65 years old.

Knowing this, along with the base rate, we can calculate that a person who is age 65 has a probability of having cancer equal to

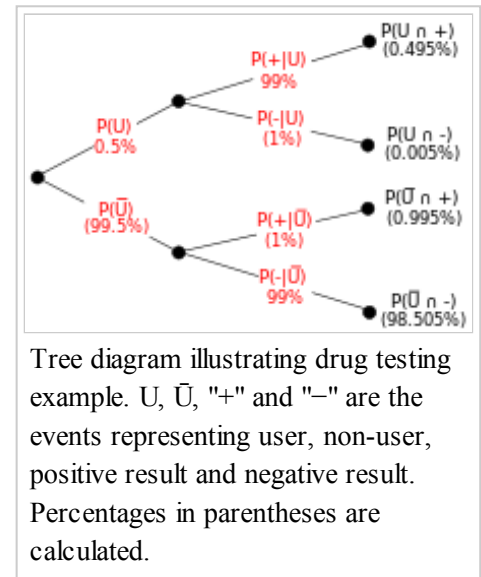
$$(0.5\% * 1\%) \div 0.2\% = 2.5\%$$

It may come as a surprise that even though being 65 years old increases the risk of having cancer, that person's probability of having cancer is still fairly low. This is because the base rate of cancer (regardless of age) is low. This illustrates both the importance of base rate, as well as that it is commonly neglected.^[3] Base rate neglect leads to serious misinterpretation of statistics; therefore, special care should be taken to avoid such mistakes. Becoming familiar with Bayes' theorem is one way to combat the natural tendency to neglect base rates.

Problems using Bayes' Theorem are often more easily grasped by applying the conditions given in the problem to a large pool of observations. Assume, for example, that a community consists of 100,000 people. According to the statement of the problem, 1% of the population, or 1,000 people will have cancer. 0.2% of the population, or 200 people, will be 65 years old. Of the 1000 people with cancer, only .5%, or 5 people, will be 65 years old. Thus, of the 200 people who are 65, only 5 can be expected to have cancer. $5/200 = 2.5\%$.

Drug testing

Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?



$$\begin{aligned}
 P(\text{User} \mid +) &= \frac{P(+ \mid \text{User})P(\text{User})}{P(+ \mid \text{User})P(\text{User}) + P(+ \mid \text{Non-user})P(\text{Non-user})} \\
 &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\
 &\approx 33.2\%
 \end{aligned}$$

Despite the apparent accuracy of the test, if an individual tests positive, it is more likely that they do *not* use the drug than that they do. This again illustrates the importance of base rates, and how the formation of policy can be egregiously misguided if base rates are neglected.

This surprising result arises because the number of non-users is very large compared to the number of users; thus the number of false positives (0.995%) outweighs the number of true positives (0.495%). To use concrete numbers, if 1000 individuals are tested, there are expected to be 995 non-users and 5 users. From the 995 non-users, $0.01 \times 995 \approx 10$ false positives are expected. From the 5 users, $0.99 \times 5 \approx 5$ true positives are expected. Out of 15 positive results, only 5, about 33%, are genuine.

Note: The importance of specificity can be illustrated by showing that even if sensitivity is 100% and specificity is at 99% the probability of the person being a drug user is $\approx 33\%$ but if the specificity is changed to 99.5% and the sensitivity is dropped down to 99% the probability of the person being a drug user rises to 49.8%.

A more complex example

The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the output, respectively. The fraction of defective items produced is this: for the first machine, 5%; for the second machine, 3%; for the third machine, 1%. If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?

A solution is as follows. Let A_i denote the event that a randomly chosen item was made by the i th machine (for $i = 1, 2, 3$). Let B denote the event that a randomly chosen item is defective. Then, we are given the following information:

$$P(A_1) = 0.2, \quad P(A_2) = 0.3, \quad P(A_3) = 0.5.$$

If the item was made by machine A_1 , then the probability that it is defective is 0.05; that is, $P(B | A_1) = 0.05$. Overall, we have

$$P(B | A_1) = 0.05, \quad P(B | A_2) = 0.03, \quad P(B | A_3) = 0.01.$$

To answer the original question, we first find $P(B)$. That can be done in the following way:

$$P(B) = \sum_i P(B | A_i) P(A_i) = (0.05)(0.2) + (0.03)(0.3) + (0.01)(0.5) = 0.024.$$

Hence 2.4% of the total output of the factory is defective.

We are given that B has occurred, and we want to calculate the conditional probability of A_3 . By Bayes' theorem,

$$P(A_3 | B) = P(B | A_3) P(A_3) / P(B) = (0.01)(0.50) / (0.024) = 5/24.$$

Given that the item is defective, the probability that it was made by the third machine is only 5/24. Although machine 3 produces half of the total output, it produces a much smaller fraction of the defective items. Hence the knowledge that the item selected was defective enables us to replace the prior probability $P(A_3) = 1/2$ by the smaller posterior probability $P(A_3 | B) = 5/24$.

Once again, the answer can be reached without recourse to the formula by applying the conditions to any hypothetical number of cases. For example, in 100,000 items produced by the factory, 20,000 will be produced by Machine A, 30,000 by Machine B, and 50,000 by Machine C. Machine A will produce 1000 defective items, Machine B 900, and Machine C 500. Of the total 2400 defective items, only 500, or 5/24 were produced by Machine C.

Interpretations

The interpretation of Bayes' theorem depends on the interpretation of probability ascribed to the terms. The two main interpretations are described below.

Bayesian interpretation

In the Bayesian (or epistemological) interpretation, probability measures a *degree of belief*. Bayes' theorem then links the degree of belief in a proposition before and after accounting for evidence. For example, suppose it is believed with 50% certainty that a coin is twice as likely to land heads than tails. If the coin is flipped a number of times and the outcomes observed, that degree of belief may rise, fall or remain the same depending on the results.

For proposition A and evidence B ,

- $P(A)$, the *prior*, is the initial degree of belief in A .
- $P(A | B)$, the *posterior*, is the degree of belief having accounted for B .
- the quotient $\frac{P(B | A)}{P(B)}$ represents the support B provides for A .

For more on the application of Bayes' theorem under the Bayesian interpretation of probability, see Bayesian inference.

Frequentist interpretation

In the frequentist interpretation, probability measures a *proportion of outcomes*. For example, suppose an experiment is performed many times. $P(A)$ is the proportion of outcomes with property A , and $P(B)$ that with property B . $P(B|A)$ is the proportion of outcomes with property B out of outcomes with property A , and $P(A|B)$ the proportion of those with A out of those with B .


The role of Bayes' theorem is best visualized with tree diagrams, as shown to the right. The two diagrams partition the same outcomes by A and B in opposite orders, to obtain the inverse probabilities. Bayes' theorem serves as the link between these different partitionings.


Example


An entomologist spots what might be a rare subspecies of beetle, due to the pattern on its back. In the rare subspecies, 98% have the pattern, or $P(\text{Pattern}|\text{Rare}) = 98\%$. In the common subspecies, 5% have the pattern. The rare subspecies accounts for only 0.1% of the population. How likely is the beetle having the pattern to be rare, or what is $P(\text{Rare}|\text{Pattern})$?

From the extended form of Bayes' theorem (since any beetle can be only rare or common),


Relative size	Case B	Case \bar{B}	Total
Condition A	w	x	w+x
Condition \bar{A}	y	z	y+z
Total	w+y	x+z	w+x+y+z

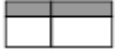


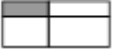




$$P(A|B) \times P(B) = \frac{w}{w+y} \times \frac{w+y}{w+x+y+z} = \frac{w}{w+x+y+z}$$







$$P(B|A) \times P(A) = \frac{w}{w+x} \times \frac{w+x}{w+x+y+z} = \frac{w}{w+x+y+z}$$

A geometric visualisation of Bayes' theorem. In the table, the values w , x , y and z give the relative weights of each corresponding condition and case. The figures denote the cells of the table involved in each metric, the probability being the fraction of each figure that is shaded. This shows that $P(A|B)P(B) = P(B|A)P(A)$ i.e. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. Similar reasoning can be used to show that $P(\bar{A}|B) = \frac{P(B|\bar{A})P(\bar{A})}{P(B)}$ and so forth.

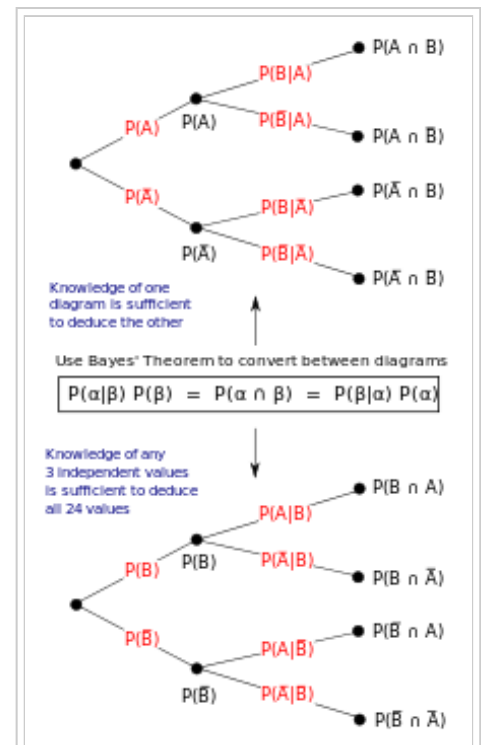
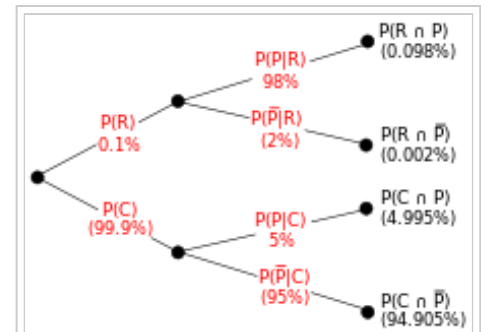


Illustration of frequentist interpretation with tree diagrams. Bayes' theorem connects conditional probabilities to their inverses.



Tree diagram illustrating frequentist example. R, C, P and P bar are the events representing rare, common, pattern and no pattern. Percentages in parentheses are calculated. Note that three independent values are given, so it is possible to calculate the inverse tree (see figure above).

$$\begin{aligned}
 P(\text{Rare} \mid \text{Pattern}) &= \frac{P(\text{Pattern} \mid \text{Rare})P(\text{Rare})}{P(\text{Pattern} \mid \text{Rare})P(\text{Rare}) + P(\text{Pattern} \mid \text{Common})P(\text{Common})} \\
 &= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.05 \times 0.999} \\
 &\approx 1.9\%.
 \end{aligned}$$

Forms

Events

Simple form

For events A and B , provided that $P(B) \neq 0$,

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}.$$

In many applications, for instance in Bayesian inference, the event B is fixed in the discussion, and we wish to consider the impact of its having been observed on our belief in various possible events A . In such a situation the denominator of the last expression, the probability of the given evidence B , is fixed; what we want to vary is A . Bayes' theorem then shows that the posterior probabilities are proportional to the numerator:

$$P(A \mid B) \propto P(A) \cdot P(B \mid A) \text{ (proportionality over } A \text{ for given } B).$$

In words: **posterior is proportional to prior times likelihood**.^[4]

If events A_1, A_2, \dots , are mutually exclusive and exhaustive, i.e., one of them is certain to occur but no two can occur together, and we know their probabilities up to proportionality, then we can determine the proportionality constant by using the fact that their probabilities must add up to one. For instance, for a given event A , the event A itself and its complement $\neg A$ are exclusive and exhaustive. Denoting the constant of proportionality by c we have

$$P(A \mid B) = c \cdot P(A) \cdot P(B \mid A) \text{ and } P(\neg A \mid B) = c \cdot P(\neg A) \cdot P(B \mid \neg A).$$

Adding these two formulas we deduce that

$$c = \frac{1}{P(A) \cdot P(B \mid A) + P(\neg A) \cdot P(B \mid \neg A)}.$$

Alternative form

Another form of Bayes' Theorem that is generally encountered when looking at two competing statements or hypotheses is:

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B \mid A) P(A) + P(B \mid \neg A) P(\neg A)}.$$

For an epistemological interpretation:

For proposition A and evidence or background B ,^[5]

- $P(A)$, the prior probability, is the initial degree of belief in A .
- $P(\neg A)$, is the corresponding probability of the initial degree of belief against A : $1 - P(A) = P(\neg A)$
- $P(B \mid A)$, the conditional probability or likelihood, is the degree of belief in B , given that the proposition A is true.
- $P(B \mid \neg A)$, the conditional probability or likelihood, is the degree of belief in B , given that the proposition A is false.
- $P(A \mid B)$, the posterior probability, is the probability for A after taking into account B for and against A .

Extended form

Often, for some partition $\{A_j\}$ of the sample space, the event space is given or conceptualized in terms of $P(A_j)$ and $P(B \mid A_j)$. It is then useful to compute $P(B)$ using the law of total probability:

$$P(B) = \sum_j P(B \mid A_j) P(A_j),$$

$$\Rightarrow P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{\sum_j P(B \mid A_j) P(A_j)}.$$

In the special case where A is a binary variable:

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B \mid A) P(A) + P(B \mid \neg A) P(\neg A)}.$$

Random variables

Consider a sample space Ω generated by two random variables X and Y . In principle, Bayes' theorem applies to the events $A = \{X = x\}$ and $B = \{Y = y\}$. However, terms become 0 at points where either variable has finite probability density. To remain useful, Bayes' theorem may be formulated in terms of the relevant densities (see Derivation).

Simple form

If X is continuous and Y is discrete,

$$f_X(x | Y = y) = \frac{P(Y = y | X = x) f_X(x)}{P(Y = y)}.$$

If X is discrete and Y is continuous,

$$P(X = x | Y = y) = \frac{f_Y(y | X = x) P(X = x)}{f_Y(y)}.$$

If both X and Y are continuous,

$$f_X(x | Y = y) = \frac{f_Y(y | X = x) f_X(x)}{f_Y(y)}.$$

Extended form

A continuous event space is often conceptualized in terms of the numerator terms. It is then useful to eliminate the denominator using the law of total probability. For $f_Y(y)$, this becomes an integral:

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y | X = \xi) f_X(\xi) d\xi.$$

Bayes' rule

Bayes' rule is Bayes' theorem in odds form.

$$O(A_1 : A_2 | B) = O(A_1 : A_2) \cdot \Lambda(A_1 : A_2 | B)$$

where

$$\Lambda(A_1 : A_2 | B) = \frac{P(B | A_1)}{P(B | A_2)}$$

is called the Bayes factor or likelihood ratio and the odds between two events is simply the ratio of the probabilities of the two events. Thus

$$O(A_1 : A_2) = \frac{P(A_1)}{P(A_2)},$$

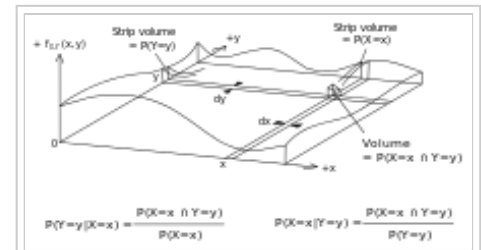


Diagram illustrating the meaning of Bayes' theorem as applied to an event space generated by continuous random variables X and Y . Note that there exists an instance of Bayes' theorem for each point in the domain. In practice, these instances might be parametrized by writing the specified probability densities as a function of x and y .

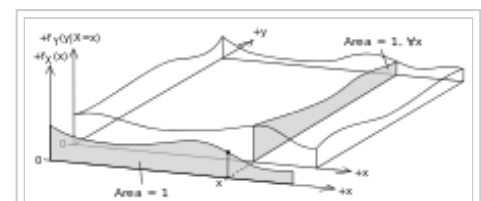


Diagram illustrating how an event space generated by continuous random variables X and Y is often conceptualized.

$$O(A_1 : A_2 | B) = \frac{P(A_1 | B)}{P(A_2 | B)},$$

So the rule says that the posterior odds are the prior odds times the Bayes factor, or in other words, posterior is proportional to prior times likelihood.

Derivation

For events

Bayes' theorem may be derived from the definition of conditional probability:

$$\begin{aligned} P(A | B) &= \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0, \\ P(B | A) &= \frac{P(A \cap B)}{P(A)}, \text{ if } P(A) \neq 0, \\ \Rightarrow P(A \cap B) &= P(A | B) P(B) = P(B | A) P(A), \\ \Rightarrow P(A | B) &= \frac{P(B | A) P(A)}{P(B)}, \text{ if } P(B) \neq 0. \end{aligned}$$

For random variables

For two continuous random variables X and Y , Bayes' theorem may be analogously derived from the definition of conditional density:

$$\begin{aligned} f_X(x | Y = y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ f_Y(y | X = x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\ \Rightarrow f_X(x | Y = y) &= \frac{f_Y(y | X = x) f_X(x)}{f_Y(y)}. \end{aligned}$$

History

Bayes' theorem was named after the Reverend Thomas Bayes (1701–61), who studied how to compute a distribution for the probability parameter of a binomial distribution (in modern terminology). Bayes' unpublished manuscript was significantly edited by Richard Price before it was posthumously read at the Royal Society. Price edited^[6] Bayes' major work *An Essay towards solving a Problem in the Doctrine of Chances* (1763), which appeared in *Philosophical Transactions*,^[7] and contains Bayes' Theorem. Price wrote an introduction to the paper which provides some of the philosophical basis of Bayesian statistics. In 1765 he was elected a Fellow of the Royal Society in recognition of his work on the legacy of Bayes.^{[8][9]}

The French mathematician Pierre-Simon Laplace reproduced and extended Bayes' results in 1774, apparently quite unaware of Bayes' work.^{[10][11]} Stephen Stigler suggested in 1983 that Bayes' theorem was discovered by Nicholas Saunderson some time before Bayes;^[12] that interpretation, however, has been disputed.^[13]

Martyn Hooper^[14] and Sharon McGrayne^[15] have argued that Richard Price's contribution was substantial:

By modern standards, we should refer to the Bayes–Price rule. Price discovered Bayes' work, recognized its importance, corrected it, contributed to the article, and found a use for it. The modern convention of employing Bayes' name alone is unfair but so entrenched that anything else makes little sense.

— [15]

See also

- Bayesian inference
- Inductive probability

Notes

- Jeffreys, Harold (1973). *Scientific Inference* (3rd ed.). Cambridge University Press. p. 31. ISBN 978-0-521-18078-8.
- Stuart, A.; Ord, K. (1994), *Kendall's Advanced Theory of Statistics: Volume I—Distribution Theory*, Edward Arnold, §8.7.
- Daniel Kahneman (25 October 2011). *Thinking, Fast and Slow*. Macmillan. ISBN 978-1-4299-6935-2. Retrieved 8 April 2012.
- Lee, Peter M. (2012). "Chapter 1". *Bayesian Statistics*. Wiley. ISBN 978-1-1183-3257-3.
- "Bayes Theorem: Introduction". *Trinity University*.
- Richard Allen (1999). *David Hartley on Human Nature*. SUNY Press. pp. 243–4. ISBN 978-0-7914-9451-6. Retrieved 16 June 2013.
- Bayes, Thomas, and Price, Richard (1763). "An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S." (PDF). *Philosophical Transactions of the Royal Society of London* **53** (0): 370–418. doi:10.1098/rstl.1763.0053.
- Holland, pp. 46–7.
- Richard Price (1991). *Price: Political Writings*. Cambridge University Press. p. xxiii. ISBN 978-0-521-40969-8. Retrieved 16 June 2013.
- Laplace refined Bayes' theorem over a period of decades:
 - Laplace announced his independent discovery of Bayes' theorem in: Laplace (1774) "Mémoire sur la probabilité des causes par les événements," *Mémoires de l'Académie royale des Sciences de MI (Savants étrangers)*, **4**: 621–656. Reprinted in: Laplace, *Oeuvres complètes* (Paris, France: Gauthier-Villars et fils, 1841), vol. 8, pp. 27–65. Available on-line at: Gallica (<http://gallica.bnf.fr/ark:/12148/bpt6k77596b/f32.image>). Bayes' theorem appears on p. 29.
 - Laplace presented a refinement of Bayes' theorem in: Laplace (read: 1783 / published: 1785) "Mémoire sur les approximations des formules qui sont fonctions de très grands nombres," *Mémoires de l'Académie royale des Sciences de Paris*, 423–467. Reprinted in: Laplace, *Oeuvres complètes* (Paris, France: Gauthier-Villars et fils, 1844), vol. 10, pp. 295–338. Available on-line at: Gallica (<http://gallica.bnf.fr/ark:/12148/bpt6k775981/f218.image.langEN>). Bayes' theorem is stated on page 301.
 - See also: Laplace, *Essai philosophique sur les probabilités* (Paris, France: Mme. Ve. Courcier [Madame veuve (i.e., widow) Courcier], 1814), page 10 (<https://books.google.com/books?id=rDUJAAAIAAJ&pg=PA10#v=onepage&q&f=false>). English translation: Pierre Simon, Marquis de Laplace with F. W. Truscott and F. L. Emory, trans., *A Philosophical Essay on Probabilities* (New York, New York: John Wiley & Sons, 1902), page 15 (<https://google.com/books?id=WxoPAAAIAAJ&pg=PA15#v=onepage&q&f=false>).
- Daston, Lorraine (1988). *Classical Probability in the Enlightenment*. Princeton Univ Press. p. 268. ISBN 0-691-08497-1.

12. Stigler, Stephen M (1983). "Who Discovered Bayes' Theorem?". *The American Statistician* **37** (4): 290–296. doi:10.1080/00031305.1983.10483122.
13. Edwards, A. W. F. (1986). "Is the Reference in Hartley (1749) to Bayesian Inference?". *The American Statistician* **40** (2): 109–110. doi:10.1080/00031305.1986.10475370.
14. Hooper, Martyn (2013). "Richard Price, Bayes' theorem, and God". *Significance* **10** (1): 36–39. doi:10.1111/j.1740-9713.2013.00638.x.
15. McGrayne, S. B. (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines & Emerged Triumphant from Two Centuries of Controversy*. Yale University Press. ISBN 978-0-300-18822-6.

Further reading

- Bruss, F. Thomas (2013), "250 years of 'An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.' ", DOI 10.1365/s13291-013-0077-z, Jahresbericht der Deutschen Mathematiker-Vereinigung, Springer Verlag, Vol. 115, Issue 3-4 (2013), 129-133.
- Gelman, A, Carlin, JB, Stern, HS, and Rubin, DB (2003), "Bayesian Data Analysis", Second Edition, CRC Press.
- Grinstead, CM and Snell, JL (1997), "Introduction to Probability (2nd edition)", American Mathematical Society (free pdf available) [1]
(http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html).
- Hazewinkel, Michiel, ed. (2001), "Bayes formula", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- McGrayne, SB (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines & Emerged Triumphant from Two Centuries of Controversy*. Yale University Press. ISBN 978-0-300-18822-6.
- Laplace, P (1774/1986), "Memoir on the Probability of the Causes of Events", *Statistical Science* 1(3):364–378.
- Lee, PM (2012), "Bayesian Statistics: An Introduction", Wiley.
- Rosenthal, JS (2005), "Struck by Lightning: the Curious World of Probabilities". Harper Collings.
- Stigler, SM (1986). "Laplace's 1774 Memoir on Inverse Probability". *Statistical Science* **1** (3): 359–363. doi:10.1214/ss/1177013620.
- Stone, JV (2013), download chapter 1 of "Bayes' Rule: A Tutorial Introduction to Bayesian Analysis" (<http://jim-stone.staff.shef.ac.uk/BookBayes2012/BayesRuleBookMain.html>), Sebtel Press, England.

External links

- Bayes' theorem (<http://www.britannica.com/EBchecked/topic/56808>) at *Encyclopædia Britannica*
- The Theory That Would Not Die by Sharon Bertsch McGrayne
(<http://www.nytimes.com/2011/08/07/books/review/the-theory-that-would-not-die-by-sharon-bertsch-mcgrayne-book-review.html>) New York Times Book Review by John Allen Paulos on 5 August 2011
- Visual explanation of Bayes using trees (<https://www.youtube.com/watch?v=Zxm4Xxvzohk>) (video)
- Bayes' frequentist interpretation explained visually (<http://www.youtube.com/watch?v=D8VZqxcu0I0>) (video)
- Earliest Known Uses of Some of the Words of Mathematics (B) (<http://jeff560.tripod.com/b.html>). Contains origins of "Bayesian", "Bayes' Theorem", "Bayes Estimate/Risk/Solution", "Empirical Bayes", and "Bayes Factor".
- Weisstein, Eric W., "Bayes' Theorem" (<http://mathworld.wolfram.com/BayesTheorem.html>), *MathWorld*.
- Bayes' theorem (<http://planetmath.org/BayesTheorem>) at PlanetMath.org.
- Bayes Theorem and the Folly of Prediction (<http://rldinvestments.com/Articles/BayesTheorem.html>)
- A tutorial on probability and Bayes' theorem devised for Oxford University psychology students (<http://www.celiagreen.com/charlesmccreery/statistics/bayestutorial.pdf>)

- An Intuitive Explanation of Bayes' Theorem by Eliezer S. Yudkowsky (<http://yudkowsky.net/rational/bayes>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Bayes%27_theorem&oldid=707492466"

Categories: Bayesian statistics | Probability theorems | Statistical theorems

- This page was last modified on 29 February 2016, at 04:53.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.