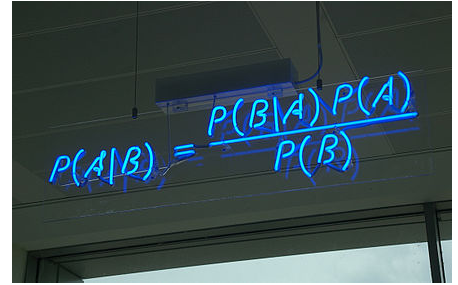WIKIPEDIA

Bayes' theorem

In probability theory and statistics, **Bayes' theorem** (alternatively **Bayes' law** or **Bayes' rule**) describes the probability of an event, based on prior knowledge of conditions that might be related to the event. For example, if cancer is related to age, then, using Bayes' theorem, a person's age can be used to more accurately assess the probability that they have cancer than can be done without knowledge of the person's age.

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. When applied, the probabilities involved in Bayes' theorem may have different probability interpretations. With the Bayesian probability interpretation the theorem expresses how a degree of belief, expressed as a probability, should rationally change to account for availability of related evidence. Bayesian inference is fundamental to Bayesian statistics.

A blue neon sign showing the simple statement of Bayes' theorem at the offices of HP Autonomy

Bayes' theorem is named after Reverend Thomas Bayes (/beɪz/; 1701?–1761), who first used conditional probability to provide an algorithm (his Proposition 9) that uses evidence to calculate limits on an unknown parameter, published as *An Essay towards solving a Problem in the Doctrine of Chances* (1763). In what he called a scholium, Bayes extended his algorithm to any unknown prior cause. Independently of Bayes, Pierre-Simon Laplace in 1774, and later in his 1812 "Théorie analytique des probabilités" used conditional probability to formulate the relation of an updated posterior probability from a prior probability, given evidence. Sir Harold Jeffreys put Bayes's 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

Statement of theorem

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

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

where A and B are events and $P(B) \neq 0$.

- $P(A \mid B)$ is a conditional probability: the likelihood of event A occurring given that B is true.
- $P(B \mid A)$ is also a conditional probability: the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other; this is known as the marginal probability.
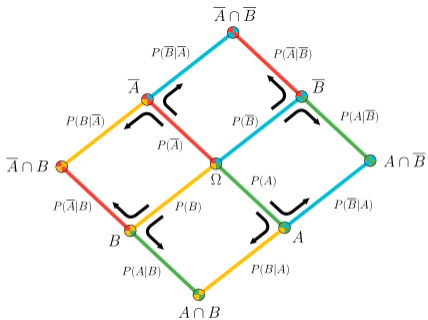
Visualization of Bayes' theorem by superposition of two event tree diagrams.

Examples

Drug testing

Suppose that a test for using a particular drug 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. What is the probability that a randomly selected individual with a positive test is a drug user?
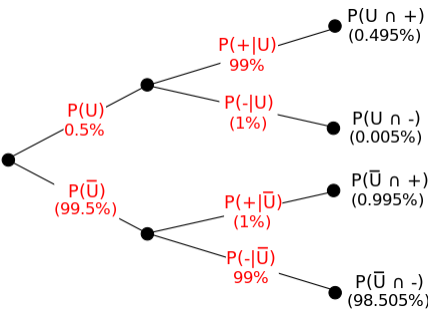
Tree diagram illustrating drug testing example. U, Ū, "+" and "–" are the events representing user, non-user, positive result and negative result. Percentages in parentheses are calculated.

$$P(\text{User} \mid +) = \frac{P(+ \mid \text{User})P(\text{User})}{P(+)}$$

$$= \frac{P(+ \mid \text{User})P(\text{User})}{P(+ \mid \text{User})P(\text{User}) + P(+ \mid \text{Non-user})P(\text{Non-user})}$$

$$= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995}$$

$$\approx 33.2\%$$

Even if an individual tests positive, it is more likely that they do not use the drug than that they do. This is because the number of non-users is large compared to the number of users. The number of false positives outweighs the number of true positives. For example, 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 are genuine.

The importance of specificity in this example can be seen by calculating that even if sensitivity is raised to 100% and specificity remains at 99% then the probability of the person being a drug user only rises from 33.2% to 33.4%, but if the sensitivity is held at 99% and the specificity is increased to 99.5% then the probability of the person being a drug user rises to about 49.9%.

A more complicated example

The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the factory output. The fraction of defective items produced is 5% for the first machine; 3% for the second machine; and 1% for the third machine. 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?

Once again, the answer can be reached without recourse to the formula by applying the conditions to any hypothetical number of cases. For example, if 100,000 items are 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.

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

$$P(X_A) = 0.2, \quad P(X_B) = 0.3, \quad P(X_C) = 0.5.$$

If the item was made by the first machine, then the probability that it is defective is 0.05; that is, $P(Y \mid X_A)$ = 0.05. Overall, we have

$$P(Y \mid X_A) = 0.05, \quad P(Y \mid X_B) = 0.03, \quad P(Y \mid X_C) = 0.01.$$

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

$$P(Y) = \sum_i P(Y \mid X_i)P(X_i) = (0.05)(0.2) + (0.03)(0.3) + (0.01)(0.5) = 0.024.$$

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

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

$$P(X_C \mid Y) = \frac{P(Y \mid X_C)P(X_C)}{P(Y)} = \frac{0.01 \cdot 0.50}{0.024} = \frac{5}{24}$$

Given that the item is defective, the probability that it was made by the third machine is only 5/24. Although machine C 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 $P(X_C)$ = 1/2 by the smaller posterior probability $P(X_C \mid Y)$ = 5/24.

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

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

A geometric visualisation of Bayes' theorem. The figures denote the cells of the table involved in each metric, the probability being the fraction of each figure that is shaded. Similar reasoning shows that $P(\bar{A}|B) = \frac{P(B|\bar{A})P(\bar{A})}{P(B)}$ and so on.

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. $P(B \mid A)$ is the proportion of outcomes with property B *out of* outcomes with property A, and $P(A \mid B)$ 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(Pattern | 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(Rare | Pattern)?

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

P(A ∩ B)

P(B|A)

P(A) P(A)

P(B̄|A)

P(A ∩ B̄)

P(Ā)

P(Ā ∩ B)

P(B|Ā)

P(Ā) P(B̄|Ā)

P(Ā ∩ B̄)

Knowledge of one
diagram is sufficient
to deduce the other

Use Bayes' Theorem to convert between diagrams

$$P(\alpha|\beta)\,P(\beta) \;=\; P(\alpha \cap \beta) \;=\; P(\beta|\alpha)\,P(\alpha)$$

Knowledge of any
3 independent values
is sufficient to deduce
all 24 values

P(B ∩ A)

P(A|B)

P(B) P(B)

P(Ā|B)

P(B ∩ Ā)

P(B̄)

P(B̄ ∩ A)

P(A|B̄)

P(B̄) P(Ā|B̄)

P(B̄ ∩ Ā)
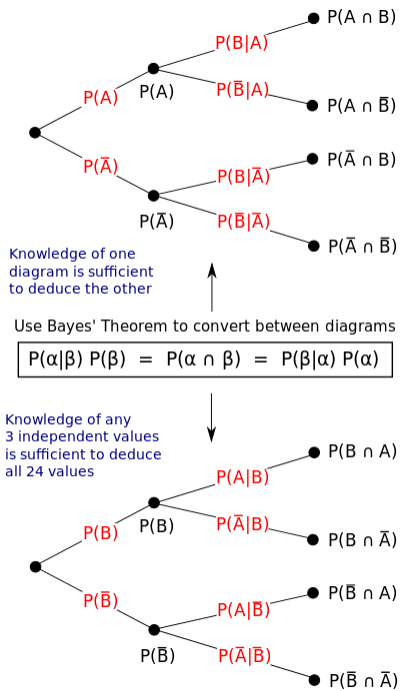
Illustration of frequentist
interpretation with tree diagrams.
Bayes' theorem connects
conditional probabilities to their
inverses.

P(R ∩ P)
(0.098%)

P(P|R)
98%

P(P̄|R)
(2%)

P(R ∩ P̄)
(0.002%)

P(R)
0.1%

P(C ∩ P)
(4.995%)

P(P|C)
5%

P(C)
(99.9%)

P(P̄|C)
(95%)

P(C ∩ P̄)
(94.905%)
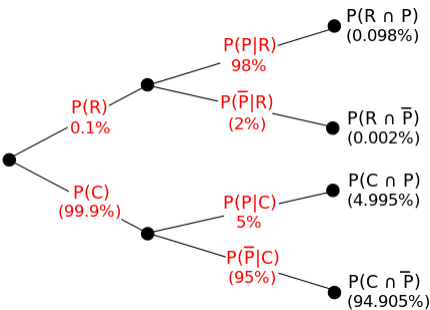
Tree diagram illustrating frequentist
example. R, C, P and P bar are the
events representing rare, common,
pattern and no pattern. Percentages
in parentheses are calculated.
Three independent values are
given, so it is possible to calculate
the inverse tree (see figure above).

$$P(\text{Rare} \mid \text{Pattern}) = \frac{P(\text{Pattern} \mid \text{Rare})P(\text{Rare})}{P(\text{Pattern})}$$

$$= \frac{P(\text{Pattern} \mid \text{Rare})P(\text{Rare})}{P(\text{Pattern} \mid \text{Rare})P(\text{Rare}) + P(\text{Pattern} \mid \text{Common})P(\text{Common})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.05 \times 0.999}$$

$$\approx 1.9\%$$

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:

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

The posterior is proportional to the prior times the likelihood.[3]

If events A_1, A_2, ..., 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 ¬*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

$$1 = c \cdot (P(B \mid A) \cdot P(A) + P(B \mid \neg A) \cdot P(\neg A)),$$

or

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

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,[4]

- $P(A)$ is 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, where $1 - P(A) = P(\neg A)$
- $P(B \mid A)$ is the conditional probability or likelihood, is the degree of belief in B, given that the proposition A is true.
- $P(B \mid \neg A)$ is the conditional probability or likelihood, is the degree of belief in B, given that the proposition A is false.
- $P(A \mid B)$ is 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 $B = \{Y = y\}$.

$$P(X = x \mid Y = y) = \frac{P(Y = y \mid X = x)P(X = x)}{P(Y = y)}$$

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).

Diagram illustrating the meaning of Bayes' theorem as applied to an event space generated by continuous random variables X and Y. 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.

Simple form

If X is continuous and Y is discrete,

$$f_{X \mid Y=y}(x) = \frac{P(Y = y \mid X = x)f_X(x)}{P(Y = y)}$$

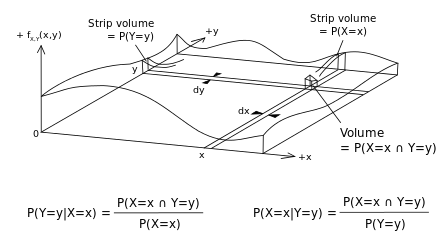where each f is a density function.

If X is discrete and Y is continuous,

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

If both X and Y are continuous,

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

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:

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

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

Bayes' rule

Bayes' theorem in odds form is:

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

where

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

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

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

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

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.

In the special case that $A_1 = A$ and $A_2 = \neg A$, one writes $O(A) = O(A : \neg A) = P(A)/(1 - P(A))$, and uses a similar abbreviation for the Bayes factor and for the conditional odds. The odds on A is by definition the odds for and against A. Bayes' rule can then be written in the abbreviated form

$$O(A|B) = O(A) \cdot \Lambda(A|B),$$

or in words: the posterior odds on A equals the prior odds on A times the likelihood ratio for A given information B. In short, **posterior odds equals prior odds times likelihood ratio.**
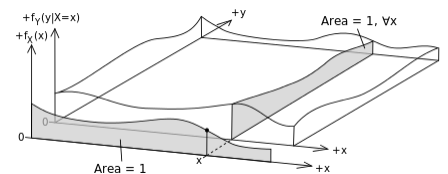
Derivation

For events

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0,$$

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}, \text{ if } P(A) \neq 0,$$

where $P(A \cap B)$ is the joint probability of both A and B being true, because

$$P(B \cap A) = P(A \cap B)$$
$$\Rightarrow P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$
$$\Rightarrow P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \text{ if } P(B) \neq 0.$$

For random variables

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

$$f_{X \mid Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_{Y \mid X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Therefore,

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

Correspondence to other mathematical frameworks

Propositional logic

Bayes' theorem represents a generalisation of contraposition which in propositional logic can be expressed as:

$$(\neg A \to \neg B) \to (B \to A).$$

The corresponding formula in terms of probability calculus is Bayes' theorem which in its expanded form is expressed as:

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

In the equation above the conditional probability $P(B \mid A)$ generalizes the logical statement $(A \to B)$, i.e. in addition to assigning TRUE or FALSE we can also assign any probability to the statement. The term $a(A)$ denotes the prior probability (aka. the base rate) of A. Assume that $P(A \mid B) = 1$ is equivalent to $(B \to A)$ being TRUE, and that $P(A \mid B) = 0$ is equivalent to $(B \to A)$ being FALSE. It is then easy to see that $P(A \mid B) = 1$ when $P(\neg B \mid \neg A) = 1$

i.e. when $(\neg A \to \neg B)$ is TRUE. This is because $P(B \mid \neg A) = 1 - P(\neg B \mid \neg A) = 0$ so that the fraction on the right-hand side of the equation above is equal to 1, and hence $P(A \mid B) = 1$ which is equivalent to $(B \to A)$ being TRUE. Hence, Bayes' theorem represents a generalization of contraposition.[5]

Subjective logic

Bayes' theorem represents a special case of conditional inversion in subjective logic expressed as:

$$(\omega^S_{A\tilde{|}B}, \omega^S_{A\tilde{|}\neg B}) = (\omega^S_{B|A}, \omega^S_{B|\neg A}) \widetilde{\phi} a_A,$$

where $\widetilde{\phi}$ denotes the operator for conditional inversion. The argument $(\omega^S_{B|A}, \omega^S_{B|\neg A})$ denotes a pair of binomial conditional opinions given by source S, and the argument a_A denotes the prior probability (aka. the base rate) of A. The pair of inverted conditional opinions is denoted $(\omega^S_{A\tilde{|}B}, \omega^S_{A\tilde{|}\neg B})$. The conditional opinion $\omega^S_{A|B}$ generalizes the probabilistic conditional $P(A \mid B)$, i.e. in addition to assigning a probability the source S can assign any subjective opinion to the conditional statement $(A \mid B)$. A binomial subjective opinion ω^S_A is the belief in the truth of statement A with degrees of uncertainty, as expressed by source S. Every subjective opinion has a corresponding projected probability $P(\omega^S_A)$. The projected probability of opinions applied to Bayes' theorem produces a homomorphism so that Bayes' theorem can be expressed in terms of the projected probabilities of opinions:

$$P(\omega^S_{A\tilde{|}B}) = \frac{P(\omega^S_{B|A})a(A)}{P(\omega^S_{B|A})a(A) + P(\omega^S_{B|\neg A})a(\neg A)}.$$

Hence, the subjective Bayes' theorem represents a generalization of Bayes' theorem.[6]

Generalizations

Conditioned version

A conditioned version of the Bayes' theorem[7] results from the addition of a third event C on which all probabilities are conditioned:

$$P(A \mid B \cap C) = \frac{P(B \mid A \cap C)\, P(A \mid C)}{P(B \mid C)}$$

Derivation

Using the chain rule

$$P(A \cap B \cap C) = P(A \mid B \cap C)\, P(B \mid C)\, P(C)$$

And, on the other hand

$$P(A \cap B \cap C) = P(B \cap A \cap C) = P(B \mid A \cap C)\, P(A \mid C)\, P(C)$$

The desired result is obtained by identifying both expressions and solving for $P(A \mid B \cap C)$.

History

Bayes' theorem was named after Thomas Bayes (1701–1761), who studied how to compute a distribution for the probability parameter of a binomial distribution (in modern terminology). Bayes's unpublished manuscript was significantly edited by Richard Price before it was posthumously read at the Royal Society. Price edited[8] Bayes's major work "An Essay towards solving a Problem in the Doctrine of Chances" (1763), which appeared in *Philosophical Transactions*,[9] 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.[10][11]

The French mathematician Pierre-Simon Laplace reproduced and extended Bayes's results in 1774, apparently unaware of Bayes's work.[note 1][12] The Bayesian interpretation of probability was developed mainly by Laplace.[13]

Stephen Stigler used a Bayesian argument to conclude that Bayes' theorem was discovered by Nicholas Saunderson, a blind English mathematician, some time before Bayes;[14][15] that interpretation, however, has been disputed.[16] Martyn Hooper[17] and Sharon McGrayne[18] 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.[18]

See also

- Quantum Bayesianism
- Bayesian inference
- Bayesian probability
- Inductive probability

Notes

1. 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=rDUJAAAAIAAJ&pg=PA10#v=onepage). 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=WxoPAAAAIAAJ&pg=PA15#v=onepage).

References

1. Jeffreys, Harold (1973). *Scientific Inference* (3rd ed.). Cambridge University Press. p. 31. ISBN 978-0-521-18078-8.
2. Stuart, A.; Ord, K. (1994), *Kendall's Advanced Theory of Statistics: Volume I—Distribution Theory*, Edward Arnold, §8.7

3. Lee, Peter M. (2012). "Chapter 1". *Bayesian Statistics* (http://www-users.york.ac.uk/~pml1/bayes/book.htm). Wiley. ISBN 978-1-1183-3257-3.

4. "Bayes' Theorem: Introduction" (https://web.archive.org/web/20040821012342/http://www.trinity.edu/cbrown/bayesweb/). *Trinity University*. Archived from the original (http://www.trinity.edu/cbrown/bayesweb/) on 21 August 2004. Retrieved 5 August 2014.

5. Audun Jøsang, 2016, *Subjective Logic; A formalism for Reasoning Under Uncertainty.* Springer, Cham, ISBN 978-3-319-42337-1

6. Audun Jøsang, 2016, *Generalising Bayes' Theorem in Subjective Logic (http://folk.uio.no/josang/papers/Josang2016-MFI.pdf)*. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016), Baden-Baden, September 2016

7. Koller, D.; Friedman, N. (2009). *Probabilistic Graphical Models* (http://pgm.stanford.edu/). Massachusetts: MIT Press. p. 1208. ISBN 978-0-262-01319-2.

8. Allen, Richard (1999). *David Hartley on Human Nature* (https://books.google.com/books?id=NCu6HhGlAB8C&pg=PA243). SUNY Press. pp. 243–4. ISBN 978-0-7914-9451-6. Retrieved 16 June 2013.

9. Bayes, Thomas & 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." (https://web.archive.org/web/20110410085940/http://www.stat.ucla.edu/history/essay.pdf) (PDF). *Philosophical Transactions of the Royal Society of London*. **53** (0): 370–418. doi:10.1098/rstl.1763.0053 (https://doi.org/10.1098%2Frstl.1763.0053). Archived from the original (http://www.stat.ucla.edu/history/essay.pdf) (PDF) on 2011-04-10. Retrieved 2003-12-27.

10. Holland, pp. 46–7.

11. Price, Richard (1991). *Price: Political Writings* (https://books.google.com/books?id=xdH-gjy2vzUC&pg=PR23). Cambridge University Press. p. xxiii. ISBN 978-0-521-40969-8. Retrieved 16 June 2013.

12. Daston, Lorraine (1988). *Classical Probability in the Enlightenment* (https://books.google.com/books?id=oq8XNbKyUewC&pg=PA268). Princeton Univ Press. p. 268. ISBN 0-691-08497-1.

13. Stigler, Stephen M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Chapter 3.

14. Stigler, Stephen M. (1983). "Who Discovered Bayes' Theorem?". *The American Statistician*. **37** (4): 290–296. doi:10.1080/00031305.1983.10483122 (https://doi.org/10.1080%2F00031305.1983.10483122).

15. de Vaux, Richard; Velleman, Paul; Bock, David (2016). *Stats, Data and Models* (4th ed.). Pearson. pp. 380–381. ISBN 978-0-321-98649-8.

16. 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 (https://doi.org/10.1080%2F00031305.1986.10475370).

17. Hooper, Martyn (2013). "Richard Price, Bayes' theorem, and God". *Significance*. **10** (1): 36–39. doi:10.1111/j.1740-9713.2013.00638.x (https://doi.org/10.1111%2Fj.1740-9713.2013.00638.x).

18. 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 (https://doi.org/10.1365%2Fs13291-013-0077-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) [1994], "Bayes formula" (https://www.encyclopediaofmath.org/index.php?title=p/b015380), *Encyclopedia of Mathematics*, Springer Science+Business Media B.V. / Kluwer Academic Publishers, 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, Peter M (2012), "Bayesian Statistics: An Introduction," 4th edition. Wiley. ISBN 978-1-118-33257-3.
- Puga JL, Krzywinski M, Altman N (31 March 2015). "Bayes' theorem" (http://www.nature.com/nmeth/journal/v12/n 4/full/nmeth.3335.html). *Nature Methods*. **12** (4): 277–278.
- Rosenthal, Jeffrey S (2005), "Struck by Lightning: The Curious World of Probabilities." HarperCollins. (Granta, 2008. ISBN 9781862079960).
- Stigler, SM (1986). "Laplace's 1774 Memoir on Inverse Probability". *Statistical Science*. **1** (3): 359–363. doi:10.1214/ss/1177013620 (https://doi.org/10.1214%2Fss%2F1177013620).
- Stone, JV (2013), download chapter 1 of "Bayes' Rule: A Tutorial Introduction to Bayesian Analysis" (http://jim-ston e.staff.shef.ac.uk/BookBayes2012/HTML_BayesRulev5EbookHTMLFiles/ops/xhtml/ch01BayesJVSone.html), Sebtel Press, England.
- *Bayesian Reasoning for Intelligent People (http://santafe.edu/~simon/br.pdf)*, An introduction and tutorial to the use of Bayes' theorem in statistics and cognitive science.
- Morris, Dan (2016), Read first 6 chapters for free of "Bayes' Theorem Examples: A Visual Introduction For Beginners (https://www.bayestheorem.net/)" Blue Windmill ISBN 978-1549761744. A short tutorial on how to understand problem scenarios and find P(B), P(A), and P(B|A).

External links

- Bayes' theorem (https://www.britannica.com/EBchecked/topic/56808) at the *Encyclopædia Britannica*
- The Theory That Would Not Die by Sharon Bertsch McGrayne (https://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 (https://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 (https://planetmath.org/BayesTheorem) at PlanetMath.org.
- Bayes Theorem and the Folly of Prediction (https://web.archive.org/web/20140202131520/http://rldinvestments.co m/Articles/BayesTheorem.html)
- A tutorial on probability and Bayes' theorem devised for Oxford University psychology students (http://www.celiagr een.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=917938455"

This page was last edited on 26 September 2019, at 06:24 (UTC).