Bayes' theorem

In probability theory and statistics, Bayes' theorem (alternatively Bayes's law or Bayes's rule) is a theorem with two distinct interpretations. In the Bayesian interpretation, it expresses how a subjective degree of belief should rationally change to account for evidence. In the frequentist interpretation, it relates inverse representations of the probabilities concerning two events. In the Bayesian interpretation, Bayes' theorem is fundamental to Bayesian statistics, and has applications in fields including science, engineering, medicine and law. The application of Bayes' theorem to update beliefs is called Bayesian inference.

Bayes' theorem is named for Thomas Bayes (English pronunciation: /ˈbeɪz/), who first suggested using the theorem to update beliefs. However, his work was published posthumously. His ideas gained limited exposure until they were independently rediscovered and further developed by Laplace, who first published the modern formulation in his 1812 Théorie analytique des probabilités. Until the second half of the 20th century, the Bayesian interpretation attracted widespread dissent from the mathematics community who generally held frequentist views, rejecting Bayesianism as unscientific. However, it is now widely accepted. This may have been due to the development of computing, which enabled the successful application of Bayesianism to many complex problems.[1]

Introductory example

If someone told you he had a nice conversation in the train, the probability it was a woman he spoke with is 50%. If he told you the person he spoke to was going to visit a quilt exhibition, it is far more likely than 50% it is a woman. Call $W$ the event he spoke to a woman, and $Q$ the event "a visitor of the quilt exhibition". Then: $P(W) = 0.50$

$P(W|Q)$, but with the knowledge of $Q$ the updated value is $P(W|Q)$ that may be calculated with Bayes' formula as:

$P(W|Q) = \frac{P(Q|W)P(W)}{P(Q)} = \frac{P(Q|W)P(W)}{P(Q|W)P(W) + P(Q|M)P(M)}$

in which $M$ (man) is the complement of $W$. As $P(M) = P(W) = 0.5$ and $P(Q|W) \gg P(Q|M)$, the updated value will be quite close to 1.

Statement and interpretation

Mathematically, Bayes' theorem gives the relationship between the probabilities of $A$ and $B$: $P(A)$ and $P(B)$, and the conditional probabilities of $A$ given $B$ and $B$ given $A$, $P(A|B)$ and $P(B|A)$. In its most common form, it is:

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

The meaning of this statement depends on the interpretation of probability ascribed to the terms:
### 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 somebody proposes that a biased coin is twice as likely to land heads than tails. Degree of belief in this might initially be 50%. The coin is then flipped a number of times to collect evidence. Belief may rise to 70% if the evidence supports the proposition.

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 $B$.
- $P(B|A)/P(B)$ 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 is defined with respect to a large number of trials, each producing one outcome from a set of possible outcomes, $\Omega$. An event is a subset of $\Omega$. The probability of event $A$, $P(A)$, is the proportion of trials producing an outcome in $A$. Similarly for $B$, $P(B)$. If we consider only trials in which $A$ occurs, the proportion in which $B$ also occurs is $P(B|A)$. If we consider only trials in which $B$ occurs, the proportion in which $A$ also occurs is $P(A|B)$. Bayes’ theorem is a fixed relationship between these quantities.

This situation may be more fully visualised with tree diagrams, shown to the right. The two diagrams represent the same information in different ways. For example, suppose that $A$ is having a risk factor for a medical condition, and $B$ is having the condition. In a population, the proportion with the condition depends whether those with or without the risk factor are examined. The proportion having the risk factor depends whether those with or without the condition are examined. Bayes' theorem links these inverse representations.

### Forms

#### For events

**Simple form**

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

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

In a Bayesian inference step, the probability of evidence $B$ is constant for all models $A_m$. The posterior may then be expressed as proportional to the numerator:

$$P(A_m|B) \propto P(B|A_m)P(A_m).$$
**Extended form**

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

\[
P(B) = \sum_j P(B|A_j)P(A_j).
\]

\[\Rightarrow P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}.
\]

In the special case of a binary partition,

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}.
\]

**Three or more events**

Extensions to Bayes' theorem may be found for three or more events. For example, for three events, two possible tree diagrams branch in the order BCA and ABC. By repeatedly applying the definition of conditional probability:

\[
P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(C|A \cap B)P(A \cap B)}{P(C|B)P(B)} = \frac{P(C|A \cap B)P(B|A)P(A)}{P(C|B)P(B)}.
\]

As previously, the law of total probability may be substituted for unknown marginal probabilities.

**For random variables**

Consider a sample space \(\Omega\) 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:

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

Bayes' rule

Under the Bayesian interpretation of probability, Bayes' rule may be thought of as Bayes' theorem in odds form.

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

Where

\[
\Lambda(A_1 : A_2 \mid B) = \frac{P(B \mid A_1)}{P(B \mid A_2)}.
\]

Derivation

For general events

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

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

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

\[
\Rightarrow P(A \cap B) = P(A \mid B) P(B) = P(B \mid A) P(A).
\]

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

For random variables

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

\[
f_X(x \mid Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}
\]

\[
f_Y(y \mid X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}
\]

\[
\Rightarrow f_X(x \mid Y = y) = \frac{f_Y(y \mid X = x) \, f_X(x)}{f_Y(y)}.
\]
Examples

Frequentist 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.
In the common subspecies, 5% have the pattern. The rare subspecies
accounts for only 0.1% of the population. How likely is the beetle to be
rare?

From the extended form of Bayes' theorem,

\[
P(\text{Rare}|\text{Pattern}) = \frac{P(\text{Pattern}|\text{Rare})P(\text{Rare})}{P(\text{Pattern}|\text{Rare})P(\text{Rare}) + P(\text{Pattern}|\text{Common})P(\text{Common})} = \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.05 \times 0.999} \approx 1.9%\]

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 they are a user?

\[
P(\text{User}+) = \frac{P(+|\text{User})P(\text{User})}{P(+|\text{User})P(\text{User}) + P(+|\text{Non-user})P(\text{Non-user})} = \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \approx 33.2%\]
Bayes' theorem

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 surprising result arises because the number of non-users is very large compared to the number of users, such that 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 × 995 ≈ 10 false positives are expected. From the 5 users, 0.99 × 5 ≈ 5 true positives are expected. Out of 15 positive results, only 5, about 33%, are genuine.

History

Bayes' theorem was named after the Reverend Thomas Bayes (1702–61), who studied how to compute a distribution for the probability parameter of a binomial distribution (in modern terminology). His friend Richard Price edited and presented this work in 1763, after Bayes' death, as *An Essay towards solving a Problem in the Doctrine of Chances*. The French mathematician Pierre-Simon Laplace reproduced and extended Bayes' results in 1774, apparently quite unaware of Bayes' work. Stephen Stigler suggested in 1983 that Bayes' theorem was discovered by Nicholas Saunderson some time before Bayes; however, this interpretation has been disputed.

Stephen Fienberg describes the evolution from "inverse probability" at the time of Bayes and Laplace, a term still used by Harold Jeffreys (1939), to "Bayesian" in the 1950s. Ironically, Ronald A. Fisher introduced the "Bayesian" label in a derogatory sense.

Notes


Further reading

**External links**

- Bayesian Statistics (http://www.scholarpedia.org/article/Bayesian_statistics) summary from Scholarpedia.
- A tutorial on probability and Bayes’ theorem devised for Oxford University psychology students (http://www.celiagreen.com/charlesmccreery/statistics/bayestutorial.pdf)
Article Sources and Contributors


Image Sources, Licenses and Contributors


License

Creative Commons Attribution-Share Alike 3.0 Unported Attribution commons://license/by-sa/3.0/