1 Bayes' theorem

Bayes' theorem (also known as Bayes' rule or Bayes' law) is a result in probability theory that relates conditional probabilities. If A and B denote two events, $P(A|B)$ denotes the conditional probability of A occurring, given that B occurs. The two conditional probabilities $P(A|B)$ and $P(B|A)$ are in general different. Bayes theorem gives a relation between $P(A|B)$ and $P(B|A)$.

An important application of Bayes' theorem is that it gives a rule how to update or revise the strengths of evidence-based beliefs in light of new evidence a posteriori.

As a formal theorem, Bayes' theorem is valid in all interpretations of probability. However, it plays a central role in the debate around the foundations of statistics: frequentist and Bayesian interpretations disagree about the kinds of things to which probabilities should be assigned in applications. Whereas frequentists assign probabilities to random events according to their frequencies of occurrence or to subsets of populations as proportions of the whole, Bayesians assign probabilities to propositions that are uncertain. A consequence is that Bayesians have more frequent occasion to use Bayes' theorem. The articles on Bayesian probability and frequentist probability discuss these debates at greater length.

2 Statement of Bayes' theorem

Bayes' theorem relates the conditional and marginal probabilities of stochastic events A and B:

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

Each term in Bayes' theorem has a conventional name:

- P(A) is the prior probability or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the conditional probability of B given A.
- P(B) is the prior or marginal probability of B, and acts as a normalizing constant.

3 Bayes' theorem in terms of likelihood

Bayes' theorem can also be interpreted in terms of likelihood:

$P(A|B) \propto L(A|B)\,P(A)$.

Here $L(A|B)$ is the likelihood of A given fixed B. The rule is then an immediate consequence of the relationship $P(B|A) = L(A|B)$. In many contexts the likelihood function L can be multiplied by a constant factor, so that it is proportional to, but does not equal the conditional probability P.

With this terminology, the theorem may be paraphrased as

posterior = $\frac{\text{likelihood}\times\text{prior}}{\text{normalizing constant}}$

In words: the posterior probability is proportional to the product of the prior probability and the likelihood.

In addition, the ratio $L(A|B)/P(B)$ is sometimes called the standardized likelihood or normalized likelihood, so the theorem may also be paraphrased as

posterior = normalized likelihood × prior.

4 Derivation from conditional probabilities

To derive the theorem, we start from the definition of conditional probability. The probability of event A given event B is

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

Likewise, the probability of event B given event A is

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

Rearranging and combining these two equations, we find

$P(A|B)\,P(B) = P(A\cap B) = P(B|A)\,P(A)$.

This lemma is sometimes called the product rule for probabilities. Dividing both sides by P(B), providing that it is non-zero, we obtain Bayes' theorem:

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

5 Alternative forms of Bayes' theorem

Bayes' theorem is often embellished by noting that

$P(B) = P(A\cap B) + P(A^C\cap B) = P(B|A)P(A) + P(B|A^C)P(A^C)$

where AC is the complementary event of A (often called "not A"). So the theorem can be restated as

$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A)+P(B|A^C)P(A^C)}$.

More generally, where Ai forms a partition of the event space,

$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum_j P(B|A_j)\,P(A_j)}$,

for any Ai in the partition.

See also the law of total probability.

6 Bayes' theorem in terms of odds and likelihood ratio

Bayes' theorem can also be written neatly in terms of a likelihood ratio and odds O as

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

where $O(A|B) = \frac{P(A|B)}{P(A^C|B)}$ are the odds of A given B,

and $O(A) = \frac{P(A)}{P(A^C)}$ are the odds of A by itself,

while $\Lambda(A|B) = \frac{L(A|B)}{L(A^C|B)} = \frac{P(B|A)}{P(B|A^C)}$ is the likelihood ratio.

7 Bayes' theorem for probability densities

There is also a version of Bayes' theorem for continuous distributions. It is somewhat harder to derive, since probability densities, strictly speaking, are not probabilities, so Bayes' theorem has to be established by a limit process; see Papoulis (citation below), Section 7.3 for an elementary derivation. Bayes's theorem for probability densities is formally similar to the theorem for probabilities:

$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(y|x)\, f(x)}{f(y)}$

and there is an analogous statement of the law of total probability:

$f(x|y) = \frac{f(y|x)\, f(x)}{\int_{-\infty}^{\infty} f(y|x)\, f(x)\, dx}$.

As in the discrete case, the terms have standard names. f(x, y) is the joint distribution of X and Y, f(x—y) is the posterior distribution of X given Y=y, f(y—x) = L(x—y) is (as a function of x) the likelihood function of X given Y=y, and f(x) and f(y) are the marginal distributions of X and Y respectively, with f(x) being the prior distribution of X.

Here we have indulged in a conventional abuse of notation, using f for each one of these terms, although each one is really a different function; the functions are distinguished by the names of their arguments.