

Bayesian Probability

Nik Bear Brown

In this lesson we will discuss Bayesian probability theory. There are no data sets, or libraries to be installed.

Probability and Statistics

Probability is a measure of the likelihood of a random phenomenon or chance behavior. Probability describes the long-term proportion with which a certain outcome will occur in situations with short-term uncertainty.

The Axioms of Probability

First axiom - The probability of an event is a non-negative real number:

$$P(E) \in \mathbb{R}, P(E) \geq 0 \quad \forall E \in F$$

where F is the event space

Second axiom - unit measure:

The probability that some elementary event in the entire sample space will occur is 1.

$$P(\Omega) = 1.$$

Third axiom - the assumption of σ -additivity:

Any countable sequence of disjoint (synonymous with mutually exclusive) events E_1, E_2, \dots satisfies

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

The total probability of all possible events always sums to 1.

Consequences of these axioms

The probability of the empty set:

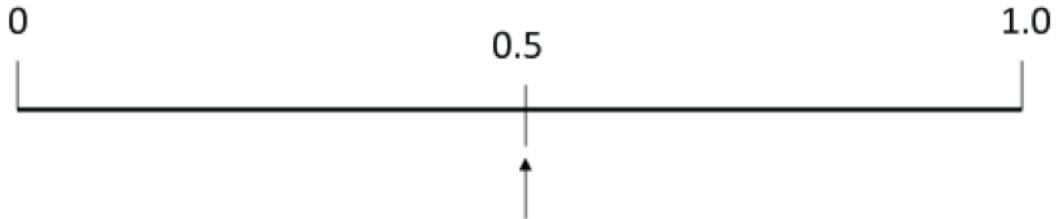
$$P(\emptyset) = 0.$$

Monotonicity

$$\text{if } A \subseteq B \text{ then } P(A) \leq P(B).$$

The numeric bound between 0 and 1:

$$0 \leq P(E) \leq 1 \quad \forall E \in F.$$



The occurrence of the event is just as likely as it is unlikely

Probability is expressed in numbers between 0 and 1.

Probability of a certain event is 1:

$$P(\text{True}) = 1$$

Probability = 1 means it always happens.

Probability of an impossible event is 0:

$$P(\text{False}) = 0$$

Probability = 0 means the event never happens.

Probability of A or B:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

or

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability of not A:

$$P(\text{not } A) = 1 - P(A)$$

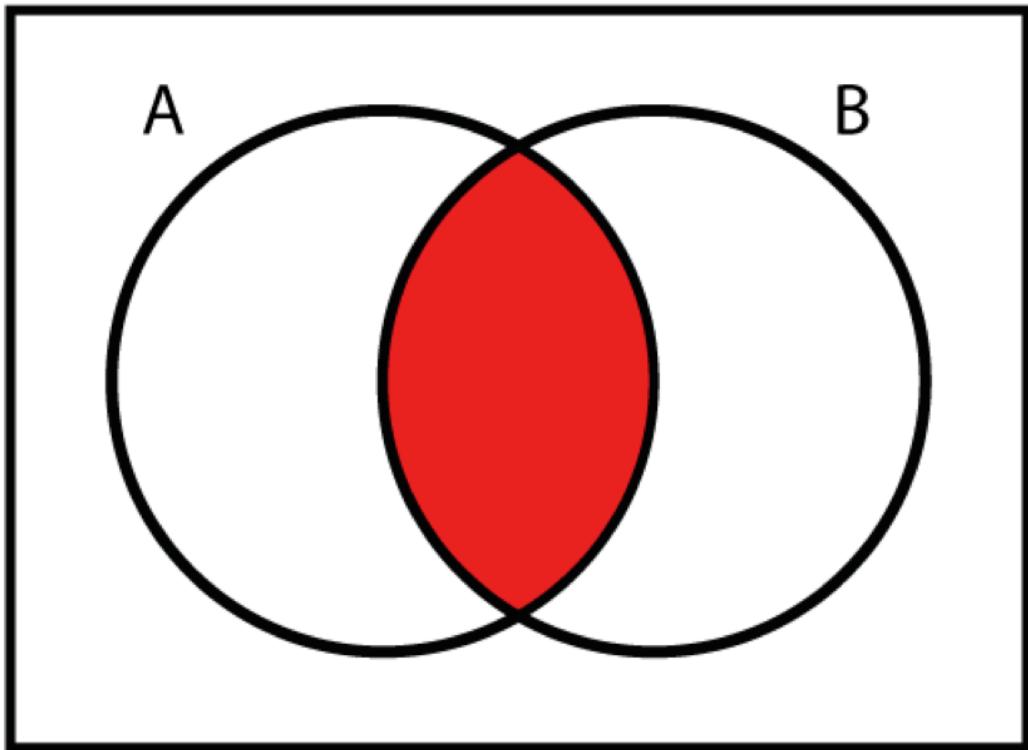
Conditional Probability

In probability theory, a **conditional probability** measures the probability of an event given that another event has occurred. That is, "the conditional probability of A given B."

the conditional probability of A given B is defined as the quotient of the probability of the joint of events A and B, and the probability of B:

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

This may be visualized using a Venn diagram.



$$P(A \cap B)$$

Corollary of Conditional Probability is The Chain Rule

If we multiply both sides by $P(B)$ then

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

becomes

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

Statistical independence

Events A and B are defined to be statistically independent if:

$$\begin{aligned}
 P(A \cap B) &= P(A)P(B) \\
 \Leftrightarrow P(A|B) &= P(A) \\
 \Leftrightarrow P(B|A) &= P(B)
 \end{aligned}$$

That is, the occurrence of A does not affect the probability of B, and vice versa

Probability of A or B for independent events $P(A \text{ and } B)$ is 0:

$$P(A \text{ or } B) = P(A) + P(B)$$

Bayes Rule

[Bayes' theorem](#) (alternatively Bayes' law or Bayes' rule) describes the probability of an event, given prior events. That is, a conditional probability.

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

where A and B are events.

- $P(A)$ and $P(B)$ are the independent probabilities of A and B.
- $P(A | B)$, 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.

Bayesian inference

[Bayesian inference](#) is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as evidence. Bayesian inference derives the posterior probability as a consequence of two antecedents, a prior probability and a "likelihood function" derived from a statistical model for the observed data.

Bayesian inference computes the posterior probability according to Bayes' theorem:

$$P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E)}$$

where,

$P(H | E)$ the posterior probability, denotes a conditional probability of H (the hypothesis) whose probability may be affected by the evidence E .

$P(H)$, the prior probability, is an estimate of the probability that a hypothesis is true, before observing the current evidence.

$P(E|H)$ is the probability of observing E given H . It indicates the compatibility of the evidence with the given hypothesis.

$P(E)$ is sometimes termed the marginal likelihood or "model evidence". This factor is the same for all possible hypotheses being considered.

Note that Bayes' rule can also be written as follows:

$$P(H|E) = \frac{P(E|H)}{P(E)} \cdot P(H)$$

where the factor $\frac{P(E|H)}{P(E)}$ represents the impact of E on the probability of H .

Bayesian probability example

Suppose a certain disease has an incidence rate of 0.01% (that is, it afflicts 0.01% of the population). A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it), but the false positive rate is 1% (that is, about 1% of people who take the test will test positive, even though they do not have the disease). Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease?

Bayes theorem would ask the question, what is the probability of disease given a positive result, or $P(disease|positive)$.

What do we know?

$P(positive|disease) = 1$ (i.e. The test does not produce false negatives.)

$P(disease) = 0.0001$ (i.e. 1/10,000 have the disease)

$P(positive|nodisease) = 0.01$ (i.e. the false positive rate is 1%. This means 1% of people who take the test will test positive, even though they do not have the disease)

Bayes' Theorem

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

which can be rewritten as

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

which in our example is

$$= \frac{P(\text{disease}|\text{positive})}{P(\text{disease}) P(\text{positive}|\text{disease})} = \frac{P(\text{disease}) P(\text{positive}|\text{disease})}{P(\text{disease}) P(\text{positive}|\text{disease}) + P(\text{no disease}) P(\text{positive}|\text{no disease})}$$

plugging in the numbers gives

$$P(\text{disease}|\text{positive}) = \frac{(0.0001)(1)}{(0.0001)(1) + (0.9999)(0.01)} \approx 0.01$$

So even though the test is 99% accurate, of all people who test positive, over 99% do not have the disease.

Bayesians versus Frequentists

Frequentist inference or frequentist statistics is a scheme for making statistical inference based on the frequency or proportion of the data. This effectively requires that conclusions should only be drawn with a set of repetitions.

Frequentists will only generate statistical inference given a large enough set of repetitions. In contrast, a Bayesian approach to inference does allow probabilities to be associated with unknown parameters.



Count Von Count

- from https://en.wikipedia.org/wiki/File:Count_von_Count_kneeling.png

While "probabilities" are involved in both approaches to inference, frequentist probability is essentially equivalent to counting. The Bayesian approach allows these estimates of probabilities to be based upon counting but also allows for subjective estimates (i.e. guesses) of prior probabilities.

Bayesian probability, also called evidential probability, or subjectivist probability, can be assigned to any statement whatsoever, even when no random process is involved. Evidential probabilities are considered to be degrees of belief, and a Bayesian can even use an un-informative prior (also called a non-informative or Jeffreys prior).

In Bayesian probability, the [Jeffreys prior](#), named after Sir Harold Jeffreys, is a non-informative (objective) prior distribution for a parameter space. The crucial idea behind the Jeffreys prior is the Jeffreys posterior. This posterior aims to reflect as best as possible the information about the parameters brought by the data, in effect "representing ignorance" about the prior. This is sometimes called the "principle of indifference." Jeffreys prior is proportional to the square root of the determinant of the Fisher information:

$$p(\vec{\theta}) \propto \sqrt{\det\mathcal{J}(\vec{\theta})}.$$

It has the key feature that it is invariant under reparameterization of the parameter vector $\vec{\theta}$.

At its essence the Bayesian can be vague or subjective about an initial guess at a prior probability, and the posterior probability be updated data point by data point. A Bayesian defines a "probability" in the same way that many non-statisticians do - namely an indication of the plausibility or belief of a proposition.

A Frequentist is someone that believes probabilities represent long run frequencies with which events occur; he or she will have a model (e.g. Gaussian, uniform, etc.) of how the sample population was generated. The observed counts are considered a random sample to estimate the true parameters of the model.

It is important to note that most Frequentist methods have a Bayesian equivalent (that is, they give the same results) when there are enough repeated trials. That is, they converge to the same result given enough data.

Resources

- [An idiot learns Bayesian analysis: Part 1](#) via @rbloggers
- [Using R for Bayesian Statistics](#)
- [Three Ways to Run Bayesian Models in R](#)
- [Bayes Rule in R | Bayesian Thinking](#)
- [R Code for Bayesian Computation with R](#)

