

Bayesian data analysis

Shengping Yang PhD, Gilbert Berdine MD

Bayesian analysis is known to be able to incorporate prior information into decision making. This can be helpful when applied to clinical data analysis. I am wondering how Bayesian differs from the frequentist's approach.

In a frequentist model, probability is the limit of the relative frequency of an event in repeated experiments. For example, the hospital mortality rate of patients with a certain disease can be estimated from observing the number of alive and expired patients at discharge, and in general, a p value and/or a confidence interval will be provided. While the conclusion made depends only on the data collected, which is objective, the often used significance level 0.05 is considered subjective. The frequentist model assumes that the model for the likelihood does not change over time. For example, a frequentist model of hospital mortality in COVID-19 patients based only on the gender of the patient assumes that the population composition by age either does not affect outcome or is constant over time.

As a comparison, in Bayesian statistics, "probability is orderly opinion, and that inference from data is nothing other than the revision of such opinion in the light of relevant new information." In other words, the probability of an event in Bayesian analysis can be updated with the inclusion of additional information, which is from data. This is possible because in Bayesian analysis all parameters in a model can be assumed to be random quantities instead of each parameter being a fixed value as in frequentist analysis. To effectively perform Bayesian data analyses, it is critical to have a good understanding the Bayes's rule (also called Bayes's Theorem).

Corresponding author: Shengping Yang
Contact Information: Shengping.Yang@pbrcc.edu
DOI: 10.12746/swrccc.v8i36.773

THE BAYES'S RULE

The fundamental basis of Bayesian analysis is the Bayes's rule, which was first written by Thomas Bayes (1701–1761), to describe the relationship between marginal and conditional probabilities. Specifically,

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

where A and B are events and $P(B) \neq 0$. Also, $P(A|B)$ is the conditional probability of event A occurring given that B is true, and $P(B|A)$ is the probability of event B occurring given that A is true. $P(A)$ and $P(B)$ are the probabilities of observing A and B, respectively.

To make the interpretation of Bayes's rule more intuitive, we will start with an example. As we know, diagnostic tests are almost never perfectly accurate. A good test is supposed to have both high sensitivity (also called true positive rate; test result positive given disease present) and high specificity (also called true negative rate; test result negative given disease free). While sensitivity and specificity tell how good the test results are given the disease status, they do not directly tell the probability that a subject has the disease, given the test results. This is a situation where the Bayes's rule can be applied. Specifically, let $P(A)$ be the probability that a randomly chosen subject has a specific disease in a specific population (disease prevalence), and $P(B|A)$ be sensitivity of the test, and $P(A|B)$, called positive predictive value, is the probability that subjects with a positive test result truly have the disease, which is what we are interested in. Note that $P(B)$ is the probability of having a positive test result among subjects in this population, which equals to $P(A)P(B|A) + P(A^c)P(B|A^c)$, where $P(A^c) = 1 - P(A)$ is the probability of being disease free among the subjects, and $P(B|A^c)$ is 1-specificity.

Therefore,

$$P(A|B) = \frac{P(B|A)P(A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

$$= \frac{\text{sensitivity} \times P(A)}{P(A) \times \text{sensitivity} + [1 - P(A)](1 - \text{specificity})}$$

Note that $P(A)$ is the disease prevalence, prior to taking a diagnostic test, which can be determined from existing data or expert opinion. This probability can be updated/ revised to $P(A|B)$ after incorporating the test result into the calculation.

Numerically, suppose that the sensitivity and specificity of a test are both 0.95, and assume that the prevalence for a disease is 30% for subjects 50–60 years old, then the probability that a subject has the disease is 89% if tested positive. Meanwhile, assuming that the prevalence of the disease is 1% among subjects in their 20s, and then this probability becomes 16%. The prior information $P(A)$ has a big role in Bayesian data analysis.

Bayes’s rule can also be expressed in terms of probability distributions:

$$f(\theta|data) = \frac{f(data|\theta)f(\theta)}{f(data)},$$

where $f(\theta)$ is the prior distribution of the parameter θ , $f(data|\theta)$ is the sampling density for the data, given the parameter θ , $f(data)$ is the marginal distribution of the data, and $f(\theta|data)$ is the posterior distribution of the parameter θ . We will not present the details of how the posterior distribution is calculated—comparing with the likelihood function that the frequentists use, the incorporation of the prior distribution $f(\theta)$ makes Bayesian analysis more computationally challenging.

THE PRIOR DISTRIBUTION

It is critical to choose an appropriate prior distribution in a Bayesian analysis. If data from past studies/ experiments are available, then the prior distribution can be inferred from those. Other times, a prior can be determined more subjectively by experts in the field. If little is known about the parameter to be estimated, a non-informative prior is preferable.

NON-INFORMATIVE PRIOR

As an attempt to avoid subjectivity, non-informative priors are often used, even when prior information/ opinion is available. For example, a standard uniform distribution can be used as the prior for the proportion θ parameter of a binomial model. As we know, the probability density of any value in the range of 0 to 1 is the same for a uniform distribution (Figure 1; horizontal line in red). However, although with equal probability for all these values, a uniform prior is not completely non-informative; for example, the mean of a standard uniform distribution is 0.5, which is informative. Jeffrey’s prior $f(\theta) \sim \frac{1}{\sqrt{\theta(1-\theta)}}$, which is based on the Fisher information matrix, is more widely used as a non-informative prior (Figure 1; black curve). In most situations, the posterior distributions are quite similar when using either a uniform or a Jeffrey’s prior. However, under certain circumstances, the posterior distributions can differ substantially.

Suppose that we wanted to estimate in-hospital mortality rate at a regional hospital. Data from all the patients eligible for the study were collected, and the in-hospital mortality rate can be estimated by dividing the number of patients expired at discharge by the total number of patients included. For example, if the total number of patients included in the study was $n = 100$, and $k = 95$ were alive at discharge, then

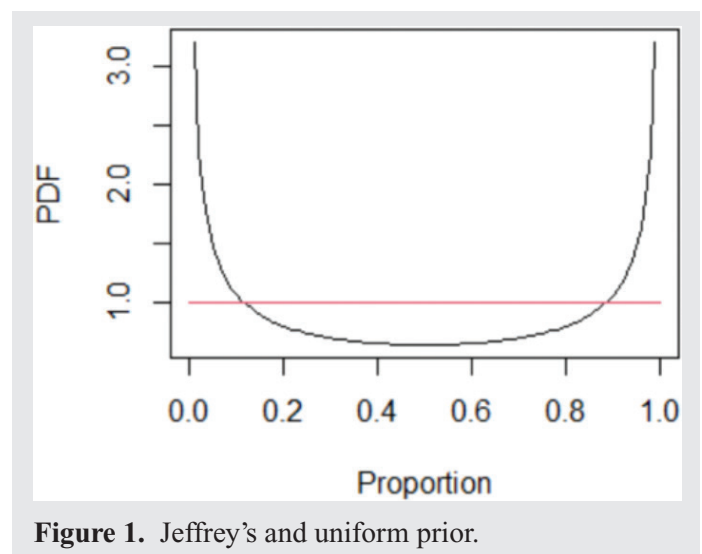


Figure 1. Jeffrey’s and uniform prior.

with a uniform prior $f(\theta) = 1$, the posterior distribution $f(\theta|data) \sim \text{Beta}(k + 1, n - k + 1)$. If a Jeffrey's prior is used, then the posterior distribution is $\text{Beta}(k + 0.5, n - k + 0.5)$. Note that if k is not too large or too small, then the means of the two posterior distributions are very close to each other. Numerically, the mortality rate estimates based on the frequentist's method is 5%, and the posterior mean by using the uniform and the Jeffrey's prior is 5.9% and 5.4%, respectively. However, if all 100 patients were alive at discharge, then the estimate based on the frequentist's method is 0%, and the posterior mean by using the uniform and the Jeffrey's prior is 1% and 0.5%, respectively. Note that 1% is twice as large as 0.5%.

INFORMATIVE PRIOR

Sometimes scientific information is available for determining the prior distribution. For example, the mortality rate of people with type I diabetes was 627 per 100,000 person-years, with a 95% confidence interval of 532–728. The systolic blood pressure is 123.5 ± 11.5 mm Hg during the day for certain healthy adults. If such information is available, it is preferable to use informative prior than a non-informative prior to gain better parameter estimations, especially for studies with a small sample size. On the other hand, an informative prior should be used with caution to avoid potential subjective bias.

As has already been mentioned, due to the incorporation of prior distribution, the computation of Bayesian posterior distribution can be challenging. For this reason, conjugate priors are widely used in practice due to their appealing computational properties.

CONJUGATE PRIOR

For some likelihood functions $f(data|\theta)$, if the posterior distributions $f(\theta|data)$ are in the same probability distribution family as the prior probability distribution $f(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function. For example, a Beta distribution is a conjugate prior to Binomial likelihood. Because the posterior and the prior are in the same distribution family, the posterior will also be a Beta

distribution. And based on well documented proof, the posterior distribution parameters can be obtained by simply adding the numbers of two potential outcomes, e.g., alive and expired, to the existing parameters of the prior distribution, respectively. Therefore, the computation becomes very straightforward. Note that if the likelihood function belongs to the exponential family, then a non-trivial conjugate prior exists. This is a convenient fact, because exponential family of distributions are commonly used in data modeling.

In situations where conjugate priors are not available or a specific distribution is more suitable, then methods, such as Markov Chain Monte Carlo (MCMC) simulation can be used to approximate the posterior distribution. A number of platforms can be used for performing this analysis, including Stan (<https://mc-stan.org/>), and BUGS (<http://www.openbugs.net/w/FrontPage>).

BAYESIAN HYPOTHESIS TESTING

In Bayesian, various summaries for the posterior model parameters can be summarized, including point estimates, such as posterior means, medians, percentiles, and interval estimates known as credible intervals. There are also different approaches for hypothesis testing. For example, the maximum a posteriori (MAP) test compares the posterior probabilities of two hypotheses and accepts the hypothesis with the higher posterior probability. As an alternative, Bayes's factor, which can be interpreted as the weight of evidence provided by a set of data, is also widely used. In general, a Bayes's factor between 1 and 3 is considered as weak evidence, between 3 and 20 as positive evidence, between 20 and 150 as strong evidence, and greater than 150 as very strong evidence. We will not cover the details of these hypothesis testing methods in this article.

In summary, Bayesian analysis is a method of statistical inference that combines prior information about a parameter with additional information from data to obtain an updated parameter distribution. Non-informative priors are more often used than informative priors unless there is solid prior evidence on the distribution of the parameters of interest. Conjugate

priors are computationally appealing and thus widely used; otherwise, methods such as MCMC simulation are needed to approximate the posterior distribution.

Keywords: prior, posterior, Bayesian analysis, probability

Article citation: Yang S, Berdine G. Bayesian data analysis. *The Southwest Respiratory and Critical Care Chronicles* 2020;8(36):74–77

From: Department of Biostatistics (SY), Pennington Biomedical Research Center, Baton Rouge, LA; Department of Internal Medicine (GB), Texas Tech University Health Sciences Center, Lubbock, TX

Submitted: 10/9/2020

Accepted: 10/11/2020

Conflicts of interest: none

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

REFERENCES

1. Abramovich F, Angelini C. Bayesian maximum a posteriori multiple testing procedure. *Indian J Statistics* 2006;68(3): 436–460.
2. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine* 2005;68:145–163.
3. Faulkenberry TJ. A simple method for teaching bayesian hypothesis testing in the brain and behavioral sciences. *J Undergrad Neurosci Educ* 2018;16(2):A126–A130.
4. Gelman A. Prior Distribution. *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd, Chichester, 2002. pp. 1634–1637.
5. Jaynes E T. Highly Informative Priors, *Bayesian Statistics 2*, Elsevier Science Publishers B.V., 1985. pp. 329–360.
6. Zhu M, L, AY. (2004) The counter-intuitive non-informative prior for the Bernoulli family. *J Statistics Education* 2004; 12(2):1–10.