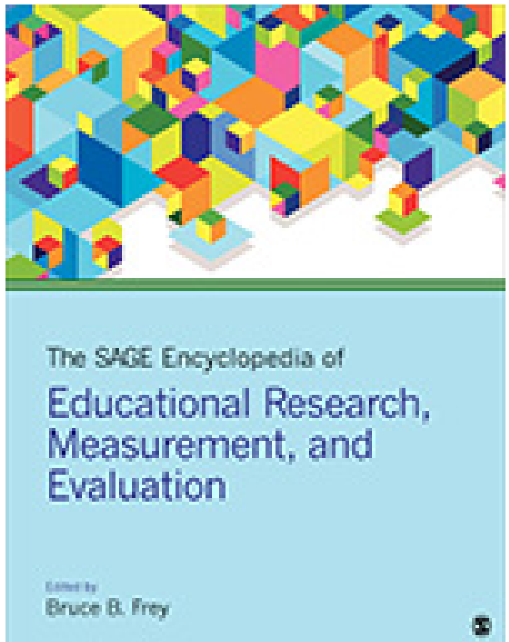


Sage Research Methods

The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation

Author: Meghan K. Cain, Zhiyong Zhang
Pub. Date: 2018
Product: Sage Research Methods
DOI: https://doi.org/10.4135/9781506326139
Methods: Educational research, Measurement

Disciplines: Education

Access Date: March 28, 2025

Publisher: SAGE Publications, Inc.

City: Thousand Oaks,

Online ISBN: 9781506326139

In Bayesian analysis, the *posterior distribution*, or *posterior*, is the distribution of a set of unknown parameters, latent variables, or otherwise missing variables of interest, conditional on the current data. The posterior distribution uses the current data to update previous knowledge, called a *prior*, about that parameter. A posterior distribution, $p(\theta|\boldsymbol{x})$, is derived using Bayes's theorem

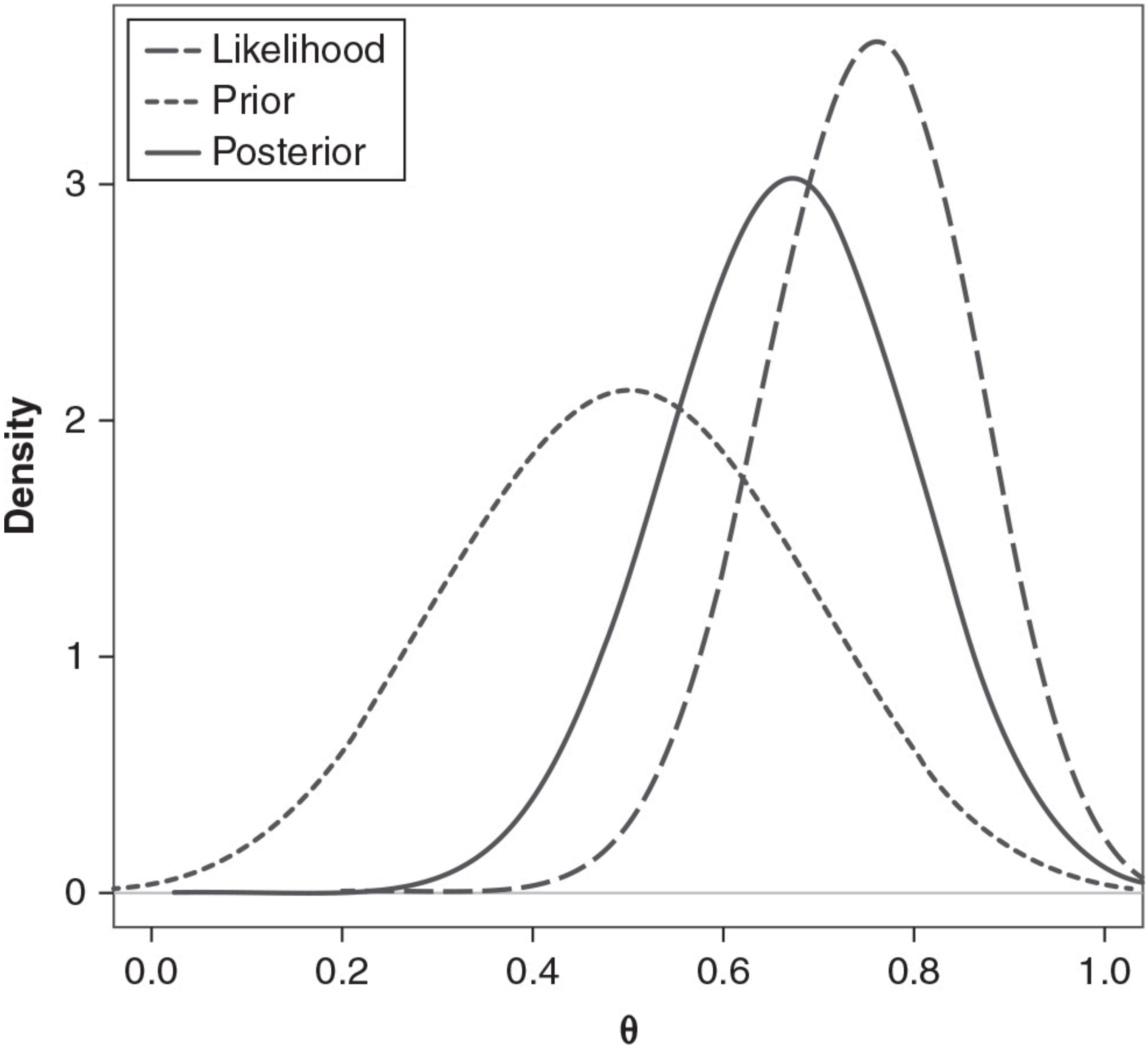
$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)} = \frac{p(x \mid \theta)p(\theta)}{\int p(x \mid \theta)p(\theta)d\theta},$$

where θ is the unknown parameter(s) and $\boldsymbol{x}$ is the current data. The probability of the data given the parameter $p(\boldsymbol{x}|\theta)$ is the likelihood $L(\theta|\boldsymbol{x})$. The prior distribution, $p(\theta)$, is user specified to represent prior knowledge about the unknown parameter(s). The last piece of Bayes's theorem, the marginal distribution of data, $p(\boldsymbol{x})$, is computed using the likelihood and the prior. The distribution of the posterior is determined by the distributions of the likelihood and the prior and scaled by the marginal distribution of the data. Therefore, the posterior can be represented as

$$\text{Posterior distribution} \propto \text{Likelihood} \times \text{Prior distribution},$$

where ∝ means "proportional to." The relationship between the posterior, the prior, and the likelihood is shown in Figure 1.

Figure 1 The likelihood and the prior determine the posterior distribution



The prior distribution is *conjugate* to the likelihood if the resulting posterior distribution has the same form as the prior distribution. The mean and variance of the posterior distribution are also determined by these two distributions. In certain situations, the posterior mean is a weighted average of the mean of the data and the prior, using the precision of each as weight. The precision, the reciprocity of variance, of the posterior is a function of the precision of the data and the prior. Thus, when a researcher is more confident in a prior, it is given more weight by specifying a smaller variance for the prior distribution.

The posterior distribution can be analytically computed by integration or it can be approximated using a Markov chain Monte Carlo algorithm. With increases in computational power, the latter is often the easier option, and the Markov chain Monte Carlo method is what is used in software such as WinBUGS and Mplus. A commonly used Markov chain Monte Carlo method is *Gibbs sampling*, which recursively generates random numbers from the conditional posterior distribution for each parameter in turn, conditional on the current values of all other parameters.

The resulting posterior distribution is what is used to make inferences about the model. The mean, median, or mode of the posterior distribution can be used as a point estimate, much like a maximum likelihood estimate (MLE) can be used within the frequentist framework. If the prior $p(\theta)$ is a constant, the mode of the posterior, if it exists, is equivalent to the MLE. *Credible intervals* can also be constructed using the posterior distribution. These are analogous to confidence intervals in the frequentist framework but differ in theory and interpretation. Credible intervals provide the $(1 - \alpha)\%$ probability that a parameter lies between a lower and upper bound. Thus, credible intervals assume the parameter is random and the lower and upper bounds are fixed, whereas confidence intervals assume the opposite.

Example

To illustrate, let's say Researcher F finds a coin in his attic. He wants to know whether the coin is fair, so he flips it 20 times and records 15 heads landings. He is a frequentist, so he would like to find an MLE of the probability of the coin landing on heads. First, he computes the likelihood using a binomial distribution,

$$L(\theta \mid x) \sim Bin(n, p) = \theta^{k}(1 - \theta)^{n-k},$$

where n is the number of tosses, p is the probability of landing on heads, and k is the number of heads landings ($n \times p$). This results in an MLE of $\hat{\theta} = 0.75$. Using a binomial test, he concludes that the coin is not fair, $z = 2.236$, $p < .05$.

The next day, Researcher F tells his colleague, Researcher B, about the coin he found that lands on heads significantly more often than it lands on tails. In response, Researcher B says "Wait a minute, shouldn't we take into account that most coins are fair?" She would like to use Bayesian analysis to investigate further by choosing a prior that is centered at 50% with small variance because she knows that most coins land on each side with equal probability. Because the β distribution is conjugate to the binomial likelihood function, she would like to use a prior with the following form:

$$p(\theta) \sim \beta(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1 - \theta)^{b-1}.$$

She chooses $p(\theta) \sim \beta(5,5)$, which is a symmetrical distribution with mean = 0.5 and variance = 0.007. Using this prior with the binomial likelihood, she calculates the posterior distribution for the proportion of heads,

$$p(\theta\,|x) = \frac{L(\theta\,|x)p(\theta)}{\int_0^1 L(\theta\,|x)p(\theta)d\theta}$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a+k)\Gamma(b+n-k)}\theta^{a+k-1}(1-\theta)^{b+n-k+1}.$$

Because she used conjugate distributions, she gets a β distribution with parameters $a + k$ and $b + n - k$. Plugging in a and b from the prior and n and k from the data, she obtains β(20,10) as the posterior. Using the mean of the posterior, the new point estimate of the fairness of the coin is $\hat{\theta} = 0.67$, slightly less extreme than it was for Researcher F. The credible interval is [0.49,0.82], and so we are not quite sure whether the coin lands on heads more than tails.

If Researcher F had instead flipped the coin 100 times and gotten 75 heads, using the same prior would have yielded the posterior β(80,30) with point estimate $\hat{\theta} = 0.73$. This is because getting 75 heads out of 100 tosses is much stronger evidence against the coin being fair than getting 15 heads out of 20 tosses. In this case, the data are given more weight in calculating the posterior, so our point estimate is closer to that of the data. The credible interval in this case would be [0.64,0.81], indicating that there is a 95% probability that the coin is not fair.

In sum, the posterior distribution is proportional to the product of the likelihood and the prior in Bayesian analysis, and it can be used to make inferences about model parameters. Any inference made with the posterior is based on prior information about a model that has been updated after collecting new data. Inferences can be made using the mean, median, or mode of the posterior as point estimates or through credible intervals, among other methods.

See also Bayes's Theorem; Bayesian Statistics; Binomial Test; Markov Chain Monte Carlo Methods; Prior Distribution

Meghan K. Cain Zhiyong Zhang

Further Readings

Gelman, A., Carlin, J. B., & Stern, H. S. (2003). *Bayesian data analysis*. London, UK: Chapman Hall.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Cambridge, MA: Academic Press.

Lee, P. M. (2012). *Bayesian statistics: An introduction* (4th ed.). Hoboken, NJ: Wiley.

https://doi.org/10.4135/9781506326139