3

Bayesian Inference

The posterior distribution summarizes all we believe about the parameter(s) after analyzing the data. It incorporates both our prior belief through the prior distribution and the information from the data through the likelihood function. In the Bayesian framework, the posterior is our entire inference about the parameter given the data. However, in the frequentist framework, there are several different types of inference: point estimation, interval estimation, and hypothesis testing. They are deeply embedded in statistical practice. Using the Bayesian approach, we can do each of these inferences based on the posterior distribution. In fact, under the Bayesian approach these inferences are more straightforward than under the frequentist approach. This is because, under the frequentist approach, the parameter is fixed but unknown and the only source of randomness is the distribution of the sample data given the unknown parameter which is the sampling distribution of the data. We are evaluating the parameter in its dimension based on a probability distribution in the data dimension. Under the Bayesian approach, we evaluate the parameter using a probability distribution in the parameter dimension, the posterior distribution. Bolstad (2007) gives a thorough discussion of the similarities and differences between Bayesian and frequentist inferences.

3.1 BAYESIAN INFERENCE FROM THE NUMERICAL POSTERIOR

In some cases we have a formula for the exact posterior. In other cases we only know the shape of the posterior using Bayes' theorem. The *posterior* is proportional to *prior* times *likelihood*. In those cases we can find the posterior density numerically by dividing through by the scale factor needed to make the integral of the posterior over its whole range of values equal to one. This scale factor is found by integrating the prior times likelihood over the whole range of parameter values. Thus the posterior is given by

$$g(\theta|y_1,\ldots,y_n) = \frac{g(\theta)\, f(y_1,\ldots,y_n|\theta)}{\int g(\theta)\, f(y_1,\ldots,y_n|\theta)\, d\theta} \tag{3.1}$$

where we have performed numerical integration in the denominator. The posterior density summarizes everything we know about the parameter given the data. The inferences are summaries of the posterior taken to answer particular questions.

Bayesian Point Estimation

The first type of inference is where a single statistic is calculated from the sample data and is used to estimate the unknown parameter. From the Bayesian perspective, point estimation is choosing a value to summarize the posterior distribution. The most important summary number of a distribution is its location. The posterior mean and the posterior median are good measures of location and hence would be good Bayesian estimators of the parameter. Generally we will use the posterior mean as our Bayesian estimator since it minimizes the posterior mean squared error

$$PMS(\hat{\theta}) = \int (\theta - \hat{\theta})^2\, g(\theta|y_1,\ldots,y_n) d\theta\,. \tag{3.2}$$

The posterior mean is the first moment (or balance point) of the posterior distribution. We find it by

$$\hat{\theta} = \int_{-\infty}^{\infty} \theta g(\theta|y_1,\ldots,y_n)\, d\theta. \tag{3.3}$$

The posterior median could also be used as a Bayesian estimator since it minimizes the posterior mean absolute deviation

$$PMAD(\hat{\theta}) = \int |\theta - \hat{\theta}|\, g(\theta|y_1,\ldots,y_n) d\theta\,. \tag{3.4}$$

We find the posterior median by numerically integrating the posterior. The posterior median $\tilde{\theta}$ is found by solving

$$.5 \;=\; \int_{-\infty}^{\tilde{\theta}} g(\theta|y_1,\ldots,y_n)\, d\theta\,. \tag{3.5}$$

The posterior median has half the posterior probability below it, and half the posterior probability above it. The posterior median can also be found from the numerical cumulative distribution function (CDF) of the posterior $G(\theta|y_1,\ldots,y_n) = \int_{-\infty}^{\theta} g(\theta|y_1,\ldots,y_n)\, d\theta$. The posterior median is the value $\tilde{\theta}$ that gives the numerical CDF value equal to .5.

Example 4 *Suppose the unscaled posterior density has shape given by*

$$g(\theta|y_1,\ldots,y_n) = .8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2}\,.$$

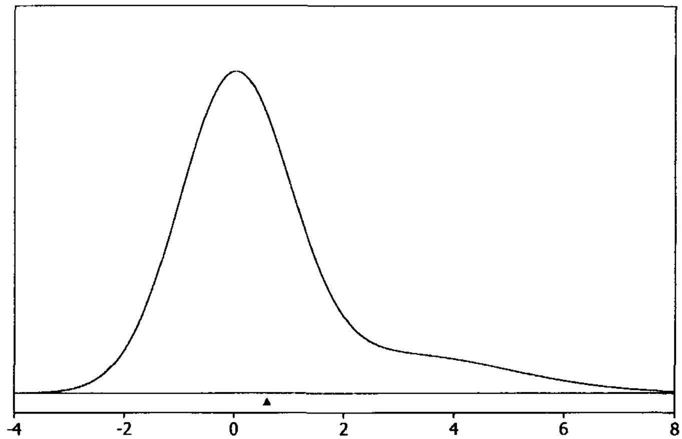

Figure 3.1 The posterior mean is the first moment or balance point of the posterior distribution.

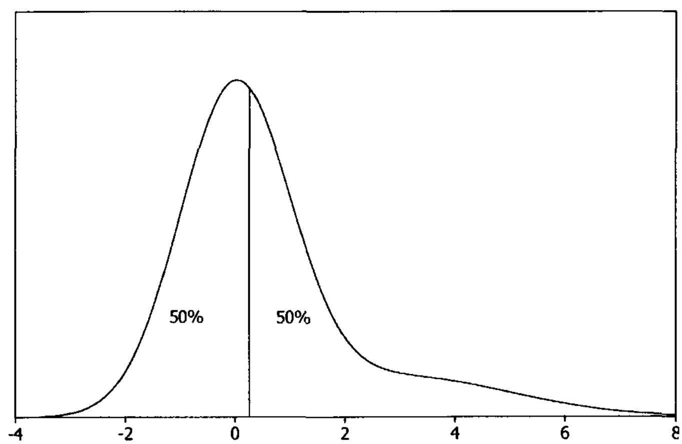

Figure 3.2 The posterior median is the point of the posterior distribution which has equal area above and below.

This is a mixture of a normal$(0, 1^2)$ *and a normal*$(3, 2^2)$*. We find the integral of the unscaled target by integrating it over the whole range*

$$\int_{-\infty}^{\infty} .8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} \, d\theta = 2.50633\,.$$

Thus the numerical density is given by

$$g(\theta|y_1, \ldots, y_n) = .39899 \times \left(.8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} \right).$$

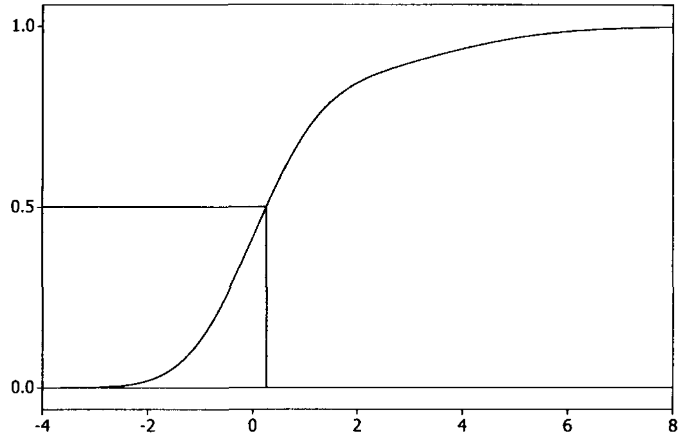

Figure 3.3 The posterior median is the θ value that has numerical CDF value equal to .5.

We find the posterior mean, $\hat{\theta}$, by integrating θ times the numerical posterior density.

$$\begin{aligned}\hat{\theta} &= E(\theta|y_1,\ldots,y_n)\\ &= \int_{-\infty}^{\infty} \theta \times .39899 \times \left(.8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2}\right) d\theta\\ &= .5999\end{aligned}$$

The target density with the posterior mean as its balance point is shown in Figure 3.1. We find the posterior median $\tilde{\theta}$ from the numerical posterior density. It is the solution of

$$\int_{-\infty}^{\tilde{\theta}} .39899 \times \left(.8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2}\right) d\theta = .5000$$

The posterior median is found to be $\tilde{\theta} = .2627$. The target density with the posterior median as the point that has equal area above and below is shown in Figure 3.2. The posterior median is shown on the graph of the CDF in Figure 3.3.

The loss-function is the cost for estimating with estimator $\hat{\theta}$ when the true parameter value is θ. The posterior mean is the Bayesian estimator that minimizes the squared-error loss-function, while the posterior median is the Bayesian estimator that minimizes the absolute value loss-function. One of the strengths of Bayesian statistics, is that we could decide on any particular loss function, and find the estimator that minimizes it. This is covered in the field of statistical decision theory. We will not pursue this topic further in this book. Readers are referred to Berger (1980) and DeGroot (1970).

Bayesian Interval Estimation

The second type of inference is where we find an interval of possible values that has a specific probability of containing the true parameter value. In the Bayesian approach, we have the posterior distribution of the parameter given the data. Hence we can calculate an interval that has the specified posterior probability of containing the random parameter θ. These are called *credible intervals*.

When we want to find a $(1-\alpha)\times 100\%$ credible interval for θ from the posterior we are looking for an interval (θ_l, θ_u) such that the posterior probability

$$\begin{aligned}(1-\alpha) &= P(\theta_l < \theta < \theta_u) \\ &= \int_{\theta_l}^{\theta_u} g(\theta|y_1,\ldots,y_n)\, d\theta\,.\end{aligned}$$

There are many possible intervals that have the required coverage probability. The shortest interval (θ_l, θ_u) with the required probability will have equal density values. That is $g(l|y_1,\ldots,y_n) = g(u|y_1,\ldots,y_n)$. However, often it is easier to find the interval (θ_l, θ_u) that has equal tail areas. This will be the interval (θ_l, θ_u) where we find θ_l and θ_u by

$$\int_{-\infty}^{\theta_l} g(\theta|y_1,\ldots,y_n)\, d\theta = \frac{\alpha}{2} \quad \text{and} \quad \int_{\theta_u}^{\infty} g(\theta|y_1,\ldots,y_n)\, d\theta = \frac{\alpha}{2}$$

respectively.

Example 4 (continued) *We find the equal tail area 95% credible interval using the numerical posterior density. The lower limit is the solution of*

$$\int_{-\infty}^{\theta_l} .39899 \times \left(.8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} \right) d\theta \; = .025$$

and the upper limit is the solution of

$$\int_{\theta_u}^{\infty} .39899 \times \left(.8 \times e^{-\frac{1}{2}(\theta)^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} \right) d\theta \; = .025\,.$$

We find the 95% credible interval is $(-1.888,\ 5.299)$. *The density with the credible interval is shown in Figure 3.4.*

Bayesian Testing of a One-Sided Hypothesis

The third type of inference that we want to do is to decide whether the parameter value for the new treatment is greater than the historical value of the parameter for the standard treatment. When we want to decide whether or not the treatment effect makes the parameter greater than the historical value it had for the standard treatment, we set this up as a one-sided hypothesis test. The historical value of the parameter for