Locally uniform, proper prior for a location parameter

Suppose we wish to estimate the mean, θ, of a distribution.

It is rare that the experimenter wouldn't at least know the range of possible values for the observations. Suppose he/she is more or less certain that the mean of the distribution is between a and b.

As a noninformative prior for θ we could use any density that is (i) constant over (a, b) and (ii) goes quickly to 0 outside (a, b).

Bayesian Inference

As noted previously, the big three of classical inference are point estimation, interval estimation and hypothesis testing.

Point estimation

We discussed point estimation in the context of a binomial experiment. The two approaches discussed are general:

1. Use the *mode* of the posterior.

2. Use Bayes principle. Choose a loss function and use the corresponding Bayes estimate. Squared error loss is particularly popular in statistics. The Bayes estimate in this case is the *mean* of the posterior.

Bayesian credible regions

Bayesian credible regions are the analogs of confidence regions in classical statistics. However, unlike the classical regions, they have a proper probability interpretation.

<u>*Definition*</u> A $100(1-\alpha)\%$ credible region for $\boldsymbol{\theta}$ is a subset C of Θ such that

$$P(\boldsymbol{\theta} \in C | \boldsymbol{Y} = \boldsymbol{y}) \geq 1 - \alpha.$$

Two principles are often used in constructing C.

- The volume of C should be as small as possible.

- The posterior density should be greater for every $\boldsymbol{\theta} \in C$ than it is for any $\boldsymbol{\theta} \notin C$.

It turns out that these two criteria are equivalent.

Definition The $100(1-\alpha)\%$ *highest posterior density* (HPD) region for $\boldsymbol{\theta}$ is a subset C of Θ such that

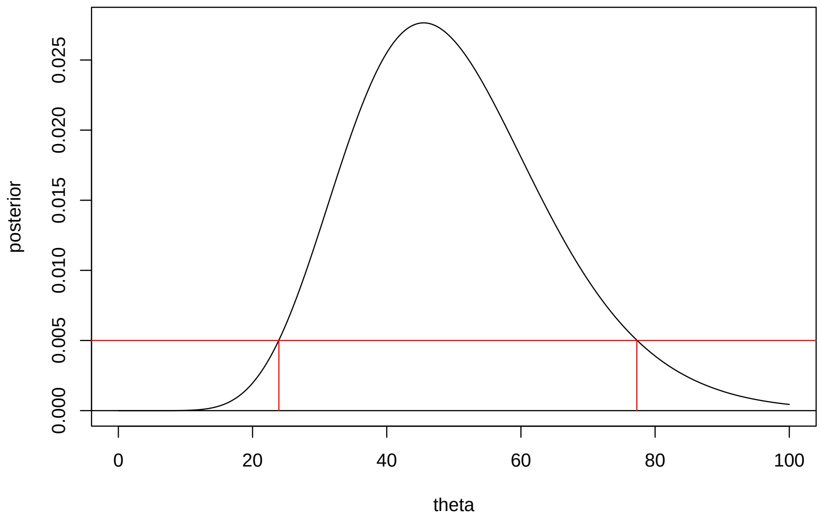
$$C = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}|\boldsymbol{y}) \geq k_\alpha\},$$

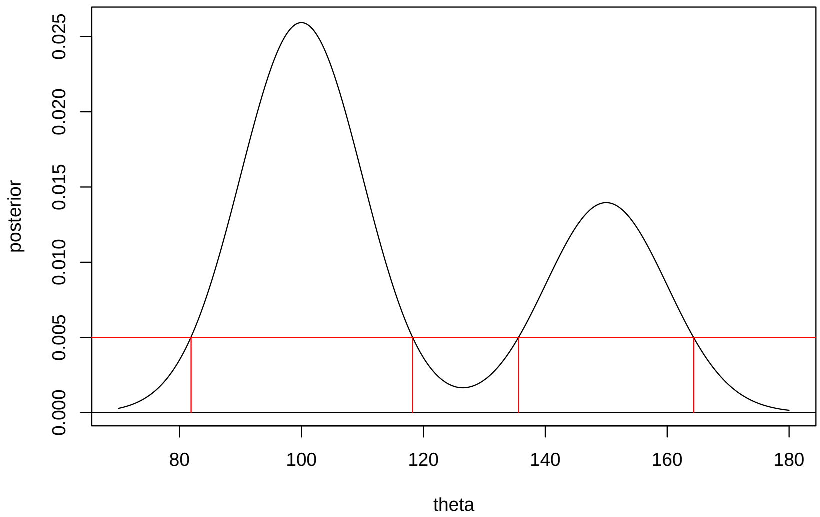
where k_α is the largest constant for which

$$P(\boldsymbol{\theta} \in C|\boldsymbol{Y} = \boldsymbol{y}) \geq 1 - \alpha.$$

An HPD region has the smallest volume of all regions with a given probability content.

Finding an HPD region





In each graph, the set of θ values at which the density exceeds the horizontal line at 0.005 forms an HPD region.

Bayesian hypothesis testing

Let $\Theta = \Theta_0 \cup \Theta_1$, where $\Theta_0 \cap \Theta_1 = \emptyset$. Suppose we want to test

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \in \Theta_1.$$

Define

$$\alpha_0 = P(\boldsymbol{\theta} \in \Theta_0 | \boldsymbol{Y} = \boldsymbol{y}) \quad \text{and} \quad \alpha_1 = 1 - \alpha_0.$$

The *posterior odds ratio* is α_0/α_1. In Bayesian hypothesis testing, one simply uses α_0/α_1 to assess the relative plausibility of H_0 and H_1.

Note that α_0/α_1 is a measure of *final precision*. Classical hypothesis tests are based on measures of *initial precision*, i.e., power and probability of type I error.

Let π_0 and π_1 denote the prior probabilities of H_0 and H_1, respectively. Another approach to testing is to use the *Bayes factor*:

$$\begin{aligned} \text{Bayes factor} &= \frac{\text{posterior odds ratio}}{\text{prior odds ratio}} \\ &= \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} \\ &= \frac{\alpha_0 \pi_1}{\pi_0 \alpha_1} = B. \end{aligned}$$

$B < 1 \Rightarrow$ degree of belief in H_0 has decreased.

It's of interest to consider testing simple hypotheses, i.e.,

$$\Theta_0 = \{\boldsymbol{\theta}_0\} \quad \text{and} \quad \Theta_1 = \{\boldsymbol{\theta}_1\}.$$

This is the setting where the Neyman-Pearson lemma provides a most powerful size α test.

In the simple vs. simple case,

$$\alpha_0 = \frac{\pi_0 f(\boldsymbol{y}|\boldsymbol{\theta}_0)}{\pi_0 f(\boldsymbol{y}|\boldsymbol{\theta}_0) + \pi_1 f(\boldsymbol{y}|\boldsymbol{\theta}_1)},$$

the posterior odds ratio is

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\boldsymbol{y}|\boldsymbol{\theta}_0)}{\pi_1 f(\boldsymbol{y}|\boldsymbol{\theta}_1)},$$

and the Bayes factor is

$$B = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}_0)}{f(\boldsymbol{y}|\boldsymbol{\theta}_1)}.$$

Two interesting facts about this Bayes factor:

- B is the likelihood ratio statistic upon which the most powerful test is based.

- B is free of any prior probabilities.

Unfortunately, the second fact doesn't generalize to more complicated testing situations.

However, it has led to the search for a "stable" Bayes factor, i.e., one which is at least approximately free of prior probabilities in most situations. See Berger, *Statistical Decision Theory and Bayesian Analysis* for more details.

Normal Data with Unknown Mean and Known Variance

Suppose $Y = (Y_1, \ldots, Y_n)$, where $Y_1, \ldots, Y_n$ are i.i.d. as $N(\theta, \sigma^2)$, where θ is unknown and σ^2 is known. For a prior, we will use $\pi \equiv N(\mu_0, \sigma_0^2)$.

By taking σ_0^2 sufficiently large, this prior will be essentially noninformative.

The posterior is

$$
\begin{aligned}
\pi(\theta|\boldsymbol{y}) &\propto \exp\left(-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2\right) \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma_0^2}(\theta^2-2\mu_0\theta)\right) \\
&\quad \times \exp\left(-\frac{n}{2\sigma^2}(\theta^2-2\bar{y}\theta)\right).
\end{aligned}
$$

By completing the square in the exponent, we have

$$\pi(\theta|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma_n^2}(\theta - \mu_n)^2\right),$$

where

$$\mu_n = \frac{\bar{y} + \mu_0\left(\frac{\sigma^2}{n\sigma_0^2}\right)}{1 + \frac{\sigma^2}{n\sigma_0^2}}$$

and

$$\sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}.$$

It follows that the posterior is $N(\mu_n, \sigma_n^2)$.

Of course, the mean and mode of the normal distribution are one and the same, and hence both of these Bayesian point estimates of θ_0 are μ_n.

Recall that θ_0 denotes the true value of the unknown parameter.

We have

$$\mu_n = w_n \bar{y} + (1 - w_n)\mu_0,$$

where

$$w_n = \left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{-1},$$

and hence this Bayes estimate has the properties cited on p. 47N.

Note also that

$$\sigma_n^2 = w_n \cdot \frac{\sigma^2}{n}.$$

HPD region for θ_0

Because the normal density is unimodal and symmetric about its mean, a $100(1-\alpha)\%$ HPD region for θ_0 has the form

$$\mu_n \pm z_{\alpha/2}\sigma_n,$$

where z_p is the $(1-p)$th quantile of the standard normal distribution.

Regardless of how small n is, if we take σ_0 to be sufficiently big (meaning that the prior for θ is noninformative), then

$$\mu_n \approx \bar{y}, \quad \sigma_n \approx \frac{\sigma}{\sqrt{n}}$$

and the HPD region is approximately

$$\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

The interesting thing here is that the last HPD region has exactly the same form as the usual frequentist confidence interval in this setting.

The Jeffreys noninformative prior for the normal model under consideration is constant over the real line, and hence improper.

However, the corresponding posterior is $N(\bar{y}, \sigma^2/n)$, and thus proper. So, by using the improper prior, an HPD region is *precisely* the usual frequentist confidence interval.

Example 9 *Hypothesis testing*

$Y_1, \ldots, Y_{100}$ are a random sample from $N(\theta,$ $1000)$. Want to test

$$H_0 : \theta < 20 \quad \text{vs.} \quad H_1 : \theta \geq 20.$$

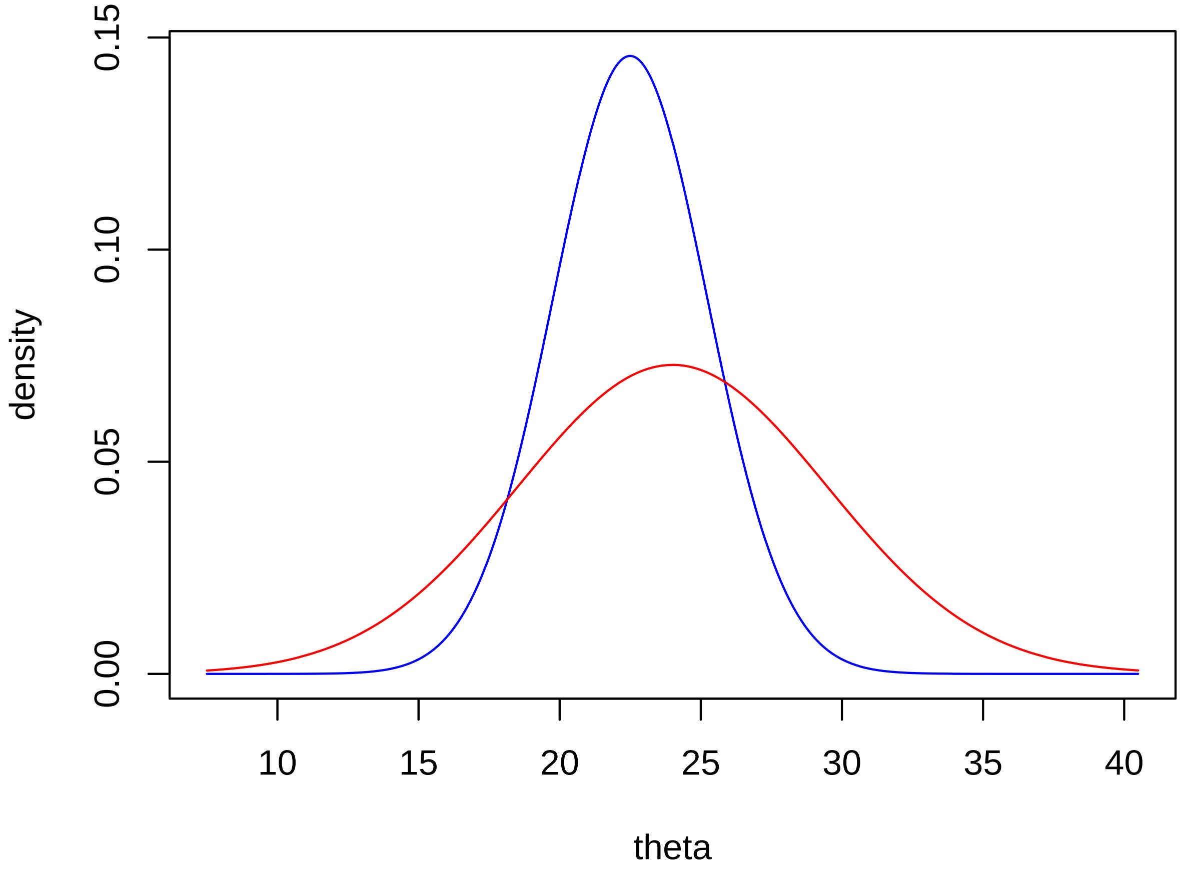
The prior for θ is $N(24, 30)$. This means

$$w_n = \left(1 + \frac{1000}{100(30)}\right)^{-1} = 3/4,$$

and hence the posterior is $N(6 + 3\bar{y}/4, 7.5)$. Using the prior, one may check that $\pi_0 = 0.233$ and $\pi_1 = 0.767$.

Suppose $\bar{y}$ is observed to be 22. Then the posterior is $N(22.5, 7.5)$.

Prior and posterior for Example 9



Prior: —— Posterior: ——

The posterior probability of H_1 is

$$\begin{aligned}\alpha_1 &= P(\theta \geq 20|\bar{Y} = 22) \\ &= P\left(Z \geq \frac{20 - 22.5}{\sqrt{7.5}}\right) \\ &= 0.819.\end{aligned}$$

So, $\alpha_0 = 0.181$ and the posterior odds ratio is

$$\frac{\alpha_0}{\alpha_1} = \frac{0.181}{0.819} = 0.221.$$

The Bayes factor is

$$B = \frac{0.221}{0.233/0.767} = 0.727.$$

The odds of H_0 being true have declined in light of the data.

As a point of interest, note that the classical P-value in this example is

$$P\left(Z > \frac{22 - 20}{\sqrt{10}}\right) = 0.264.$$