Example 10 *HPD region for exponential rate*

A company wants to obtain information about the distribution of lifetimes of a certain electronic component. It is assumed that the lifetime of a randomly selected component has density

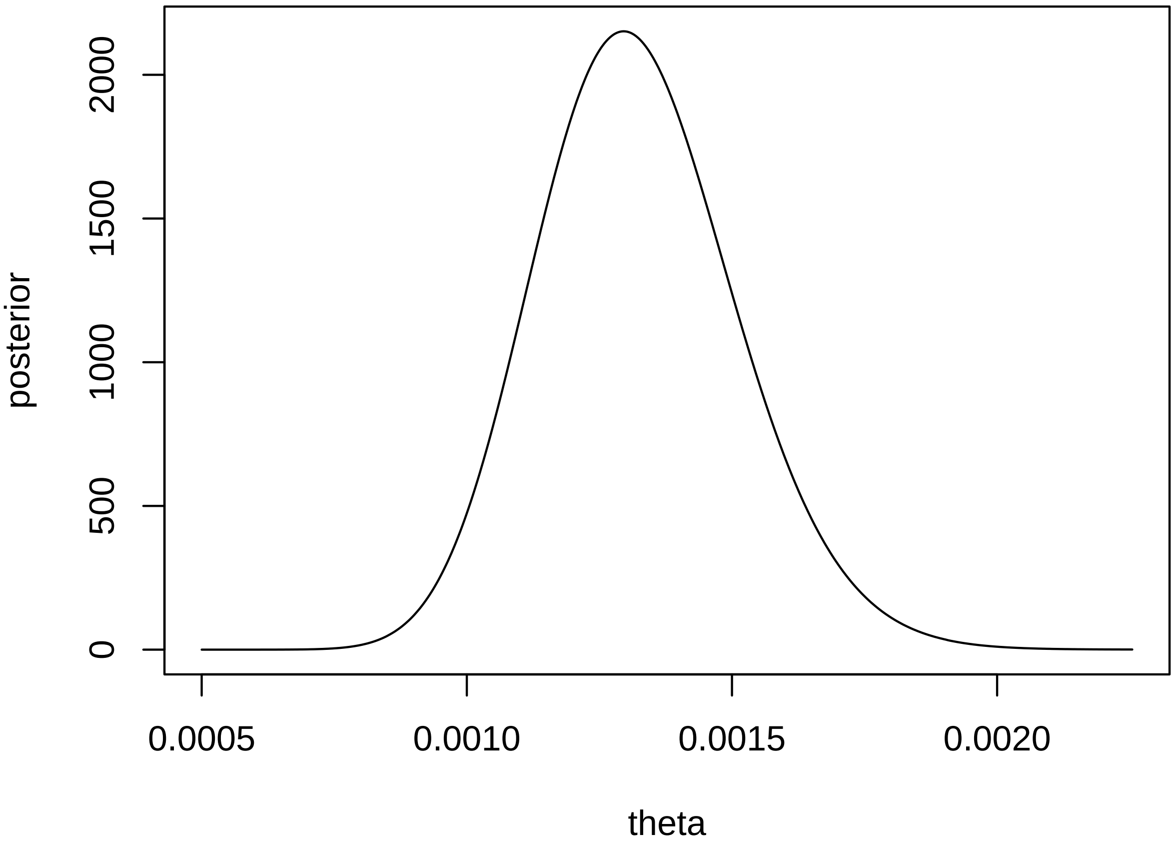
$$g(y|\theta) = \theta e^{-\theta y} I_{(0,\infty)}(y).$$

The company decides to use the noninformative (and improper) prior $\pi(\theta) = \theta^{-1}$.

Fifty components are put on test until they all fail. The average lifetime of the components was 756.3 hours.

The company wishes to find a 95% HPD region for θ.

The posterior for θ is Gamma$(50, 37815)$, which is pictured below.



The mean of this distribution is $1/756.3 = 0.001322$, which of course is the Bayes estimate of θ.

The standard deviation of the distribution is $\sqrt{50}/[50(756.3)] = 0.0001870$.

To find a 95% HPD region, one can use numerical approximation. The following algorithm was used:

- Choose an initial guess, $\tilde{k}$, for the value of $k_{.05}$ (See p. 71N.)

- Use Newton's method to solve the equation $\pi(\theta|\bar{y}) = \tilde{k}$ for its two solutions, call them $\theta_1 < \theta_2$.

- Evaluate $p = P(\theta < \theta_1|\bar{y}) + P(\theta > \theta_2|\bar{y})$.

- If p is sufficiently close to 0.05, then stop and use (θ_1, θ_2) as the HPD region. Otherwise, repeat the previous steps starting from a different guess $\tilde{k}$.

The previous algorithm was applied starting from an initial $\tilde{k}$ of 250. Initial guesses for the solutions of

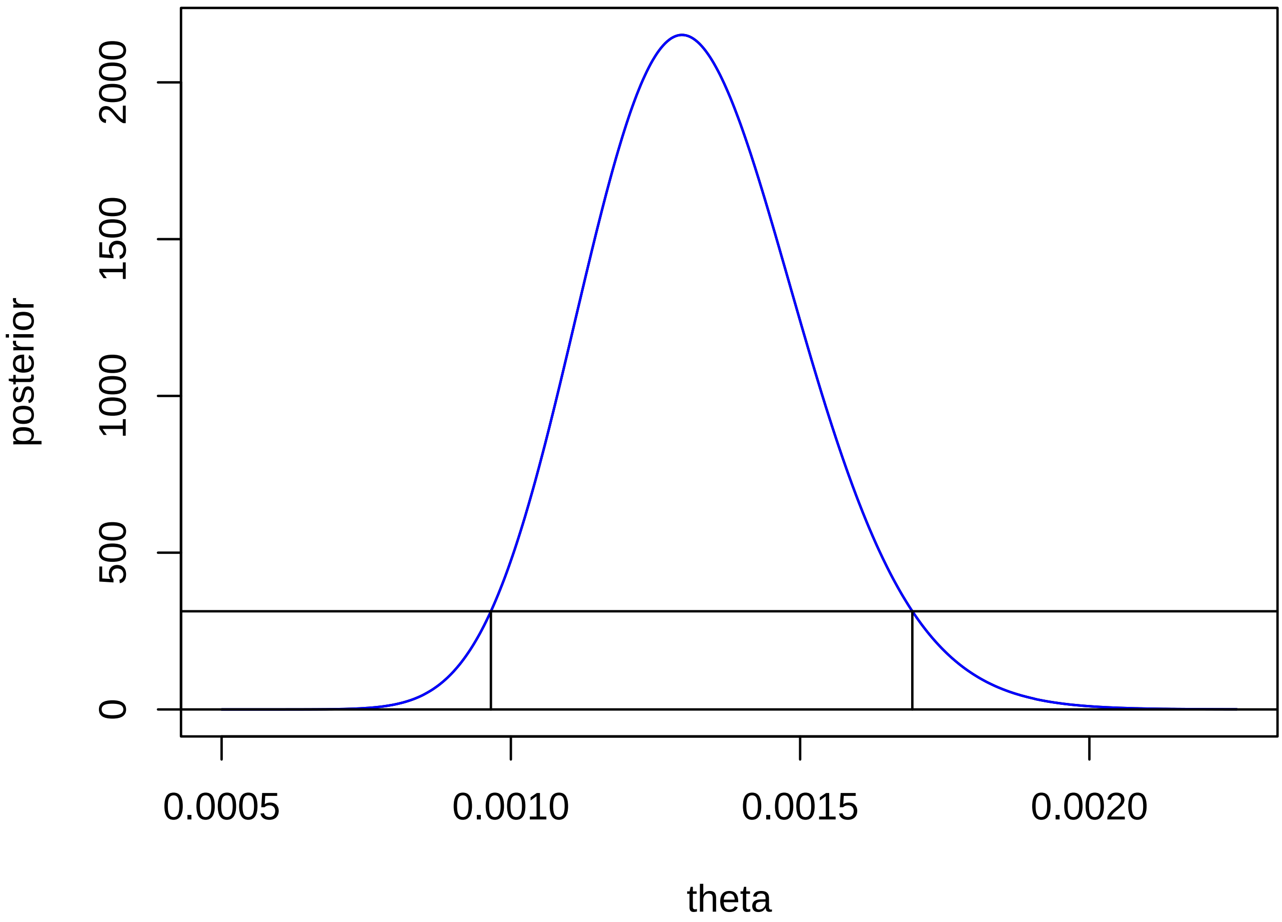
$$\pi(\theta|\bar{y}) = 250 \qquad (H)$$

were 0.001 and 0.0017.

The solutions to equation (H) were found to be 0.0009487 and 0.0017187, and the area between these two values is 0.9617. Since $1 - 0.9617 = 0.0383$ is smaller than 0.05, $k_{0.05}$ is bigger than 250.

Continuing to refine the choice of $\tilde{k}$, it was found that $k_{0.05} \approx 313.14$ and the 95% HPD interval is $(0.0009655, 0.0016940)$.

Illustration of 95% HPD region
for Example 10



At the course website you can find the *R* function used to obtain this region.

Another, simpler way, of approximating the HPD region is to use the fact that the posterior is approximately $N(1/756.3, [50(756.3)^2]^{-1})$.

So, the 95% HPD interval is approximately

$$1/756.3 \pm 1.96\left[\sqrt{50}(756.3)\right]^{-1} = \\ 0.0013222 \pm 0.0003665,$$

or $(0.0009557, 0.0016887)$. This is reasonably close to the "exact" HPD region on p. 96N.

Later in the course (Chapter 4) we will justify using normal approximations for posteriors when the sample is large.

Poisson Data

Suppose Y has a Poisson distribution with unknown mean θ.

$$f(y|\theta) = \frac{e^{-\theta}\theta^y}{y!} I_{\{0,1,\ldots\}}(y),$$

with parameter space $\Theta = (0, \infty)$.

This model could apply to a situation where one observes

- a single Poisson count, or
- n i.i.d. Poisson variables.

Why?

Conjugate family for Poisson data

It is easily verified that the Gamma(α, β) family is a conjugate family for the Poisson distribution. For a Gamma(α, β) prior, we have

$$\pi(\theta|y) \propto \theta^{y+\alpha-1} e^{-\theta(\beta+1)},$$

and hence the posterior is Gamma($y+\alpha, \beta+1$).

Jeffreys prior

The log-probability function is

$$\log f(y|\theta) = -\theta + y \log \theta - \log y!,$$

$$\frac{\partial}{\partial \theta} \log f(y|\theta) = -1 + \frac{y}{\theta}$$

and

$$\frac{\partial^2}{\partial \theta^2} \log f(y|\theta) = -\frac{y}{\theta^2},$$

It follows that

$$-E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right] = \frac{1}{\theta},$$

and so the Jeffreys noninformative prior is $\pi(\theta) \propto \theta^{-1/2}$, which is improper.

Interestingly, $\theta^{-1/2}$ is improper only because of its behavior at ∞. The function $\theta^{-1/2}$ is integrable on any finite interval $(0, a)$. This is different than in the the case of an exponential distribution, where Jeffreys prior is improper at both ∞ *and* 0.

The posterior corresponding to the Jeffreys prior is such that

$$\pi(\theta|y) \propto \theta^{y-1/2}e^{-\theta}$$

and hence is Gamma($y+1/2, 1$), which is proper regardless of the value of y.

For the Gamma(α, β) prior, the Bayes estimate corresponding to squared error loss is

$$\frac{y+\alpha}{\beta+1} = y \cdot \left(\frac{1}{\beta+1}\right) + \frac{\alpha}{\beta} \cdot \left(1 - \frac{1}{\beta+1}\right).$$

As usual, this is a linear combination of a frequentist estimate (in this case the MLE) and the prior mean.

The mode of the posterior is

$$\text{mode} = \begin{cases} \frac{(y+\alpha-1)}{(\beta+1)}, & y+\alpha > 1, \\ 0, & y+\alpha \leq 1. \end{cases}$$

For the Jeffreys prior, the very last result remains true if we take $\alpha = 1/2$ and $\beta = 0$, and the Bayes estimate wrt squared error loss is $y + 1/2$.

Testing a Point Null Hypothesis

Suppose we want to test

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0.$$

If θ is a continuous parameter, then continuous priors and posteriors assign probability 0 to θ_0 (and hence to H_0).

There are two ways around this problem:

- Change H_0 to $H_0 : \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$.

- Assign a prior probability of $\pi_0 > 0$ to θ_0, and then distribute the remaining probability $1 - \pi_0$ among the other values of θ.

In the second case, let G_1 be an absolutely continuous cdf over all of Θ, and let G_0 be the cdf that assigns probability 1 to θ_0. Then the prior cdf, G, is a mixture of G_0 and G_1:

$$G(\theta) = \pi_0 G_0(\theta) + (1 - \pi_0) G_1(\theta) \quad \forall\ \theta \in \Theta.$$

Letting g_1 be the density corresponding to G_1, the posterior probability of H_0 is

$$\alpha_0 = P(\theta_0 | \boldsymbol{Y} = \boldsymbol{y}) = \frac{\pi_0 f(\boldsymbol{y}|\theta_0)}{m(\boldsymbol{y})},$$

where

$$\begin{aligned} m(\boldsymbol{y}) &= \pi_0 f(\boldsymbol{y}|\theta_0) \\ &\quad + (1 - \pi_0) \int_\Theta f(\boldsymbol{y}|\theta) g_1(\theta)\, d\theta. \end{aligned}$$

Example 11 *Example 11, p. 150 of Berger*

$Y = (Y_1, \ldots, Y_n)$: $Y_1, \ldots, Y_n$ are a random sample from $N(\theta, \sigma^2)$, where θ is unknown and σ^2 known.

Suppose we want to test

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0,$$

and we take $\pi_0 = \pi_1 = 1/2$.

Let g_1 be $N(\theta_0, \sigma^2)$, and suppose

$$z = \frac{\bar{y} - \theta_0}{\sigma/\sqrt{n}} = 1.96.$$

The following table is Table 4.2 in Berger and applies to the above situation:

n	5	20	100	1000
α_0	0.33	0.42	0.60	0.80

Even more amazing is the fact that Table 4.3 of Berger implies that in the situation on the previous page, $\alpha_0 \geq 0.127$ for all n and all priors!

Let's be sure we know what this means:

> If the experimenter is convinced that the prior probability of H_0 is 1/2 in Example 11, then having observed $z = 1.96$ (which is usually convincing evidence *against* H_0 for a frequentist), the posterior probability of H_0 must be at least 0.127, no matter the sample size or prior. **The point is that the frequentist and Bayes answers disagree profoundly**.

Lindley's paradox

Lindley's paradox occurs when a frequentist P-value is extremely small and a Bayesian posterior probability for H_0 is "large," say 1/2.

Lindley's paradox occurs mainly when testing a point null hypothesis. Example 11 provides a good example of the phenomenon, although the paradox can be arbitrarily paradoxical.

For example, we could have $P = 10^{-10}$ and $\alpha_0 = 1/2$.

Resolution of Lindley's paradox

- In some cases, the frequentist P-value probably does overstate the significance of the test statistic. We are all aware of cases where P is very small, and yet a confidence interval for θ *practically* includes θ_0.

- A "large" prior probability (such as 1/2) for a single point is often not realistic. In a case where $P = 10^{-10}$, $\pi_0 = 1/2$ and $\alpha_0 = 0.3$, the paradox may be resolved if one admits that π_0 should be quite small.

- Substitute $H_0 : \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$ for $H_0 : \theta = \theta_0$.