1 The Bayesian Approach to Parameter Estimation

In the Bayesian approach, we suppose that the unknown parameter θ is a random variable with a prior distribution $\pi(\theta)$. For a given value of θ, the data S have a pdf or pmf $f_\theta(s)$. The joint distribution of (S, θ) is

$$f(s, \theta) = f_\theta(s)\pi(\theta).$$

The marginal distribution of S has pmf or pdf

$$m(s) = \int_\Omega f(s, \theta) d\theta = \int_\Omega f_\theta(s)\pi(\theta) d\theta.$$

For the purposes of inference, we are interested in the conditional distribution of θ given $S = s$:

$$\pi(\theta|s) = \frac{f(s, \theta)}{m(s)} = \frac{f_\theta(s)\pi(\theta)}{\int_\Omega f_\theta(s)\pi(\theta) d\theta}.$$

The distribution specified by $\pi(\theta|s)$ is called the posterior distribution of θ given $S = s$. It represents the knowledge of the statistician arising from prior information and from the data. We see that as a function of θ, we can write the posterior distribution as

$$\begin{array}{rcl} \pi(\theta|s) & \propto & f_\theta(s)\pi(\theta) \\ & \text{or} & \\ \text{Posterior density} & \propto & \text{Likelihood} \times \text{Prior density}. \end{array}$$

Example: Suppose that X is binomial(n, θ) where θ has a beta(a, b) distribution. Find the posterior distribution of θ given $X = x$.

The likelihood and the prior density are

$$f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \text{ and}$$

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

The joint distribution is given by

$$\begin{aligned} f(x,\theta) &= \binom{n}{x}\theta^x(1-\theta)^{n-x} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\ &= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\theta^{x+a-1}(1-\theta)^{n+x+b-1}. \end{aligned}$$

Then the marginal pmf of X is

$$\begin{aligned} m(x) &= \int_0^1 f(x,\theta)d\theta = \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{x+a-1}(1-\theta)^{n-x+b-1}d\theta \\ &= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}. \end{aligned}$$

The posterior density of θ given $X=x$ is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x,\theta)}{m(x)} = \frac{\binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\theta^{x+a-1}(1-\theta)^{n-x+b-1}}{\binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}} \\ &= \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)}\theta^{x+a-1}(1-\theta)^{n-x+b-1}. \end{aligned}$$

Remark: We could have saved a lot of work by using the relationship

$$\text{Posterior density} \propto \text{Likelihood} \times \text{Prior density}$$

We notice that

$$\begin{aligned}
\text{Likelihood} &\propto \theta^x(1-\theta)^{n-x}, \\
\text{Prior density} &\propto \theta^{a-1}(1-\theta)^{b-1}, \\
&\text{and} \\
\text{Posterior density} &\propto \theta^x(1-\theta)^{n-x} \times \theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}.
\end{aligned}$$

We can form the posterior density by recognizing the constant (depending on x) that makes this function of θ integrate to 1. We notice that the form of the density is that of a beta density. Hence, the constant is

$$\frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)}.$$

Back to Bill of Rights Example:

We will consider two approaches to specifying the prior distribution and look at the resulting posterior distributions. The problem of specifying the prior distribution is central to the Bayes paradigm. Ideally the prior distribution reflects the investigator's knowledge about the parameter.

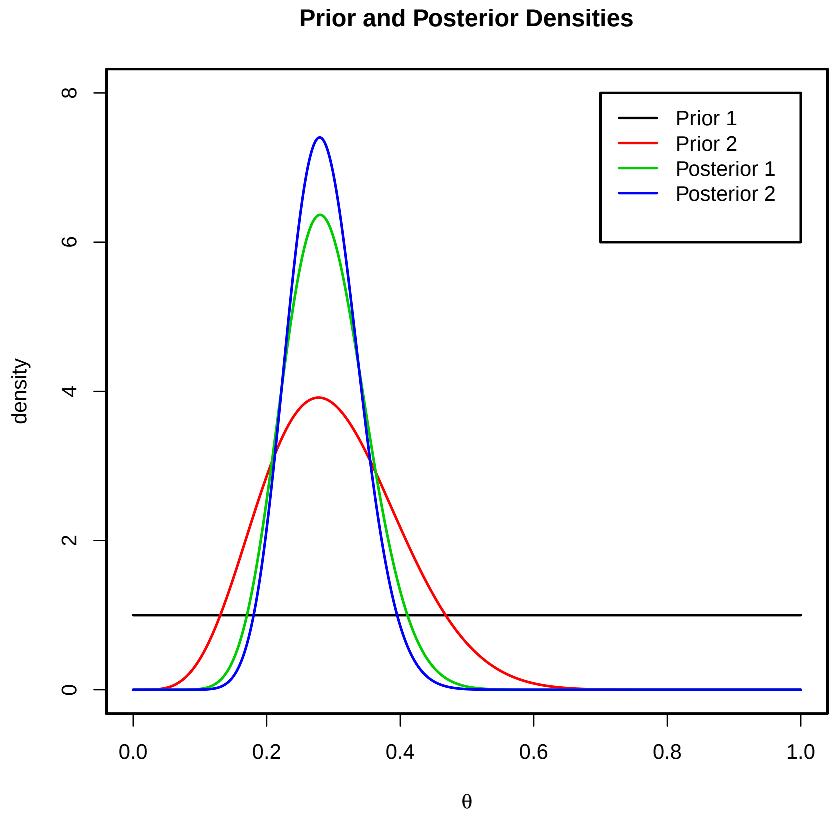
Investigator 1 desires to be "objective" and supposes that he has no prior information about θ. A uniform prior density on $(0, 1)$ is a reasonable choice for a "noninformative" prior. This corresponds to a beta $(1, 1)$ distribution.

Investigator 2 has information from an earlier study that found that the proportion knowing the bill of rights was 0.3, so she chooses a beta prior with this mean and a standard deviation of 0.1. This corresponds to a beta $(6, 14)$ distribution.

The data had $n = 50$ and $x = 14$. Thus, the two investigators ended up with the following posterior densities:

Investigator 1: Beta $(15, 37)$ Investigator 2: Beta $(20, 50)$

Plots of the prior and posterior densities are below:



1.1 Summarizing Results from a Bayesian Analysis

The posterior distribution summarizes the statistical analysis based on the prior information and the data. We can use summary measures for the distribution as Bayes estimates.

- The **posterior mode** is the mode of $\pi(\theta|s)$. This represents "the most likely" value of θ given $S = s$.
- The **posterior mean** is the mean of $\pi(\theta|s)$. This represents the average value of θ given $S = s$.
- The **posterior standard deviation and variance** are used to summarize the variability of θ given $S = s$.
- We could also report percentiles of the posterior distribution. The **posterior median** is sometimes used as a point estimate.

Back to Example We present a table with comparing the prior and posterior distributions for the two investigators:

Estimate	Prior 1	Bayes 1	Prior 2	Bayes 2
mode	none	0.28	0.2778	0.2794
mean	0.5	0.2885	0.3	0.2857
median	0.5	0.2857	0.2932	0.2837
standard deviation	0.289	0.0622	0.1	0.0533

We compare these with the mle and its estimated standard error:

$$\hat{\theta} = 0.28 \qquad SE(\hat{\theta}) = 0.0635$$

Back to Nerve Impulse Example

We assumed that $X_1, \ldots, X_n$ formed a random sample from an exponential (λ) distribution. Suppose now that λ is a rv with a gamma (a, b) distribution. Then

$$\pi(\lambda) \quad \propto \quad \lambda^{a-1} e^{-b\lambda}$$

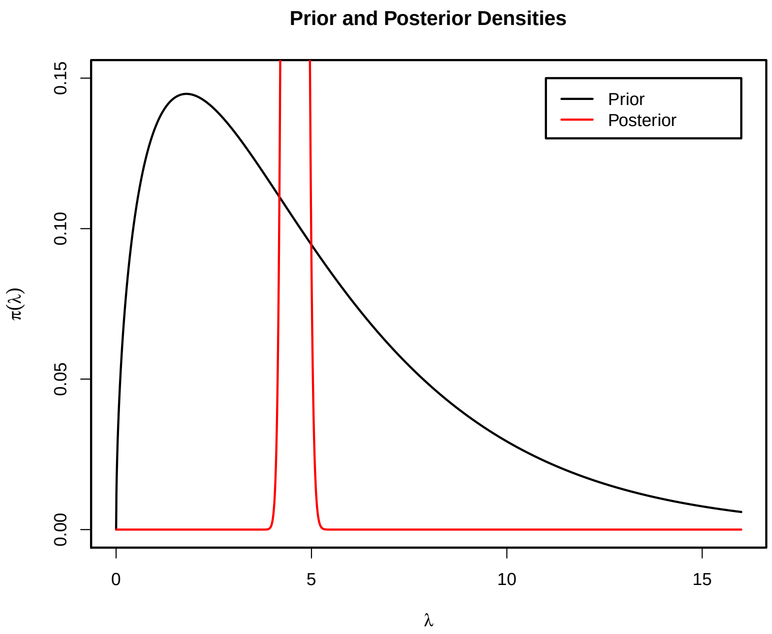
and

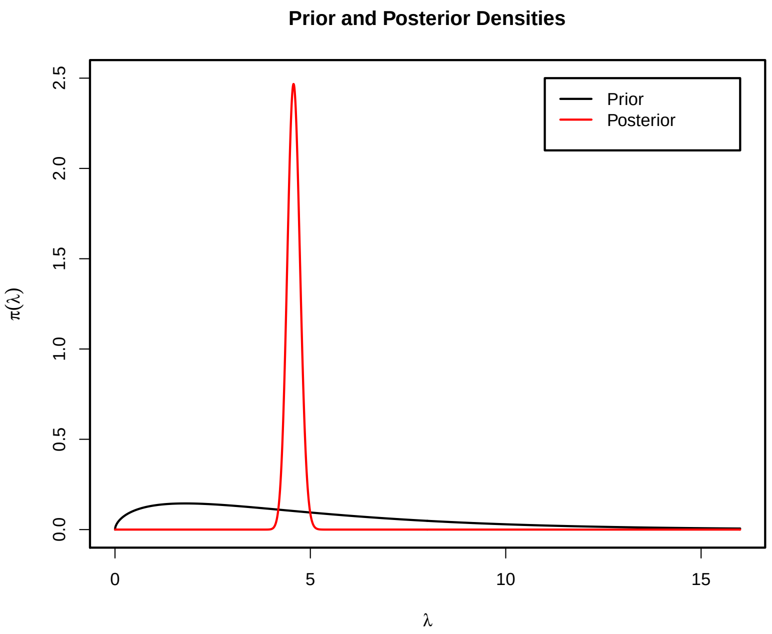
$$L(\lambda) \quad \propto \quad \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

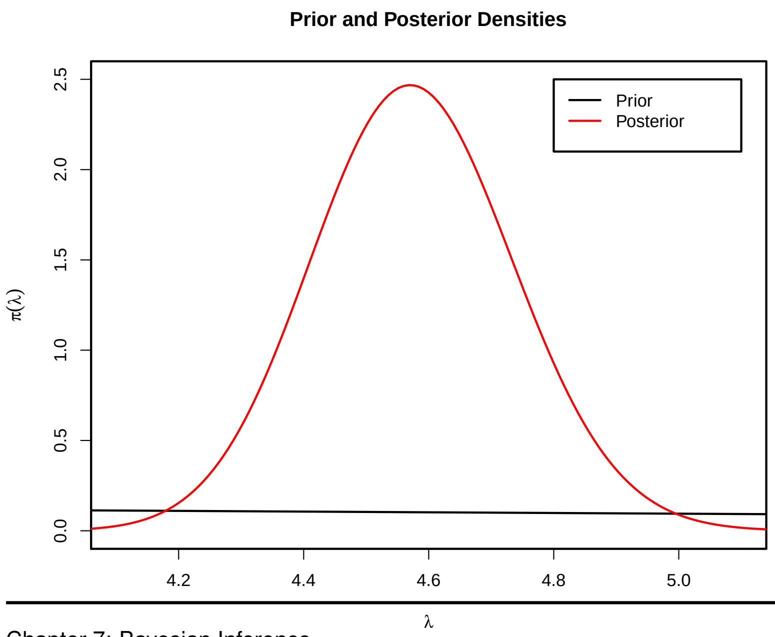
Thus, the posterior pdf can be seen to be a gamma $(a + n, b + \sum_{i=1}^n x_i)$ pdf:

$$\text{Posterior pdf} \quad \propto \quad \lambda^{a+n-1} e^{-(b+\sum_{i=1}^n x_i)\lambda}.$$

Suppose that the investigator wanted to use a prior distribution for λ with mean $= 5$ and variance $= 16$. This corresponds to a gamma $(25/16, 5/16)$ distribution. The posterior distribution given the observed data with $\sum x_i = 174.64$ and $n = 799$ is a gamma $(800.5625, 174.9525)$.


Prior and Posterior Densities
Prior
Posterior
0.00
0.05
0.10
0.15
0
5
10
15
$\pi(\lambda)$
λ


Prior and Posterior Densities
Prior
Posterior
0.0
0.5
1.0
1.5
2.0
2.5
0
5
10
15
$\pi(\lambda)$
λ


Prior and Posterior Densities
Prior
Posterior
0.0
0.5
1.0
1.5
2.0
2.5
4.2
4.4
4.6
4.8
5.0
$\pi(\lambda)$
λ

We can compare various estimates of λ:

$$\begin{aligned} E(\lambda|x_1,\ldots,x_n) &= \frac{800.5625}{174.9525} = 4.5759 \\ \text{mode of posterior} &= \frac{n+a-1}{b+\sum_{i=1}^n x_i} = \frac{799.5625}{174.9525} = 4.5702 \\ \text{median of posterior} = 4.5740 & \\ \text{mle} &= \frac{1}{\bar{x}} = 4.5751 \end{aligned}$$

1.2 Choice of Prior Distribution

The prior distribution should reflect the investigator's knowledge about the parameter.

When the investigator has limited knowledge about the parameter, a so-called noninformative prior could be used. These priors typically depend on the type of parameter being estimated.

For instance, a location parameter (i.e., $f(x|\theta) = f_0(x - \theta)$) would have a uniform distribution as the noninformative prior distribution. If the range of the parameter is infinite, this leads to an improper prior, a prior that does not have a finite integral.

Similar considerations for a scale parameter (i.e., $f(x|\sigma) = (1/\sigma)f_0(x/\sigma)$) leads to a prior density of the form $\pi(\sigma) = 1/\sigma$. Similarly, if the range of σ is infinite, this is an improper prior.

A common choice of prior distribution is a conjugate prior distribution. We say that a family Π of distributions is a conjugate family for $\mathcal{F} = \{f(x|\theta) : \theta \in \Omega\}$ if when we choose a prior distribution $\pi \in \Pi$, the posterior distribution is also in Π. Conjugate priors are mathematically convenient, but one should ascertain that they represent the investigator's knowledge of the parameter.

Distribution	Parameter	Conjugate Prior
Binomial	θ	beta
Negative binomial	θ	beta
Poisson	λ	gamma
Normal	μ	normal
Normal	σ	inverse gamma
Exponential	λ	gamma

1.3 Bayes Credible Intervals

An advantage of the Bayesian approach to inference is that one can construct intervals using the posterior distribution that have a given probability of containing the unknown parameter. We contrast this to the frequentist confidence intervals which contain an unknown parameter with a specified "level of confidence."

A level γ Bayes credible interval for the parameter $\psi(\theta)$ is an interval $C(s) = [\ell(s), u(s)]$ such that

$$P[\psi(\theta) \in C(s)|S = s] = \gamma.$$

We define a level γ highest posterior density (or HPD) interval as an interval $[\ell(s), u(s)]$ for θ that satisfies

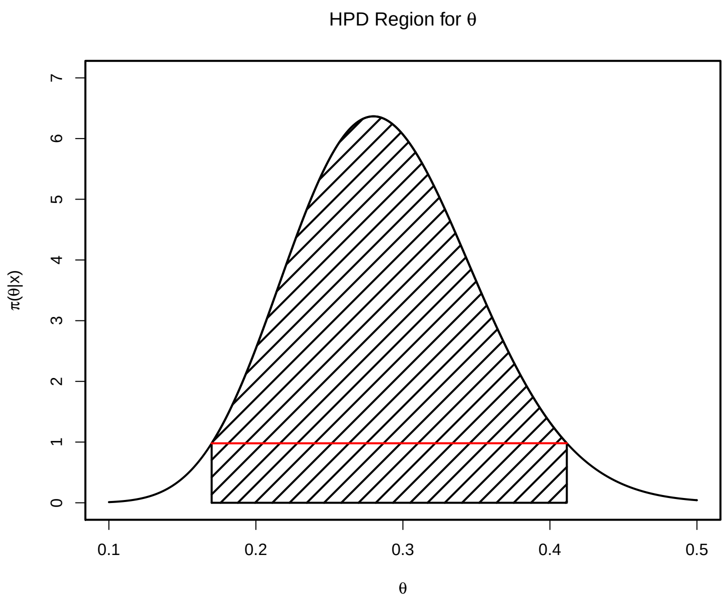
$$\int_{\ell(s)}^{u(s)} \pi(\theta|s)d\theta = \gamma$$

and

$$\pi(\theta|s) \geq \pi(\theta^*|s) \quad \text{for} \quad \theta \in [\ell(s), u(s)] \quad \text{and} \quad \theta^* \notin [\ell(s), u(s)].$$

Bill of Rights Example: Investigator 1 had a beta$(15, 37)$ posterior distribution. The following are various 95% intervals for θ:

Type of Interval	Lower	Upper
Bayes HPD	0.1699	0.4115
Bayes equal tail	0.1749	0.4174
Frequentist (Wald)	0.1555	0.4045
Frequentist (Score)	0.1747	0.4167



1.3.1 Bayes Inference for the Normal Mean

Suppose now that $X_1, \ldots, X_n$ form a random sample from a normal population with mean μ and variance σ^2. It is convenient to reparameterize the distribution in terms of the precision, $\xi = 1/\sigma^2$. The pdf becomes

$$f(x|\mu, \xi) = \left(\frac{\xi}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}\xi(x-\mu)^2\right).$$

We will examine the case where $\xi = \xi_0$ is **known**.

The conjugate prior distribution for θ is a normal $(\mu_0, \xi_{prior}^{-1})$ distribution.

Standard calculations result in the posterior distribution of μ given $X_1 = x_1, \ldots, X_n = x_n$ being normal with precision

$$\xi_{post} = n\xi_0 + \xi_{prior}.$$

The mean of the posterior distribution is

$$\begin{aligned}\mu_{post} &= \frac{n\xi_0\bar{x} + \mu_0\xi_{prior}}{n\xi_0 + \xi_{prior}} \\ &= \bar{x}\frac{n\xi_0}{n\xi_0 + \xi_{prior}} + \mu_0\frac{\xi_{prior}}{n\xi_0 + \xi_{prior}}.\end{aligned}$$

We see that the Bayes estimate, the mean, median, and mode of the posterior distribution is a weighted average of the prior mean μ_0 and the sample mean (mle) $\bar{x}$. The weights are proportion to the precision of each of the two quantities, the prior precision ξ_{prior} for μ_0 and the precision $n\xi_0 = n/\sigma_0^2$ of $\bar{x}$.

We next compare the Bayes HPD interval to the usual confidence interval:

$$\mu_{post} \pm Z_{(1+\gamma)/2}\sqrt{1/\xi_{post}},$$

$$\bar{x} \pm Z_{(1+\gamma)/2}\sqrt{\sigma^2/n}.$$

We see that the Bayes interval is shifted toward the prior mean and is shorter relative to the frequentist interval.

