Recall the definition of a conjugate family of priors (p. 43N).

For the normal model with unknown mean and known variance, we showed that for any normal prior and any set of data, the posterior for θ is also normal.

So, the family of normal distributions is a conjugate family of priors for the likelihood model defined on p. 77N.

The Posterior and Sufficient Statistics

Recall the *factorization theorem*:

A statistic $T(\boldsymbol{Y})$ is sufficient if and only if there exist functions $g(\,\cdot\,|\boldsymbol{\theta})$ and h such that, for all $\boldsymbol{y} \in \mathcal{Y}$ and all $\boldsymbol{\theta} \in \Theta$,

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = g(T(\boldsymbol{y})|\boldsymbol{\theta})h(\boldsymbol{y}).$$

Suppose $T(\boldsymbol{Y})$ is a sufficient statistic in the model $f(\boldsymbol{y}|\boldsymbol{\theta})$, and let the prior for $\boldsymbol{\theta}$ be π. Then the posterior is

$$\begin{aligned} \pi(\boldsymbol{\theta}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &\propto g(T(\boldsymbol{y})|\boldsymbol{\theta})h(\boldsymbol{y})\pi(\boldsymbol{\theta}) \\ &\propto g(T(\boldsymbol{y})|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \end{aligned}$$

This implies that *the posterior depends on the data only through the sufficient statistic.* Explicitly,

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{g(T(\boldsymbol{y})|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} g(T(\boldsymbol{y})|\boldsymbol{\eta})\pi(\boldsymbol{\eta})\,d\boldsymbol{\eta}}.$$

The last fact justifies using the notation

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \pi(\boldsymbol{\theta}|T(\boldsymbol{y}))$$

whenever $T(\boldsymbol{Y})$ is a sufficient statistic.

Likelihood Principle

An important principle in statistics, whether one is Bayesian or frequentist, is the *likelihood principle*. This principle says the following:

Let $L_1(\boldsymbol{\theta})$ and $L_2(\boldsymbol{\theta})$ be the likelihood functions for two different experiments. If

$$L_1(\boldsymbol{\theta}) = CL_2(\boldsymbol{\theta}) \quad \forall\ \boldsymbol{\theta},$$

where C is a quantity free of $\boldsymbol{\theta}$, then the inferences drawn about $\boldsymbol{\theta}$ in the two experiments should be the same.

Bayesian analysis agrees with the likelihood principle in the following sense:

> *In two experiments with likelihoods proportional to each other, if the two researchers use the same prior, they will also obtain the same posterior distribution.*

Exponential Data

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, where $Y_1, \ldots, Y_n$ are a random sample from an exponential density with unknown mean.

There are two ways of parameterizing the exponential distribution.

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

$$g(y|\theta) = \frac{1}{\theta} e^{-y/\theta} I_{(0,\infty)}(y) \qquad (B)$$

In either case the (unrestricted) parameter space is $\Theta = (0, \infty)$.

A conjugate family of priors for case (A) is the family of gamma distributions, while for case (B) the inverse-gamma family is conjugate.

Let's find the Jeffreys noninformative prior for case (A). The likelihood function is

$$\begin{aligned} f(\boldsymbol{y}|\theta) &= \theta^n \prod_{i=1}^{n} e^{-\theta y_i} \\ &= \theta^n e^{-\theta n \bar{y}}. \end{aligned}$$

(As an aside we note that the sample mean $\bar{Y}$ is a sufficient statistic.)

Now,

$$\log f(\boldsymbol{y}|\theta) = n \log \theta - \theta n \bar{y},$$

$$\frac{\partial}{\partial \theta} \log f(\boldsymbol{y}|\theta) = \frac{n}{\theta} - n\bar{y},$$

and

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

So, the Jeffreys noninformative prior for θ is proportional to θ^{-1}, and hence is improper.

Verify that the Jeffreys prior for case (B) is the same, i.e., it is proportional to θ^{-1}.

One may regared this improper prior as a limiting case of the Gamma(α, β) distribution, which is

$$\pi(\theta; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta} I_{(0,\infty)}(\theta).$$

Note that, for any $\theta > 0$,

$$\lim_{\alpha \to 0, \beta \to 0} \left(\frac{\theta^{\alpha-1} e^{-\beta\theta}}{\theta^{-1}} \right) = 1.$$

Let's compute the posterior for case (A) using a Gamma(α, β) prior.

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{\alpha-1} e^{-\beta\theta} \theta^n e^{-\theta n \bar{y}} = \theta^{n+\alpha-1} e^{-\theta(\beta+n\bar{y})}.$$

It follows that the posterior is Gamma$(n + \alpha, \beta + n\bar{y})$.

If we allow α and β to tend to 0, then the resulting posterior is Gamma$(n, n\bar{y})$, which is proper for any n and any $\bar{y} > 0$, the latter of which occurs with probability 1.

Gamma$(n, n\bar{y})$ is the posterior one obtains with the improper prior $\pi(\theta) = \theta^{-1}$. If one insisted on a proper noninformative prior, one could always take α and β *very* small, but positive.

This seems pedantic, though, in light of the fact that the posterior based on the improper prior is (i) proper and (ii) the limit of posteriors with proper priors.

The mean of the Gamma(α, β) distribution is α/β, and hence the Bayes estimate of θ_0 wrt squared error loss is

$$\frac{n+\alpha}{\beta + n\bar{y}} = \frac{1+\alpha/n}{\bar{y}+\beta/n}.$$

Regardless of the values of α and β, when n is sufficiently big,

$$\text{Bayes estimate} \approx \frac{1}{\bar{y}} = \text{MLE}.$$

Also, if we use the Jeffreys prior, the Bayes estimate is precisely the same as the MLE.

Whenever α is bigger than 1, the mode of the Gamma(α, β) distribution is $(\alpha - 1)/\beta$, and so the mode of the posterior is guaranteed to be $(n + \alpha - 1)/(\beta + n\bar{y})$ whenever $n \geq 2$. This estimate will generally differ very little from the mean of the posterior.