Chapter 3

General results for multi-parameter problems

In this chapter we will study some general results for multi-parameter problems.

3.1 Different levels of prior knowledge

We have *substantial prior information* for $\boldsymbol{\theta}$ when the prior distribution *dominates* the posterior distribution, that is $\pi(\boldsymbol{\theta}|\boldsymbol{x}) \sim \pi(\boldsymbol{\theta})$.

When prior information about $\boldsymbol{\theta}$ is limited, this is usually represented through the use of a conjugate prior distribution, with vague prior knowledge represented by making the conjugate distribution as diffuse as possible.

If we represent prior ignorance for a single parameter θ by using uniform or improper priors then we have seen (MAS2903, section 3.4) that, in general, the prior for $g(\theta)$ is not constant and so we are not ignorant about $g(\theta)$. The same problem occurs when we have more than one parameter.

Suppose we represent prior ignorance about $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ using $\pi(\boldsymbol{\theta}) = \textit{constant}$. Let $\phi_i = g_i(\boldsymbol{\theta})$, $i = 1, \ldots, p$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$ be a 1–1 transformation. Then, in general, the prior density for $\boldsymbol{\phi}$ is not constant and this suggests that we are not ignorant about $\boldsymbol{\phi}$. However, if we are ignorant about $\boldsymbol{\theta}$ then we must also be ignorant about $g(\boldsymbol{\theta})$.

This contradiction makes it impossible to use this representation of prior ignorance.

Example 3.1

Suppose $0 < \theta_1 < 1$ and $0 < \theta_2 < 1$. If we are ignorant about $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ then show that $\theta_1\theta_2$ does not have a constant prior density.

Example 3.2

Suppose we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ unknown). Determine the Jeffreys prior for this model.

Hint: We have already seen that the likelihood function can be written as

$$f(\boldsymbol{x}|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left[-\frac{n\tau}{2}\left\{s^2 + (\bar{x} - \mu)^2\right\}\right]$$

where

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

✎ *...Solution to Example 3.2*

✎ *...Solution to Example 3.2*

3.2 Asymptotic posterior distribution

Suppose we have a statistical model for data with likelihood function $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$, together with a prior distribution with density $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. Then

$$J(\hat{\boldsymbol{\theta}})^{1/2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|\boldsymbol{x} \xrightarrow{\mathcal{D}} N_p(0, I_p) \qquad \text{as } n \to \infty,$$

where $\hat{\boldsymbol{\theta}}$ is the likelihood mode, I_p is the $p \times p$ identity matrix and $J(\boldsymbol{\theta})$ is the observed information matrix, with $(i, j)^{th}$ element

$$J_{ij} = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\boldsymbol{x}|\boldsymbol{\theta}),$$

and $A^{1/2}$ denotes the square root matrix of A.

Comments

1. This asymptotic result can give us a useful approximation to the posterior distribution for $\boldsymbol{\theta}$ when n is large:

$$\boldsymbol{\theta}|\boldsymbol{x} \sim N_p\left(\hat{\boldsymbol{\theta}}, J(\hat{\boldsymbol{\theta}})^{-1}\right) \qquad \text{approximately.}$$

2. This limiting result is similar to one for the maximum likelihood estimator in Frequentist Statistics:

$$I(\boldsymbol{\theta})^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_p(0, I_p) \qquad \text{as } n \to \infty,$$

   where $I(\boldsymbol{\theta}) = E_{\boldsymbol{X}|\boldsymbol{\theta}}[J(\boldsymbol{\theta})]$ is Fisher's information matrix. Note that this statement about the distribution of $\hat{\boldsymbol{\theta}}$ for fixed (unknown) $\boldsymbol{\theta}$, whereas the results above is a statement about the distribution of $\boldsymbol{\theta}$ for fixed (known) $\hat{\boldsymbol{\theta}}$.

Example 3.3

Suppose we now have a random sample from a $N(\mu, 1/\tau)$ distribution (with unknown precision). Determine the asymptotic posterior distribution for (μ, τ).

Hint: we have already seen that the likelihood function can be written as

$$f(\boldsymbol{x}|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left[-\frac{n\tau}{2}\left\{s^2 + (\bar{x}-\mu)^2\right\}\right]$$

where $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$.

We have

$$\begin{aligned}
\log f(\boldsymbol{x}|\mu,\tau) &= \frac{n}{2}\log\tau - \frac{n}{2}\log(2\pi) - \frac{n\tau}{2}\left\{s^2 + (\bar{x}-\mu)^2\right\} \\
\Longrightarrow \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu,\tau) &= n\tau(\bar{x}-\mu) \\
\frac{\partial}{\partial\tau}\log f(\boldsymbol{x}|\mu,\tau) &= \frac{n}{2\tau} - \frac{n}{2}\left\{s^2 + (\bar{x}-\mu)^2\right\} \\
\Longrightarrow \frac{\partial^2}{\partial\mu^2}\log f(\boldsymbol{x}|\mu,\tau) &= -n\tau \\
\frac{\partial^2}{\partial\mu\partial\tau}\log f(\boldsymbol{x}|\mu,\tau) &= n(\bar{x}-\mu) \\
\frac{\partial^2}{\partial\tau^2}\log f(\boldsymbol{x}|\mu,\tau) &= -\frac{n}{2\tau^2}.
\end{aligned}$$

✎ *...Solution to Example 3.3* Now

$$\frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu,\tau)=0 \qquad \Longrightarrow \qquad \hat{\mu}=\bar{x}$$
$$\frac{\partial}{\partial\tau}\log f(\boldsymbol{x}|\mu,\tau)=0 \qquad \Longrightarrow \qquad \hat{\tau}=\frac{1}{s^2}$$

Therefore

$$J_{11}(\hat{\mu},\hat{\tau})=-\frac{\partial^2}{\partial\mu^2}\log f(\boldsymbol{x}|\mu,\tau)\bigg|_{(\hat{\mu},\hat{\tau})}=n\hat{\tau}=\frac{n}{s^2}$$
$$J_{12}(\hat{\mu},\hat{\tau})=-\frac{\partial^2}{\partial\mu\partial\tau}\log f(\boldsymbol{x}|\mu,\tau)\bigg|_{(\hat{\mu},\hat{\tau})}=-n(\bar{x}-\hat{\mu})=0$$
$$J_{22}(\hat{\mu},\hat{\tau})=-\frac{\partial^2}{\partial\tau^2}\log f(\boldsymbol{x}|\mu,\tau)\bigg|_{(\hat{\mu},\hat{\tau})}=\frac{n}{2\hat{\tau}^2}=\frac{ns^4}{2}$$

and so

$$J(\hat{\mu},\hat{\tau})=\begin{pmatrix}\frac{n}{s^2} & 0\\ 0 & \frac{ns^4}{2}\end{pmatrix},$$

whence

$$J(\hat{\mu},\hat{\tau})^{-1}=\begin{pmatrix}\frac{s^2}{n} & 0\\ 0 & \frac{2}{ns^4}\end{pmatrix}.$$

Therefore, for large n, the (approximate) posterior distribution for (μ,τ) is

$$\begin{pmatrix}\mu\\ \tau\end{pmatrix}\bigg|\,\boldsymbol{x}\sim N_2\left\{\begin{pmatrix}\bar{x}\\ \frac{1}{s^2}\end{pmatrix},\begin{pmatrix}\frac{s^2}{n} & 0\\ 0 & \frac{2}{ns^4}\end{pmatrix}\right\}.$$

Putting this another way, for large n

$$\mu|\boldsymbol{x}\sim N(\bar{x},s^2/n), \qquad \tau|\boldsymbol{x}\sim N\{1/s^2,2/(ns^4)\},$$

independently.

3.3 Learning objectives

By the end of this chapter, you should be able to:

- understand different levels of prior information and have an appreciation for the difficulty of specifying ignorance priors in multi-parameter problems
- determine the asymptotic posterior distribution when the data are a large random sample from **any** distribution
- explain the similarities and differences between the asymptotic posterior distribution and the asymptotic distribution of the maximum likelihood estimator