Chapter 3

General results for multi-parameter problems

3.1 Different levels of prior knowledge

- Substantial prior information: $\pi(\boldsymbol{\theta}|\boldsymbol{x}) \sim \pi(\boldsymbol{\theta})$
- Limited prior knowledge: use conjugate prior where possible and determine its parameter(s) by prior knowledge
- Vague prior information: use conjugate prior where possible and make them diffuse
- Prior ignorant – use 'flat' prior? Problem?
  1. Have $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ and we are prior ignorant
  2. Use $\pi(\boldsymbol{\theta}) = \mathit{constant}$?
  3. Let $\phi_i = g_i(\theta), i = 1, \ldots, p$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)^T$
  4. Prior density for $\boldsymbol{\phi}$ is not constant!
  5. But if we are ignorant about $\boldsymbol{\theta}$ we must be ignorant about $\boldsymbol{\phi}$...

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 constant prior density.

Solution

. . .

3.2 Asymptotic posterior distribution

$$p = 1 \text{ case :} \qquad \sqrt{J(\hat{\theta})}\,(\theta - \hat{\theta})\Big|\,\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty$$

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

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

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

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

Comments

When n is large, $\boldsymbol{\theta} \mid \boldsymbol{x} \sim N_p(\hat{\boldsymbol{\theta}}, J(\hat{\boldsymbol{\theta}})^{-1})$ approximately. . . .

Example 3.2

Suppose we now have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ unknown). 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 $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$

Solution

. . .

3.3 Learning objectives

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

1. describe the different levels of prior information; determine and use vague priors
2. determine the asymptotic posterior distribution when the data are a random sample from **any** distribution
3. explain the similarities and differences between the asymptotic posterior distribution and the asymptotic distribution of the maximum likelihood estimator