Solutions Manual to MATHEMATICAL STATISTICS: Asymptotic Minimax Theory

Alexander Korostelev
Wayne State University,
Detroit, MI 48202

Olga Korosteleva
California State University,
Long Beach, CA 90840

Chapter 1

Exercise 1.1 To verify first that the representation holds, compute the second partial derivative of $\ln p(x,\theta)$ with respect to θ. It is

$$\frac{\partial^2 \ln p(x,\theta)}{\partial\theta^2} = -\frac{1}{\big[p(x,\theta)\big]^2}\Big(\frac{\partial\, p(x,\theta)}{\partial\,\theta}\Big)^2 + \frac{1}{p(x,\theta)}\,\frac{\partial^2\, p(x,\theta)}{\partial\,\theta^2}$$

$$= -\Big(\frac{\partial\,\ln p(x,\theta)}{\partial\theta}\Big)^2 + \frac{1}{p(x,\theta)}\,\frac{\partial^2\, p(x,\theta)}{\partial\,\theta^2}\,.$$

Multiplying by $p(x,\theta)$ and rearranging the terms produce the result,

$$\Big(\frac{\partial \ln\, p\,(x,\theta)}{\partial\,\theta}\Big)^2 p\,(x,\,\theta) = \frac{\partial^2\, p\,(x,\,\theta)}{\partial\theta^2} - \Big(\frac{\partial^2\,\ln\, p\,(x,\,\theta)}{\partial\theta^2}\Big)\, p\,(x,\,\theta).$$

Now integrating both sides of this equality with respect to x, we obtain

$$I_n(\theta) = n\,\mathbb{E}_\theta\Big[\Big(\frac{\partial \ln\, p\,(X,\theta)}{\partial\,\theta}\Big)^2\Big] = n\int_{\mathbb{R}}\Big(\frac{\partial \ln\, p\,(x,\theta)}{\partial\,\theta}\Big)^2 p\,(x,\theta)\,dx$$

$$= n\int_{\mathbb{R}}\frac{\partial^2\, p\,(x,\theta)}{\partial\theta^2}\,dx - n\int_{\mathbb{R}}\Big(\frac{\partial^2\,\ln\, p\,(x,\theta)}{\partial\theta^2}\Big)\, p\,(x,\theta)\,dx$$

$$= n\,\underbrace{\frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}} p(x,\theta)\,dx}_{0} - n\int_{\mathbb{R}}\Big(\frac{\partial^2\,\ln\, p\,(x,\theta)}{\partial\theta^2}\Big)\, p\,(x,\theta)\,dx$$

$$= -\,n\int_{\mathbb{R}}\Big(\frac{\partial^2\,\ln\, p\,(x,\theta)}{\partial\theta^2}\Big)\, p\,(x,\theta)\,dx = -\,n\,\mathbb{E}_\theta\Big[\frac{\partial^2\,\ln\, p\,(x,\theta)}{\partial\theta^2}\Big].$$

Exercise 1.2 The first step is to notice that θ_n^* is an unbiased estimator of θ. Indeed, $\mathbb{E}_\theta[\,\theta_n^*] = \mathbb{E}_\theta\big[(1/n)\sum_{i=1}^n (X_i-\mu)^2\big] = \mathbb{E}_\theta[(X_1-\mu)^2] = \theta$.
Further, the log-likelihood function for the $\mathcal{N}(\mu,\theta)$ distribution has the form

$$\ln p(x,\theta) = -\frac{1}{2}\ln(2\,\pi\,\theta) - \frac{(x-\mu)^2}{2\,\theta}\,.$$

Therefore,

$$\frac{\partial \ln p(x,\theta)}{\partial\theta} = -\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2}, \quad\text{and}\quad \frac{\partial^2\,\ln p(x,\theta)}{\partial\theta^2} = \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}\,.$$

Applying the result of Exercise 1.1, we get

$$I_n(\theta) = -\,n\,\mathbb{E}_\theta\Big[\frac{\partial^2\,\ln p(X,\theta)}{\partial\theta^2}\Big] = -\,n\,\mathbb{E}_\theta\Big[\frac{1}{2\theta^2} - \frac{(X-\mu)^2}{\theta^3}\Big]$$

$$= -n\left[\frac{1}{2\theta^2} - \frac{\theta}{\theta^3}\right] = \frac{n}{2\theta^2}\,.$$

Next, using the fact that $\sum_{i=1}^{n}(X_i-\mu)^2/\theta$ has a chi-squared distribution with n degrees of freedom, and, hence its variance equals to $2n$, we arrive at

$$\mathbb{V}ar_\theta\left[\theta_n^*\right] = \mathbb{V}ar_\theta\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] = \frac{2n\theta^2}{n^2} = \frac{2\theta^2}{n} = \frac{1}{I_n(\theta)}\,.$$

Thus, we have shown that θ_n^* is an unbiased estimator of θ and that its variance attains the Cramér-Rao lower bound, that is, θ_n^* is an efficient estimator of θ.

Exercise 1.3 For the Bernoulli(θ) distribution,

$$\ln p\,(x,\,\theta) = x\,\ln\,\theta + (1-x)\,\ln(1-\theta),$$

thus,

$$\frac{\partial\,\ln p(x,\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta} \quad \text{and} \quad \frac{\partial^2\,\ln p(x,\theta)}{\partial\theta^2} = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}\,.$$

From here,

$$I_n(\theta) = -n\,\mathbb{E}_\theta\Big[-\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\Big] = n\Big(\frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2}\Big) = \frac{n}{\theta(1-\theta)}\,.$$

On the other hand, $\mathbb{E}_\theta\big[\bar{X}_n\big] = \mathbb{E}_\theta\big[X\big] = \theta$ and $\mathbb{V}ar_\theta\big[\bar{X}_n\big] = \mathbb{V}ar_\theta\big[X\big]/n = \theta(1-\theta)/n = 1/I_n(\theta)$. Therefore $\theta_n^* = \bar{X}_n$ is efficient.

Exercise 1.4 In the Poisson(θ) model,

$$\ln p\,(x,\,\theta) = x\,\ln\theta - \theta - \ln x!\,,$$

hence,

$$\frac{\partial\ln p(x,\theta)}{\partial\theta} = \frac{x}{\theta} - 1 \quad \text{and} \quad \frac{\partial^2\ln p(x,\theta)}{\partial\theta^2} = -\frac{x}{\theta^2}\,.$$

Thus,

$$I_n(\theta) = -n\,\mathbb{E}_\theta\Big[-\frac{X}{\theta^2}\Big] = \frac{n}{\theta}\,.$$

The estimate $\bar{X}_n$ is unbiased with the variance $\mathbb{V}ar_\theta\big[\bar{X}_n\big] = \theta/n = 1/I_n(\theta)$, and therefore efficient.

Exercise 1.5 For the given exponential density,

$$\ln p\,(x,\,\theta) = -\ln\theta - x/\theta\,,$$

whence,

$$\frac{\partial \ln p(x,\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2} \quad \text{and} \quad \frac{\partial^2 \ln p(x,\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2x}{\theta^3}\,.$$

Therefore,

$$I_n(\theta) = -n\,\mathbb{E}_\theta\Big[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\Big] = -n\Big[\frac{1}{\theta^2} - \frac{2\theta}{\theta^3}\Big] = \frac{n}{\theta^2}\,.$$

Also, $\mathbb{E}_\theta\big[\bar{X}_n\big] = \theta$ and $\mathbb{V}ar_\theta\big[\bar{X}_n\big] = \theta^2/n = 1/I_n(\theta)$. Hence efficiency holds.

Exercise 1.6 If $X_1, \dots, X_n$ are independent exponential random variables with the mean $1/\theta$, their sum $Y = \sum_{i=1}^n X_i$ has a gamma distribution with the density

$$f_Y(y) = \frac{y^{n-1}\,\theta^n\,e^{-y\theta}}{\Gamma(n)}\,,\quad y > 0\,.$$

Consequently,

$$\mathbb{E}_\theta\Big[\frac{1}{\bar{X}_n}\Big] = \mathbb{E}_\theta\Big[\frac{n}{Y}\Big] = n\int_0^\infty \frac{1}{y}\,\frac{y^{n-1}\theta^n\,e^{-y\theta}}{\Gamma(n)}\,dy$$

$$= \frac{n\,\theta}{\Gamma(n)}\int_0^\infty y^{n-2}\,\theta^{n-1}\,e^{-y\theta}\,dy = \frac{n\,\theta\,\Gamma(n-1)}{\Gamma(n)}$$

$$= \frac{n\,\theta\,(n-2)!}{(n-1)!} = \frac{n\,\theta}{n-1}\,.$$

Also,

$$\mathbb{V}ar_\theta\big[\,1/\bar{X}_n\,\big] = \mathbb{V}ar_\theta\big[\,n/Y\,\big] = n^2\Big(\mathbb{E}_\theta\big[1/Y^2\big] - \big(\mathbb{E}_\theta\big[\,1/Y\,\big]\big)^2\Big)$$

$$= n^2\Big[\frac{\theta^2\Gamma(n-2)}{\Gamma(n)} - \frac{\theta^2}{(n-1)^2}\Big] = n^2\,\theta^2\Big[\frac{1}{(n-1)(n-2)} - \frac{1}{(n-1)^2}\Big]$$

$$= \frac{n^2\,\theta^2}{(n-1)^2(n-2)}\,.$$

Exercise 1.7 The trick here is to notice the relation

$$\frac{\partial \ln p_0(x-\theta)}{\partial\theta} = \frac{1}{p_0(x-\theta)}\,\frac{\partial p_0(x-\theta)}{\partial\theta}$$

$$= -\frac{1}{p_0(x-\theta)}\,\frac{\partial p_0(x-\theta)}{\partial x} = -\frac{p_0{}'(x-\theta)}{p_0(x-\theta)}\,.$$

Thus we can write

$$I_n(\theta) = n\,\mathbb{E}_\theta\Big[\Big(-\frac{p_0{}'(X-\theta)}{p_0(X-\theta)}\Big)^2\Big] = n\int_{\mathbb{R}} \frac{\big(p_0{}'(y)\big)^2}{p_0(y)}\,dy\,,$$

which is a constant independent of θ.

Exercise 1.8 Using the expression for the Fisher information derived in the previous exercise, we write

$$I_n(\theta) = n\int_{\mathbb{R}} \frac{\big(p_0{}'(y)\big)^2}{p_0(y)}\,dy = n\int_{-\pi/2}^{\pi/2} \frac{\big(-C\alpha\,\cos^{\alpha-1} y\,\sin y\big)^2}{C\cos^{\alpha} y}\,dy$$

$$= n\,C\,\alpha^2\int_{-\pi/2}^{\pi/2} \sin^2 y\cos^{\alpha-2} y\,dy = n\,C\,\alpha^2\int_{-\pi/2}^{\pi/2} (1-\cos^2 y)\cos^{\alpha-2} y\,dy$$

$$= n\,C\,\alpha^2\int_{-\pi/2}^{\pi/2} \big(\cos^{\alpha-2} y - \cos^{\alpha} y\big)\,dy\,.$$

Here the first term is integrable if $\alpha - 2 > -1$ (equivalently, $\alpha > 1$), while the second one is integrable if $\alpha > -1$. Therefore, the Fisher information exists when $\alpha > 1$.

Chapter 2

Exercise 2.9 By Exercise 1.4, the Fisher information of the Poisson(θ) sample is $I_n(\theta) = n/\theta$. The joint distribution of the sample is

$$p(X_1, \dots X_n, \theta) = C_n \, \theta^{\sum X_i} e^{-n\theta}$$

where $C_n = C_n(X_1, \dots, X_n)$ is the normalizing constant independent of θ. As a function of θ, this joint probability has the algebraic form of a gamma distribution. Thus, if we select the prior density to be a gamma density, $\pi(\theta) = C(\alpha, \beta)\, \theta^{\alpha-1} e^{-\beta\theta}$, $\theta > 0$, for some positive α and β, then the weighted posterior density is also a gamma density,

$$\tilde{f}(\theta \,|\, X_1, \dots, X_n) = I_n(\theta)\, C_n\, \theta^{\sum X_i} e^{-n\theta} C(\alpha, \beta)\, \theta^{\alpha-1} e^{-\beta\theta}$$
$$= \tilde{C}_n\, \theta^{\sum X_i + \alpha - 2} e^{-(n+\beta)\theta}, \; \theta > 0,$$

where $\tilde{C}_n = n\, C_n(X_1, \dots, X_n)\, C(\alpha, \beta)$ is the normalizing constant. The expected value of the weighted posterior gamma distribution is equal to

$$\int_0^\infty \theta\, \tilde{f}(\theta \,|\, X_1, \dots, X_n)\, d\theta = \frac{\sum X_i + \alpha - 1}{n + \beta}.$$

Exercise 2.10 As shown in Example 1.10, the Fisher information $I_n(\theta) = n/\sigma^2$. Thus, the weighted posterior distribution of θ can be found as follows:

$$\tilde{f}\big(\theta \,|\, X_1, \dots, X_n\big) = C\, I_n(\theta)\, \exp\Big\{ -\frac{\sum \big(X_i - \theta\big)^2}{2\sigma^2} - \frac{(\theta - \mu)^2}{2\sigma_\theta^2} \Big\}$$
$$= C\, \frac{n}{\sigma^2} \exp\Big\{ -\Big(\frac{\sum X_i^2}{2\sigma^2} - \frac{2\theta \sum X_i}{2\sigma^2} + \frac{n\theta^2}{2\sigma^2} + \frac{\theta^2}{2\sigma_\theta^2} - \frac{2\theta\mu}{2\sigma_\theta^2} + \frac{\mu^2}{2\sigma_\theta^2} \Big) \Big\}$$
$$= C_1 \exp\Big\{ -\frac{1}{2}\Big[\theta^2 \big(\frac{n}{\sigma^2} + \frac{1}{\sigma_\theta^2} \big) - 2\,\theta \big(\frac{n\bar{X}_n}{\sigma^2} + \frac{\mu}{\sigma_\theta^2} \big) \Big] \Big\}$$
$$= C_2 \exp\Big\{ -\frac{1}{2} \big(\frac{n}{\sigma^2} + \frac{1}{\sigma_\theta^2} \big) \Big(\theta - \big(n\,\sigma_\theta^2\, \bar{X}_n + \mu\sigma^2\big) / \big(n\,\sigma_\theta^2 + \sigma^2\big) \Big)^2 \Big\}.$$

Here C, C_1, and C_2 are the appropriate normalizing constants. Thus, the weighted posterior mean is $\big(n\,\sigma_\theta^2\, \bar{X}_n + \mu\sigma^2\big) / \big(n\,\sigma_\theta^2 + \sigma^2\big)$ and the variance is $\big(n/\sigma^2 + 1/\sigma_\theta^2\big)^{-1} = \sigma^2\sigma_\theta^2 / \big(n\,\sigma_\theta^2 + \sigma^2\big)$.

Exercise 2.11 First, we derive the Fisher information for the exponential model. We have

$$\ln p(x, \theta) = \ln \theta - \theta\, x, \quad \frac{\partial \ln p(x, \theta)}{\partial \theta} = \frac{1}{\theta} - x,$$

and

$$\frac{\partial^2 \ln p(x,\theta)}{\partial \theta^2} = -\frac{1}{\theta^2}.$$

Consequently,

$$I_n(\theta) = -n\mathbb{E}_\theta\Big[-\frac{1}{\theta^2}\Big] = \frac{n}{\theta^2}.$$

Further, the joint distribution of the sample is

$$p(X_1, \dots X_n, \theta) = C_n\, \theta^{\sum X_i} e^{-\theta \sum X_i}$$

with the normalizing constant $C_n = C_n(X_1, \dots, X_n)$ independent of θ. As a function of θ, this joint probability belongs to the family of gamma distributions, hence, if we choose the conjugate prior to be a gamma distribution, $\pi(\theta) = C(\alpha, \beta)\, \theta^{\alpha-1} e^{-\beta\theta}$, $\theta > 0$, with some $\alpha > 0$ and $\beta > 0$, then the weighted posterior is also a gamma,

$$\begin{aligned}\tilde{f} = (\theta \,|\, X_1, \dots, X_n) &= I_n(\theta)\, C_n\, \theta^{\sum X_i} e^{-\theta \sum X_i} C(\alpha, \beta)\, \theta^{\alpha-1} e^{-\beta\theta} \\ &= \tilde{C}_n\, \theta^{\sum X_i + \alpha - 3} e^{-(\sum X_i + \beta)\,\theta}\end{aligned}$$

where $\tilde{C}_n = n\, C_n(X_1, \dots, X_n)\, C(\alpha, \beta)$ is the normalizing constant. The corresponding weighted posterior mean of the gamma distribution is equal to

$$\int_0^\infty \theta\, \tilde{f}(\theta \,|\, X_1, \dots, X_n)\, d\theta = \frac{\sum X_i + \alpha - 2}{\sum X_i + \beta}.$$

Exercise 2.12 (i) The joint density of n independent Bernoulli(θ) observations $X_1, \dots, X_n$ is

$$p(X_1, \dots X_n, \theta) = \theta^{\sum X_i} (1-\theta)^{n - \sum X_i}.$$

Using the conjugate prior $\pi(\theta) = C\big[\theta(1-\theta)\big]^{\sqrt{n}/2 - 1}$, we obtain the non-weighted posterior density $f(\theta \,|\, X_1, \dots, X_n) = C\, \theta^{\sum X_i + \sqrt{n}/2 - 1} (1-\theta)^{n - \sum X_i + \sqrt{n}/2 - 1}$, which is a beta density with the mean

$$\theta_n^* = \frac{\sum X_i + \sqrt{n}/2}{\sum X_i + \sqrt{n}/2 + n - \sum X_i + \sqrt{n}/2} = \frac{\sum X_i + \sqrt{n}/2}{n + \sqrt{n}}.$$

(ii) The variance of θ_n^* is

$$\mathbb{V}ar_\theta\,[\theta_n^*] = \frac{n\,\mathbb{V}ar_\theta(X_1)}{(n+\sqrt{n})^2} = \frac{n\theta(1-\theta)}{(n+\sqrt{n})^2},$$

and the bias equals to

$$b_n(\theta, \theta_n^*) = \mathbb{E}_\theta[\theta_n^*] - \theta = \frac{n\theta + \sqrt{n}/2}{n + \sqrt{n}} - \theta = \frac{\sqrt{n}/2 - \sqrt{n}\,\theta}{n + \sqrt{n}}.$$

Consequently, the non-normalized quadratic risk of θ_n^* is

$$\mathbb{E}_\theta\big[(\theta_n^* - \theta)^2\big] \;=\; \mathbb{V}ar_\theta\big[\,\theta_n^*\big] \;+\; b_n^2(\theta, \theta_n^*)$$

$$= \frac{n\theta(1-\theta) + \big(\sqrt{n}/2 - \sqrt{n}\,\theta\big)^2}{(n+\sqrt{n})^2} \;=\; \frac{n/4}{(n+\sqrt{n})^2} \;=\; \frac{1}{4(1+\sqrt{n})^2}\,.$$

(iii) Let $t_n = t_n(X_1, \dots, X_n)$ be the Bayes estimator with respect to a non-normalized risk function

$$R_n(\theta, \hat{\theta}_n, w) = \mathbb{E}_\theta\big[w(\hat{\theta}_n - \theta)\big].$$

The statement and the proof of Theorem 2.5 remain exactly the same if the non-normalized risk and the corresponding Bayes estimator are used. Since θ_n^* is the Bayes estimator for a constant non-normalized risk, it is minimax.

Exercise 2.13 In Example 2.4, let $\alpha = \beta = 1 + 1/b$. Then the Bayes estimator assumes the form

$$t_n(b) = \frac{\sum X_i + 1/b}{n + 2/b}$$

where X_i's are independent Bernoulli(θ) random variables. The normalized quadratic risk of $t_n(b)$ is equal to

$$R_n\big(\theta, t_n(b), w\big) \;=\; \mathbb{E}_\theta\Big[\,\big(\sqrt{I_n(\theta)}\,(t_n(b) - \theta)\,\big)^2\,\Big]$$

$$= I_n(\theta)\,\Big[\mathbb{V}ar_\theta\big[t_n(b)\big] + b_n^2\big(\theta, t_n(b)\big)\Big]$$

$$= I_n(\theta)\,\Big[\,\frac{n\mathbb{V}ar_\theta[X_1]}{(n+2/b)^2} \;+\; \Big(\frac{n\mathbb{E}_\theta[X_1] + 1/b}{n+2/b} \;-\; \theta\Big)^2\,\Big]$$

$$= \frac{n}{\theta(1-\theta)}\,\Big[\,\frac{n\theta(1-\theta)}{(n+2/b)^2} \;+\; \Big(\frac{n\theta + 1/b}{n+2/b} \;-\; \theta\Big)^2\,\Big]$$

$$= \frac{n}{\theta(1-\theta)}\,\Big[\,\frac{n\theta(1-\theta)}{(n+2/b)^2} \;+\; \underbrace{\frac{(1-2\theta)^2}{b^2\,(n+2/b)^2}}_{\to 0}\,\Big]$$

$$\to\; \frac{n}{\theta(1-\theta)}\,\frac{n\theta(1-\theta)}{n^2} \;=\; 1 \;\text{ as }\; b \to \infty.$$

Thus, by Theorem 2.8, the minimax lower bound is equal to 1. The normalized quadratic risk of $\bar{X}_n = \lim_{b\to\infty} t_n(b)$ is derived as

$$R_n\big(\theta, \bar{X}_n, w\big) = \mathbb{E}_\theta\,\Big[\big(\sqrt{I_n(\theta)}\,(\bar{X}_n - \theta)\big)^2\Big]$$

$$= \;I_n(\theta)\,\mathbb{V}ar_\theta\big[\bar{X}_n\big] \;=\; \frac{n}{\theta(1-\theta)}\,\frac{\theta(1-\theta)}{n} = 1.$$

That is, it attains the minimax lower bound, and hence $\bar{X}_n$ is minimax.

Chapter 3

EXERCISE 3.14 Let $X \sim \text{Binomial}(n\,, \theta^2)$. Then

$$\mathbb{E}_\theta\Big[\,\big|\sqrt{X/n}\,-\,\theta\,\big|\,\Big] \;=\; \mathbb{E}_\theta\Big[\,\frac{\big|\,X/n\,-\,\theta^2\,\big|}{\big|\sqrt{X/n}\,+\,\theta\,\big|}\,\Big]$$

$$\leq\; \frac{1}{\theta}\,\mathbb{E}_\theta\Big[\,\big|\,X/n\,-\,\theta^2\,\big|\,\Big] \;\leq\; \frac{1}{\theta}\sqrt{\mathbb{E}_\theta\Big[\,\big(\,X/n\,-\,\theta^2\,\big)^2\,\Big]}$$

(by the Cauchy-Schwarz inequality)

$$=\; \frac{1}{\theta}\sqrt{\frac{\theta^2(1-\theta^2)}{n}} \;=\; \sqrt{\frac{1-\theta^2}{n}} \;\to 0 \;\text{ as }\; n\to\infty\,.$$

EXERCISE 3.15 First we show that the Hodges estimator $\hat{\theta}_n$ is asymptotically unbiased. To this end write

$$\mathbb{E}_\theta\big[\,\hat{\theta}_n\,-\,\theta\,\big] \;=\; \mathbb{E}_\theta\big[\,\hat{\theta}_n-\bar{X}_n\,+\,\bar{X}_n-\theta\,\big] \;=\; \mathbb{E}_\theta\big[\,\hat{\theta}_n-\bar{X}_n\,\big]$$

$$=\; \mathbb{E}_\theta\Big[\,-\,\bar{X}_n\,\mathbb{I}\big(|\bar{X}_n|<n^{-1/4}\big)\,\Big] \;<\; n^{-1/4} \;\to 0 \;\text{ as }\; n\to\infty.$$

Next consider the case $\theta \neq 0$. We will check that

$$\lim_{n\to\infty}\,\mathbb{E}_\theta\Big[\,n\,\big(\hat{\theta}_n\,-\,\theta\,\big)^2\,\Big] \;=\; 1\,.$$

Firstly, we show that

$$\mathbb{E}_\theta\Big[\,n\,\big(\hat{\theta}_n\,-\,\bar{X}_n\,\big)^2\,\Big] \;\to 0 \;\text{ as }\; n\to\infty.$$

Indeed,

$$\mathbb{E}_\theta\Big[\,n\,\big(\hat{\theta}_n\,-\,\bar{X}_n\,\big)^2\,\Big] \;=\; n\,\mathbb{E}_\theta\Big[\,(-\bar{X}_n)^2\,\mathbb{I}\big(|\bar{X}_n|\;<\;n^{-1/4}\big)\,\Big]$$

$$\leq\; n^{1/2}\,\mathbb{P}_\theta\big(|\bar{X}_n|\;<\;n^{-1/4}\big) \;=\; n^{1/2}\int_{-n^{1/4}-\theta\,n^{1/2}}^{n^{1/4}-\theta\,n^{1/2}} \frac{1}{\sqrt{2\pi}}\,e^{-z^2/2}\,dz$$

$$=\; n^{1/2}\int_{-n^{1/4}}^{n^{1/4}} \frac{1}{\sqrt{2\pi}}\,e^{-\,(u\,-\,\theta\,n^{1/2})^2/2}\,du.$$

Here we made a substitution $u \;=\; z\,+\,\theta\,n^{1/2}$. Now, since $|u| \;\leq\; n^{1/4}$, the exponent can be bounded from above as follows

$$-\,\big(u-\theta\,n^{1/2}\big)^2/2 \;=\; -\,u^2/2+u\,\theta\,n^{1/2}-\theta^2\,n/2 \;\leq\; -\,u^2/2+\theta\,n^{3/4}-\theta^2\,n/2,$$

and, thus, for all sufficiently large n, the above integral admits the upper bound

$$n^{1/2}\int_{-n^{1/4}}^{n^{1/4}}\frac{1}{\sqrt{2\pi}}\,e^{-(u-\theta\,n^{1/2})^2/2}\,du$$

$$\leq\; n^{1/2}\int_{-n^{1/4}}^{n^{1/4}}\frac{1}{\sqrt{2\pi}}\,e^{-u^2/2+\theta\,n^{3/4}-\theta^2\,n/2}\,du$$

$$\leq\; e^{-\theta^2\,n/4}\int_{-n^{1/4}}^{n^{1/4}}\frac{1}{\sqrt{2\pi}}\,e^{-u^2/2}\,du\;\to 0\;\text{ as }\;n\to\infty.$$

Further, we use the Cauchy-Schwarz inequality to write

$$\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\theta\big)^2\,\Big]\;=\;\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\bar X_n\,+\bar X_n\,-\,\theta\big)^2\,\Big]$$

$$=\;\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\bar X_n\big)^2\,\Big]\;+\;2\,\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\bar X_n\big)\big(\bar X_n\,-\,\theta\big)\,\Big]\;+\;\mathbb{E}_\theta\Big[\,n\,\big(\bar X_n\,-\,\theta\big)^2\,\Big]$$

$$\leq\;\underbrace{\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\bar X_n\big)^2\,\Big]}_{\to 0}\;+2\;\underbrace{\Big\{\,\mathbb{E}_\theta\Big[\,n\,\big(\hat\theta_n\,-\,\bar X_n\big)^2\Big]\,\Big\}^{1/2}}_{\to 0}\;\times$$

$$\times\;\underbrace{\Big\{\,\mathbb{E}_\theta\Big[\,n\,\big(\bar X_n\,-\,\theta\big)^2\,\Big]\,\Big\}^{1/2}}_{=1}\;+\;\underbrace{\mathbb{E}_\theta\Big[\,n\,\big(\bar X_n\,-\,\theta\big)^2\,\Big]}_{=1}\;\to 1\;\text{ as }\;n\to\infty\,.$$

Consider now the case $\theta=0$. We will verify that

$$\lim_{n\to\infty}\;\mathbb{E}_\theta\big[\,n\,\hat\theta_n^2\,\big]\;=\;0\,.$$

We have

$$\mathbb{E}_\theta\big[\,n\,\hat\theta_n^2\,\big]\;=\;\mathbb{E}_\theta\Big[\,n\,\bar X_n^2\,\mathbb{I}\,\big(|\bar X_n|\;\geq\;n^{-1/4}\,\big)\,\Big]$$

$$=\;\mathbb{E}_\theta\Big[\,\big(\sqrt{n}\bar X_n\big)^2\,\mathbb{I}\,\big(|\sqrt{n}\bar X_n|\;\geq\;n^{1/4}\,\big)\,\Big]\;=\;2\int_{n^{1/4}}^{\infty}\frac{z^2}{\sqrt{2\pi}}\,e^{-z^2/2}\,dz$$

$$\leq\;2\int_{n^{1/4}}^{\infty}e^{-z}\,dz\;=\;2\,e^{-n^{1/4}}\;\to\;0\;\text{ as }\;n\to\infty.$$

EXERCISE 3.16 The following lower bound holds:

$$\sup_{\theta\in\mathbb{R}}\;\mathbb{E}_\theta\Big[\,I_n(\theta)\,(\hat\theta_n-\theta)^2\,\Big]\;\geq\;n\,I_*\;\max_{\theta\in\{\theta_0,\,\theta_1\}}\;\mathbb{E}_\theta\Big[(\hat\theta_n-\theta)^2\,\Big]$$

$$\geq\;\frac{n\,I_*}{2}\,\Big\{\mathbb{E}_{\theta_0}\Big[(\hat\theta_n-\theta_0)^2\,\Big]\;+\;\mathbb{E}_{\theta_1}\Big[(\hat\theta_n-\theta_1)^2\,\Big]\,\Big\}$$

$$=\;\frac{n\,I_*}{2}\;\mathbb{E}_{\theta_0}\Big[\,(\hat\theta_n-\theta_0)^2\;+\;(\hat\theta_n-\theta_1)^2\;\exp\big\{\Delta L_n(\theta_0\,,\,\theta_1)\big\}\,\Big]\qquad\text{(by (3.8))}$$

$$\geq \frac{n\,I_*}{2}\,\mathbb{E}_{\theta_0}\Big[\Big((\hat\theta_n-\theta_0)^2\;+\;(\hat\theta_n-\theta_1)^2\,\exp\{z_0\}\Big)\,\mathbb{I}\big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\,\big)\Big]$$

$$\geq \frac{n\,I_*\,\exp\{z_0\}}{2}\,\mathbb{E}_{\theta_0}\Big[\Big((\hat\theta_n-\theta_0)^2\exp\{-z_0\}+(\hat\theta_n-\theta_1)^2\Big)\,\mathbb{I}\big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\,\big)\Big]$$

$$\geq \frac{n\,I_*\,\exp\{z_0\}}{2}\,\mathbb{E}_{\theta_0}\Big[\Big((\hat\theta_n-\theta_0)^2\;+\;(\hat\theta_n-\theta_1)^2\Big)\,\mathbb{I}\big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\,\big)\Big],$$

since $\exp\{-z_0\}\geq 1$ for z_0 is assumed negative,

$$\geq \frac{n\,I_*\,\exp\{z_0\}}{2}\,\frac{(\theta_1-\theta_0)^2}{2}\,\mathbb{P}_{\theta_0}\Big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\Big)$$

$$\geq \frac{n\,I_*\,p_0\,\exp\{z_0\}}{4}\Big(\frac{1}{\sqrt{n}}\Big)^2\;=\;\frac{1}{4}\,I_*\,p_0\,\exp\{z_0\}\,.$$

Exercise 3.17 First we show that the inequality stated in the hint is valid. For any x it is necessarily true that either $|x|\;\geq\;1/2$ or $|x-1|\;\geq\;1/2$, because if the contrary holds, then $-1/2\;<\;x\;<1/2$ and $-1/2\;<\;1-x\;<\;1/2$ imply that $1=x+(1-x)\;<1/2+1/2=1$, which is false.
Further, since $w(x)=w(-x)$ we may assume that $x>0$. And suppose that $x\;\geq\;1/2$ (as opposed to the case $x-1\;\geq\;1/2$). In view of the facts that the loss function w is everywhere nonnegative and is increasing on the positive half-axis, we have

$$w(x)\;+\;w(x-1)\;\geq\;w(x)\;\geq\;w(1/2).$$

Next, using the argument identical to that in Exercise 3.16, we obtain

$$\sup_{\theta\in\mathbb{R}}\,\mathbb{E}_\theta\Big[\,w\big(\sqrt{n}\,(\hat\theta_n-\theta)\big)\Big]\;\geq\;\frac{1}{2}\,\exp\{z_0\}\,\mathbb{E}_{\theta_0}\Big[\Big(\,w\big(\sqrt{n}\,(\hat\theta_n-\theta_0)\big)\,+$$

$$+\,w\big(\sqrt{n}\,(\hat\theta_n-\theta_1)\big)\Big)\,\mathbb{I}\big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\,\big)\Big]\,.$$

Now recall that $\theta_1\;=\;\theta_0\;+\;1/\sqrt{n}$ and use the inequality proved earlier to continue

$$\geq\;\frac{1}{2}\,w(1/2)\exp\{z_0\}\,\mathbb{P}_{\theta_0}\Big(\,\Delta L_n(\theta_0\,,\theta_1)\;\geq\;z_0\Big)\;\geq\;\frac{1}{2}\,w(1/2)\,p_0\,\exp\{z_0\}\,.$$

Exercise 3.18 It suffices to prove the assertion (3.14) for an indicator function, that is, for the bounded loss function $w(u)\;=\;\mathbb{I}\big(\,|u|\;>\;\gamma\,\big)$, where γ is a fixed constant. We write

$$\int_{-(b-a)}^{b-a}w\big(\,c-u\,\big)\,e^{-u^2/2}\,du\;=\;\int_{-(b-a)}^{b-a}\mathbb{I}\big(\,|c-u|\;>\;\gamma\,\big)\,e^{-u^2/2}\,du$$

$$= \int_{-(b-a)}^{c-\gamma} e^{-u^2/2}\, du \, + \int_{c+\gamma}^{b-a} e^{-u^2/2}\, du \, .$$

To minimize this expression over values of c, take the derivative with respect to c and set it equal to zero to obtain

$$e^{-(c-\gamma)^2} \, - \, e^{-(c+\gamma)^2} \, = \, 0, \ \ \text{or,} \ \ \text{equivalently,} \ \ (c-\gamma)^2 \, = \, (c+\gamma)^2 \, .$$

The solution is $c = 0$.

Finally, the result holds for any loss function w since it can be written as a limit of linear combinations of indicator functions,

$$\int_{-(b-a)}^{b-a} w(c-u)\, e^{-u^2/2}\, du \, = \, \lim_{n\to\infty} \sum_{i=1}^{n} \Delta w_i \int_{-(b-a)}^{b-a} \mathbb{I}\big(\, |c-u| \, > \, \gamma_i \,\big)\, e^{-u^2/2}\, du$$

where

$$\gamma_i = \frac{b-a}{n}\, i, \ \ \Delta w_i = w(\gamma_i) - w(\gamma_{i-1})\, .$$

Exercise 3.19 We will show that for both distributions the representation (3.15) takes place.

(i) For the exponential model, as shown in Exercise 2.11, the Fisher information $I_n(\theta) = n/\theta^2$, hence,

$$L_n\big(\,\theta_0 \, + \, t/\sqrt{I_n(\theta_0)}\,\big) \, - \, L_n(\theta_0) \, = \, L_n\big(\,\theta_0 \, + \, \frac{\theta_0\, t}{\sqrt{n}}\,\big) \, - \, L_n(\theta_0)$$

$$= n \ln\big(\,\theta_0 + \frac{\theta_0\, t}{\sqrt{n}}\,\big) \, - \, \big(\,\theta_0 + \frac{\theta_0\, t}{\sqrt{n}}\,\big)\, n\bar{X}_n \, - \, n\ln(\theta_0) \, + \, \theta_0\, n\, \bar{X}_n$$

$$= \cancel{n\ln(\theta_0)} \, + \, n\ln\big(1 + \frac{t}{\sqrt{n}}\big) \, - \, \cancel{\theta_0 n \bar{X}_n} \, - \, t\,\theta_0\,\sqrt{n}\,\bar{X}_n \, - \, \cancel{n\ln(\theta_0)} \, + \, \cancel{\theta_0 n \bar{X}_n}.$$

Using the Taylor expansion, we get that for large n,

$$n\ln\big(1 + \frac{t}{\sqrt{n}}\big) \, = \, n\,\Big(\, \frac{t}{\sqrt{n}} \, - \, \frac{t^2}{2\,n} \, + \, o_n\big(\frac{1}{n}\big)\Big) \, = \, t\,\sqrt{n} \, - \, t^2/2 \, + \, o_n(1).$$

Also, by the Central Limit Theorem, for all sufficiently large n, $\bar{X}_n$ is approximately $\mathcal{N}\big(1/\theta_0,\, 1/(n\theta_0^2)\big)$, that is, $(\bar{X}_n \, - \, 1/\theta_0)\theta_0\,\sqrt{n} = (\theta_0\,\bar{X}_n \, - \, 1)\sqrt{n}$ is approximately $\mathcal{N}(0,1)$. Consequently, $Z \, = \, -\,(\theta_0\,\bar{X}_n \, - \, 1)\sqrt{n}$ is approximately standard normal as well. Thus, $n\ln\big(\,1 \, + \, t/\sqrt{n}\,\big) \, - \, t\,\theta_0\,\sqrt{n}\,\bar{X}_n \, = \, t\,\sqrt{n} \, - \, t^2/2 \, + \, o_n(1) \, - \, t\,\theta_0\,\sqrt{n}\,\bar{X}_n \, = \, -\,t\,(\theta_0\,\bar{X}_n \, - \, 1)\sqrt{n} \, - \, t^2/2 \, + \, o_n(1) \, = \, t\,Z \, - \, t^2/2 \, + \, o_n(1)\, .$

(ii) For the Poisson model, by Exercise 1.4, $I_n(\theta) = n/\theta$, thus,

$$L_n\big(\,\theta_0 \,+\, t/\sqrt{I_n(\theta_0)}\,\big) \,-\, L_n(\theta_0) \,=\, L_n\big(\,\theta_0 \,+\, t\,\sqrt{\frac{\theta_0}{n}}\,\big) \,-\, L_n(\theta_0)$$

$$= n\,\bar{X}_n\,\ln\big(\,\theta_0 + t\,\sqrt{\frac{\theta_0}{n}}\,\big) \,-\, n\,\big(\,\theta_0 + t\,\sqrt{\frac{\theta_0}{n}}\,\big) \,-\, n\,\bar{X}_n\,\ln(\theta_0) \,+\, n\,\theta_0$$

$$= n\,\bar{X}_n\,\ln\big(1+\frac{t}{\sqrt{\theta_0\,n}}\big) - t\,\sqrt{\theta_0\,n} \,=\, n\,\bar{X}_n\,\Big(\,\frac{t}{\sqrt{\theta_0\,n}} - \frac{t^2}{2\,\theta_0\,n} + o_n\big(\frac{1}{n}\big)\,\Big) - t\,\sqrt{\theta_0\,n}$$

$$= t\,\bar{X}_n\,\sqrt{\frac{n}{\theta_0}} \,-\, t\,\sqrt{\theta_0\,n} \,-\, \frac{\bar{X}_n}{\theta_0}\,\frac{t^2}{2} \,+\, o_n(1)$$

$$= t\,Z \,-\, \big(1 \,+\, \frac{Z}{\sqrt{\theta_0\,n}}\big)\,\frac{t^2}{2} \,+\, o_n(1) \,=\, t\,Z \,-\, \frac{t^2}{2} \,+\, o_n(1).$$

Here we used the fact that by the CLT, for all large enough n, $\bar{X}_n$ is approximately $\mathcal{N}(\theta_0,\,\theta_0/n)$, and hence,

$$Z \,=\, \frac{\bar{X}_n - \theta_0}{\sqrt{\theta_0/n}} \,=\, \bar{X}_n\,\sqrt{\frac{n}{\theta_0}} \,-\, \sqrt{\theta_0\,n}$$

is approximately $\mathcal{N}(0,1)$ random variable. Also,

$$\frac{\bar{X}_n}{\theta_0} \,=\, \frac{\big(\sqrt{\theta_0\,n} \,+\, Z\big)\sqrt{\theta_0/n}}{\theta_0} \,=\, 1 \,+\, \frac{Z}{\sqrt{\theta_0\,n}} \,=\, 1 \,+\, o_n(1).$$

EXERCISE 3.20 Consider a truncated loss function $w_C(u) = \min(w(u), C)$ for some $C > 0$. As in the proof of Theorem 3.8, we write

$$\sup_{\theta\in\mathbb{R}} \mathbb{E}_\theta\Big[\,w_C\big(\,\sqrt{nI(\theta)}\,(\hat{\theta}_n \,-\, \theta)\,\big)\,\Big]$$

$$\geq \frac{\sqrt{nI(\theta)}}{2b}\int_{-b/\sqrt{nI(\theta)}}^{b/\sqrt{nI(\theta)}} \mathbb{E}_\theta\Big[\,w_C\big(\,\sqrt{nI(\theta)}\,(\hat{\theta}_n \,-\, \theta)\,\big)\,\Big]\,d\theta$$

$$= \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_{t/\sqrt{nI(\theta)}}\Big[\,w_C\big(\,\sqrt{nI(\theta)}\,\hat{\theta}_n \,-\, t\,\big)\,\Big]\,dt$$

where we used a change of variables $t = \sqrt{nI(\theta)}$. Let $a_n = nI(t/\sqrt{nI(0)})$. We continue

$$= \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\,w_C\big(\,\sqrt{a_n}\,\hat{\theta}_n \,-\, t\,\big)\,\exp\big\{\Delta L_n\big(0, t/\sqrt{nI(0)}\big)\big\}\,\Big]\,dt\,.$$

Applying the LAN condition (3.16), we get

$$= \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\, w_C\big(\sqrt{a_n}\,\hat\theta_n - t\,\big)\, \exp\big\{\, z_n(0)\,t \;-\; t^2/2 \,+\, \varepsilon_n(0,t)\,\big\}\,\Big]\,dt\,.$$

An elementary inequality $|x| \geq |y| - |x-y|$ for any x and $y \in \mathbb{R}$ implies that

$$\geq \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\, w_C\big(\sqrt{a_n}\,\hat\theta_n - t\,\big)\, \exp\big\{\, \tilde z_n(0)\,t \;-\; t^2/2\,\big\}\,dt +$$

$$+\frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\, w_C\big(\sqrt{a_n}\,\hat\theta_n - t\,\big)\,\Big|\, \exp\big\{\, z_n(0)\,t \;-\; t^2/2 \,+\, \varepsilon_n(0,t)\,\big\}$$

$$-\,\exp\big\{\, \tilde z_n(0)\,t \;-\; t^2/2\,\big\}\,\Big|\,\Big]\,dt\,.$$

Now, by Theorem 3.11, and the fact that $w_C \leq C$, the second term vanishes as n grows, and thus is $o_n(1)$ as $n \to \infty$. Hence, we obtain the following lower bound

$$\sup_{\theta\in\mathbb{R}} \mathbb{E}_\theta\Big[\, w_C\big(\sqrt{nI(\theta)}\,(\hat\theta_n - \theta)\,\big)\,\Big]$$

$$\geq \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\, w_C\big(\sqrt{a_n}\,\hat\theta_n - t\,\big)\, \exp\big\{\, \tilde z_n(0)\,t - t^2/2\,\big\}\,\Big]\,dt$$

$$+\,o_n(1)\,.$$

Put $\eta_n = \sqrt{a_n}\,\hat\theta_n - \tilde z_n(0)$. We can rewrite the bound as

$$\geq \frac{1}{2b}\int_{-b}^{b} \mathbb{E}_0\Big[\, \exp\big\{\tfrac{1}{2}\tilde z_n^2(0)\big\}\, w_C\big(\,\eta_n - (t - \tilde z_n(0))\,\big)\, \exp\big\{ -\tfrac{1}{2}\big(t - \tilde z_n(0)\big)^2\,\big\}\,\Big]\,dt$$

$$+\,o_n(1)$$

which, after the substitution $u = t - \tilde z_n(0)$ becomes

$$\geq \frac{1}{2b}\int_{-(b-a)}^{b-a} \mathbb{E}_0\Big[\, \exp\big\{\tfrac{1}{2}\tilde z_n^2(0)\big\}\, \mathbb{I}(|\tilde z_n(0)| \leq a)\, w_C(\eta_n - u)\, \exp\big\{ -\, \tfrac{1}{2}u^2\,\big\}\,\Big]\,du$$

$$+\,o_n(1)\,.$$

As in the proof of Theorem 3.8, for $n \to \infty$,

$$\mathbb{E}_0\Big[\, \exp\big\{\tilde z_n^2(0)\big\}\, \mathbb{I}(|\tilde z_n(0)| \leq a)\,\Big] \;\to\; \frac{2a}{\sqrt{2\pi}}\,,$$

and, by an argument similar to the proof of Theorem 3.9,

$$\int_{-(b-a)}^{b-a} w_C(\eta_n \,-\, u)\, \exp\big\{ -\, \frac{1}{2}\,u^2\,\big\}\,du \;\geq\; \int_{-(b-a)}^{b-a} w_C(u)\exp\big\{ -\, \frac{1}{2}\,u^2\,\big\}\,du\,.$$

Putting $a = b - \sqrt{b}$ and letting b, C and n go to infinity, we arrive at the conclusion that

$$\sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta \Big[w_C \big(\sqrt{nI(\theta)} \, (\hat{\theta}_n - \theta) \big) \Big] \geq \int_{-\infty}^{\infty} \frac{w(u)}{\sqrt{2\pi}} \, e^{-u^2/2} \, du \, .$$

EXERCISE 3.21 Note that the distorted parabola can be written in the form

$$zt - t^2/2 + \varepsilon(t) = -(1/2)(t-z)^2 + z^2/2 + \varepsilon(t) \, .$$

The parabola $-(1/2)(t-z)^2 + z^2/2$ is maximized at $t = z$. The value of the distorted parabola at $t = z$ is bounded from below by

$$-(1/2)(z-z)^2 + z^2/2 + \varepsilon(z) = z^2/2 + \varepsilon(z) \geq z^2/2 - \delta \, .$$

On the other hand, for all t such that $|\, t - z \,| > 2\sqrt{\delta}$, this function is strictly less than $z^2/2 - \delta$. Indeed,

$$\begin{aligned} -(1/2)(t-z)^2 + z^2/2 + \varepsilon(t) &< -(1/2)(2\sqrt{\delta})^2 + z^2/2 + \varepsilon(t) \\ &< -2\delta + z^2/2 + \delta = z^2/2 - \delta \, . \end{aligned}$$

Thus, the value $t = t^*$ at which the function is maximized must satisfy $|\, t^* - z \,| \leq 2\sqrt{\delta}$.

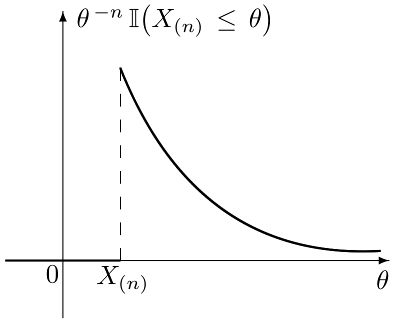
Chapter 4

EXERCISE 4.22 (i) The likelihood function has the form

$$\prod_{i=1}^{n} p(X_i, \theta) = \theta^{-n} \prod_{i=1}^{n} \mathbb{I}\big(0 \le X_i \le \theta\big)$$

$$= \theta^{-n}\,\mathbb{I}\big(0 \le X_1 \le \theta,\, 0 \le X_2 \le \theta, \dots, 0 \le X_n \le \theta\big) = \theta^{-n}\,\mathbb{I}\big(X_{(n)} \le \theta\big).$$

Here $X_{(n)} = \max(X_1, \dots, X_n)$. As depicted in the figure below, function θ^{-n} decreases everywhere, attaining its maximum at the left-most point. Therefore, the MLE of θ is $\hat{\theta}_n = X_{(n)}$.



(ii) The c.d.f. of $X_{(n)}$ can be found as follows:

$$F_{X_{(n)}}(x) = \mathbb{P}_\theta\big(X_{(n)} \le x\big) = \mathbb{P}_\theta\big(X_1 \le x, X_2 \le x, \dots, X_n \le x\big)$$

$$= \mathbb{P}_\theta\big(X_1 \le x\big)\,\mathbb{P}_\theta\big(X_2 \le x\big) \dots \mathbb{P}_\theta\big(X_n \le x\big) \ \text{(by independence)}$$

$$= \Big[\mathbb{P}(X_1 \le x)\Big]^n = \Big(\frac{x}{\theta}\Big)^n, \ \ 0 \le x \le \theta.$$

Hence the density of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = \Big(\frac{x^n}{\theta^n}\Big)' = \frac{n\,x^{n-1}}{\theta^n}.$$

The expected value of $X_{(n)}$ is computed as

$$\mathbb{E}_\theta[X_{(n)}] = \int_0^\theta x\,\frac{n\,x^{n-1}}{\theta^n}\,dx = \frac{n}{\theta^n}\int_0^\theta x^n\,dx = \frac{n\,\theta^{n+1}}{(n+1)\theta^n} = \frac{n\,\theta}{n+1},$$

and therefore,

$$\mathbb{E}_\theta\big[\theta_n^*\big] = \mathbb{E}_\theta\Big[\frac{n+1}{n}\,X_{(n)}\Big] = \frac{n+1}{n}\,\frac{n\,\theta}{n+1} = \theta.$$

(iii) The variance of $X_{(n)}$ is

$$\mathbb{V}ar_\theta\big[X_{(n)}\big] = \int_0^\theta x^2 \frac{n\,x^{n-1}}{\theta^n}\,dx - \Big(\frac{n\,\theta}{n+1}\Big)^2$$

$$= \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,dx - \Big(\frac{n\,\theta}{n+1}\Big)^2 = \frac{n\,\theta^{n+2}}{(n+2)\,\theta^n} - \Big(\frac{n\,\theta}{n+1}\Big)^2$$

$$= \frac{n\,\theta^2}{n+2} - \frac{n^2\,\theta^2}{(n+1)^2} = \frac{n\,\theta^2}{(n+1)^2\,(n+2)}.$$

Consequently, the variance of θ_n^* is

$$\mathbb{V}ar_\theta\big[\theta_n^*\big] = \mathbb{V}ar_\theta\Big[\frac{n+1}{n}X_{(n)}\Big] = \frac{(n+1)^2}{n^2}\,\frac{n\,\theta^2}{(n+1)^2\,(n+2)} = \frac{\theta^2}{n\,(n+2)}\,.$$

Exercise 4.23 (i) The likelihood function can be written as

$$\prod_{i=1}^n p(X_i,\theta) = \exp\Big\{-\big(\sum_{i=1}^n X_i - n\,\theta\big)\Big\}\prod_{i=1}^n \mathbb{I}\big(X_i \geq \theta\big)$$

$$= \exp\Big\{-\sum_{i=1}^n X_i + n\,\theta\Big\}\,\mathbb{I}\big(X_1 \geq \theta,\, X_2 \geq \theta, \dots, X_n \geq \theta\big)$$

$$= \exp\big\{n\,\theta\big\}\,\mathbb{I}\big(X_{(1)} \geq \theta\big)\,\exp\Big\{-\sum_{i=1}^n X_i\Big\}$$

with $X_{(1)} = \min(X_1,\dots,X_n)$. The second exponent is constant with respect to θ and may be disregarded for maximization purposes. The function $\exp\{n\,\theta\}$ is increasing and therefore reaches its maximum at the right-most point $\hat\theta_n = X_{(1)}$.

(ii) The c.d.f. of the minimum can be found by the following argument:

$$1 - F_{X_{(1)}}(x) = \mathbb{P}_\theta\big(X_{(1)} \geq x\big) = \mathbb{P}_\theta\big(X_1 \geq x,\, X_2 \geq x, \dots, X_n \geq x\big)$$

$$= \mathbb{P}_\theta\big(X_1 \geq x\big)\,\mathbb{P}_\theta\big(X_2 \geq x\big)\dots\mathbb{P}_\theta\big(X_n \geq x\big) \text{ (by independence)}$$

$$= \Big[\mathbb{P}_\theta\big(X_1 \geq x\big)\Big]^n = \Big[\int_x^\infty e^{-(y-\theta)}\,dy\Big]^n = \Big[e^{-(x-\theta)}\Big]^n = e^{-n(x-\theta)},$$

whence

$$F_{X_{(1)}}(x) = 1 - e^{-n(x-\theta)}.$$

Therefore, the density of $X_{(1)}$ is derived as

$$f_{X_{(1)}}(x) = F'_{X_{(1)}}(x) = \Big[1 - e^{-n(x-\theta)}\Big]' = n\,e^{-n(x-\theta)},\;\; x \geq \theta.$$

The expected value of $X_{(1)}$ is equal to

$$\mathbb{E}_\theta\big[\,X_{(1)}\,\big] \;=\; \int_\theta^\infty x\,n\,e^{-\,n\,(x-\theta)}\,dx$$

$$=\int_0^\infty \big(\,\frac{y}{n}\,+\,\theta\big)\,e^{-y}\,dy\;\; \big(\text{after substitution } \; y\,=\,n(x-\theta)\,\big)$$

$$=\,\frac{1}{n}\underbrace{\int_0^\infty y\,e^{-y}\,dy}_{=1}\,+\theta\underbrace{\int_0^\infty e^{-y}\,dy}_{=1}\;=\;\frac{1}{n}\,+\,\theta\,.$$

As a result, the estimator $\theta_n^* = X_{(1)} - 1/n$ is an unbiased estimator of θ.

(iii) The variance of $X_{(1)}$ is computed as

$$\mathbb{V}ar_\theta\big[\,X_{(1)}\,\big] \;=\; \int_\theta^\infty x^2\,n\,e^{-\,n\,(x-\theta)}\,dx\;-\;\Big(\frac{1}{n}\,+\,\theta\Big)^2$$

$$=\int_0^\infty \big(\,\frac{y}{n}\,+\,\theta\big)^2\,e^{-y}\,dy\;-\;\Big(\frac{1}{n}\,+\,\theta\Big)^2$$

$$=\,\frac{1}{n^2}\underbrace{\int_0^\infty y^{\,2}\,e^{-y}\,dy}_{=2}\,+\,\frac{2\,\theta}{n}\underbrace{\int_0^\infty y\,e^{-y}\,dy}_{=1}\,+\theta^2\underbrace{\int_0^\infty e^{-y}\,dy}_{=1}\,-$$

$$-\,\frac{1}{n^2}\,-\,\frac{2\,\theta}{n}\,-\,\theta^2\;=\;\frac{1}{n^2}\,.$$

EXERCISE 4.24 We will show that the squared L_2 - norm of $\sqrt{p(\,\cdot\,,\,\theta+\Delta\theta)} - \sqrt{p(\,\cdot\,,\,\theta)}$ is equal to $\Delta\theta + o(\Delta\theta)$ as $\Delta\theta\,\to\,0$. Then by Theorem 4.3 and Example 4.4 it will follow that the Fisher information does not exist. By definition, we obtain

$$\big\|\,\sqrt{p(\,\cdot\,,\,\theta+\Delta\theta)}\;-\;\sqrt{p(\,\cdot\,,\,\theta)}\,\big\|_2^2\;=$$

$$=\,\int_{\mathbb{R}}\Big[\,e^{-\,(x-\theta-\Delta\theta)/2}\,\mathbb{I}\big(\,x\,\geq\,\theta+\Delta\theta\,\big)\;-\;e^{-\,(x-\theta)/2}\,\mathbb{I}\big(\,x\,\geq\,\theta\,\big)\Big]^2\,dx$$

$$=\,\int_\theta^{\theta+\Delta\theta} e^{-\,(x-\theta)}\,dx\;+\;\int_{\theta+\Delta\theta}^\infty\Big(e^{-\,(x-\theta-\Delta\theta)/2}\;-\;e^{-\,(x-\theta)/2}\Big)^2\,dx$$

$$=\,\int_\theta^{\theta+\Delta\theta} e^{-\,(x-\theta)}\,dx\;+\;\Big(e^{\Delta\theta/2}\;-\;1\Big)^2\int_{\theta+\Delta\theta}^\infty e^{-\,(x-\theta)}\,dx$$

$$=\,1\;-\;e^{-\Delta\theta}\;+\;\Big(e^{\Delta\theta/2}\;-\;1\Big)^2\,e^{-\Delta\theta}$$

$$= 2 - 2\,e^{-\Delta\theta/2} = \Delta\theta + o(\Delta\theta) \text{ as } \Delta\theta \to 0\,.$$

EXERCISE 4.25 First of all, we find the values of c_- and c_+ as functions of θ. By our assumption, $c_+ - c_- = \theta$. Also, since the density integrates to one, $c_+ + c_- = 1$. Hence, $c_- = (1-\theta)/2$ and $c_+ = (1+\theta)/2$.

Next, we use the formula proved in Theorem 4.3 to compute the Fisher information. We have

$$I(\theta) = 4 \left\| \partial \sqrt{p(\,\cdot\,,\theta)}/\partial\theta \right\|_2^2 =$$

$$= 4 \Big[\int_{-1}^{0} \Big(\frac{\partial\sqrt{(1-\theta)/2}}{\partial\theta} \Big)^2 dx + \int_{0}^{1} \Big(\frac{\partial\sqrt{(1+\theta)/2}}{\partial\theta} \Big)^2 dx \Big]$$

$$= 4 \Big[\frac{1}{8(1-\theta)} + \frac{1}{8(1+\theta)} \Big] = \frac{1}{1-\theta^2}\,.$$

EXERCISE 4.26 In the case of the shifted exponential distribution we have

$$Z_n\big(\theta, \theta + u/n\big) = \prod_{i=1}^{n} \frac{\exp\big\{ -X_i + (\theta + u/n) \big\}\, \mathbb{I}\big(X_i \geq \theta + u/n \big)}{\exp\big\{ -X_i + \theta \big\}\, \mathbb{I}\big(X_i \geq \theta \big)}$$

$$= \frac{\exp\big\{ -\sum_{i=1}^{n} X_i + n\,(\theta + u/n) \big\}\, \mathbb{I}\big(X_{(1)} \geq \theta + u/n \big)}{\exp\big\{ -\sum_{i=1} X_i + n\,\theta \big\}\, \mathbb{I}\big(X_{(1)} \geq \theta \big)}$$

$$= e^{u}\, \frac{\mathbb{I}\big(X_{(1)} \geq \theta + u/n \big)}{\mathbb{I}\big(X_{(1)} \geq \theta \big)} = e^{u}\, \frac{\mathbb{I}\big(u \leq T_n \big)}{\mathbb{I}\big(X_{(1)} \geq \theta \big)} \text{ where } T_n = n\,(X_{(1)} - \theta)\,.$$

Here $\mathbb{P}_\theta\big(X_{(1)} \geq \theta \big) = 1$, and

$$\mathbb{P}_\theta\big(T_n \geq t \big) = \mathbb{P}_\theta\big(n\,(X_{(1)} - \theta) \geq t \big)$$

$$= \mathbb{P}_\theta\big(X_{(1)} \geq \theta + t/n \big) = \exp\big\{ -n\,(\theta + t/n - \theta) \big\} = \exp\big\{ -t \big\}\,.$$

Therefore, the likelihood ratio has a representation that satisfies property (ii) in the definition of an asymptotically exponential statistical experiment with $\lambda(\theta) = 1$. Note that in this case, T_n has an exact exponential distribution for any n, and $o_n(1) = 0$.

EXERCISE 4.27 (i) From Exercise 4.22, the estimator θ_n^* is unbiased and its variance is equal to $\theta^2/[n(n+2)]$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}_{\theta_0}\Big[\big(n(\theta_n^* - \theta_0) \big)^2 \Big] = \lim_{n\to\infty} n^2\, \mathbb{V}ar_{\theta_0}\big[\theta_n^* \big] = \lim_{n\to\infty} \frac{n^2\,\theta_0^2}{n(n+2)} = \theta_0^2.$$

(ii) From Exercise 4.23, θ_n^* is unbiased and its variance is equal to $1/n^2$. Hence,

$$\mathbb{E}_{\theta_0}\Big[\,\big(n(\theta_n^* - \theta_0)\big)^2\,\Big] \;=\; n^2\,\mathbb{V}ar_{\theta_0}\big[\,\theta_n^*\,\big] \;=\; \frac{n^2}{n^2} \;=\; 1.$$

EXERCISE 4.28 Consider the case $y \,\le\, 0$. Then

$$\lambda_0 \min_{y\,\le\,0} \int_0^\infty |u \,-\, y|\, e^{-\,\lambda_0\, u}\, du \;=\; \lambda_0 \min_{y\,\le\,0} \int_0^\infty (u \,-\, y)\, e^{-\,\lambda_0\, u}\, du$$

$$=\; \min_{y\,\le\,0}\Big(\,\frac{1}{\lambda_0} \,-\, y\Big) \;=\; \frac{1}{\lambda_0}\,, \quad \text{attained at } y = 0.$$

In the case $y \,\ge\, 0\,$,

$$\lambda_0 \min_{y\,\ge\,0} \int_0^\infty |u \,-\, y|\, e^{-\,\lambda_0\, u}\, du$$

$$=\; \lambda_0 \min_{y\,\ge\,0}\,\Big(\,\int_y^\infty (u \,-\, y)\, e^{-\,\lambda_0\, u}\, du \,+\, \int_0^y (y \,-\, u)\, e^{-\,\lambda_0\, u}\, du\Big)$$

$$=\; \min_{y\,\ge\,0}\,\Big(\,\frac{2\, e^{-\,\lambda_0\, y} \,-\, 1}{\lambda_0} \,+\, y\Big) \;=\; \frac{\ln 2}{\lambda_0}\,,$$

attained at $y = \ln 2/\lambda_0$.

Thus,

$$\lambda_0 \min_{y\in\mathbb{R}} \int_0^\infty |u \,-\, y|\, e^{-\,\lambda_0\, u}\, du \;=\; \min\Big(\,\frac{\ln 2}{\lambda_0}\,,\, \frac{1}{\lambda_0}\,\Big) \;=\; \frac{\ln 2}{\lambda_0}\,.$$

EXERCISE 4.29 (i) For a normalizing constant C, we write by definition

$$f_b(\theta\,|\,X_1\,, \dots\,, X_n) \;=\; C\, f(X_1,\, \theta)\, \dots\, f(X_n,\, \theta)\, \pi_b(\theta)$$

$$=\; C\, \exp\Big\{\,-\,\sum_{i\,=\,1}^{n}(X_i - \theta)\,\Big\}\, \mathbb{I}(X_1 \,\ge\, \theta)\; \dots\; \mathbb{I}(X_n \,\ge\, \theta)\, \frac{1}{b}\, \mathbb{I}(0 \,\le\, \theta \,\le\, b)$$

$$=\; C_1\, e^{n\,\theta}\, \mathbb{I}(\,X_{(1)} \,\ge\, \theta\,)\, \mathbb{I}(0 \,\le\, \theta \,\le\, b) \;=\; C_1\, e^{n\,\theta}\, \mathbb{I}(0 \,\le\, \theta \,\le\, Y)$$

where

$$C_1 \;=\; \Big(\,\int_0^Y e^{n\,\theta}\, d\theta\Big)^{-1} \;=\; \frac{n}{\exp\{\,n\,Y\,\} \,-\, 1}\,, \quad Y = \min(X_{(1)}, b).$$

(ii) The posterior mean follows by direct integration,

$$\theta_n^*(b) = \int_0^Y \frac{n\,\theta\, e^{n\,\theta}}{\exp\{\,nY\,\} - 1}\, d\theta = \frac{1}{n}\,\frac{1}{\exp\{\,nY\,\} - 1}\int_0^{nY} t\,e^t\,dt$$

$$= \frac{1}{n}\,\frac{nY\,\exp\{\,nY\,\} - \big(\exp\{\,nY\,\} - 1\big)}{\exp\{\,nY\,\} - 1} = Y - \frac{1}{n} + \frac{Y}{\exp(\,nY\,) - 1}. \quad \square$$

(iii) Consider the last term in the expression for the estimator $\theta_n^*(b)$. Since by our assumption $\theta \geq \sqrt{b}$, we have that $\sqrt{b} \leq Y \leq b$. Therefore, for all large enough b, the deterministic upper bound holds with $\mathbb{P}_\theta$ - probability 1:

$$\frac{Y}{\exp\{\,nY\,\} - 1} \leq \frac{b}{\exp\{\,n\sqrt{b}\,\} - 1} \to 0 \text{ as } b \to \infty\,.$$

Hence the last term is negligible. To prove the proposition, it remains to show that

$$\lim_{b\to\infty} \mathbb{E}_\theta\Big[\, n^2\Big(\,Y - \frac{1}{n} - \theta\Big)^2\Big] = 1\,.$$

Using the definition of Y and the explicit formula for the distribution of $X_{(1)}$, we get

$$\mathbb{E}_\theta\Big[\, n^2\Big(\,Y - \frac{1}{n} - \theta\Big)^2\Big] =$$

$$= \mathbb{E}_\theta\Big[\, n^2\Big(\,X_{(1)} - \frac{1}{n} - \theta\Big)^2\,\mathbb{I}\big(X_{(1)} \leq b\big) + n^2\Big(\,b - \frac{1}{n} - \theta\Big)^2\,\mathbb{I}\big(X_{(1)} \geq b\big)\Big]$$

$$= n^2\int_\theta^b \Big(y - \frac{1}{n} - \theta\Big)^2 n\,e^{-n\,(y-\theta)}\,dy + n^2\Big(\,b - \frac{1}{n} - \theta\Big)^2\,\mathbb{P}_\theta\big(X_{(1)} \geq b\big)$$

$$= \int_0^{n(b-\theta)} (t-1)^2\,e^{-t}\,dt + \Big(\,n(b-\theta) - 1\Big)^2\,e^{-\,n(b-\theta)} \to 1 \text{ as } b \to \infty\,.$$

Here the first term tends to 1, while the second one vanishes as $b \to \infty$, uniformly in $\theta \in [\,\sqrt{b}\,,\, b - \sqrt{b}\,]$.

(iv) We write

$$\sup_{\theta\in\mathbb{R}} \mathbb{E}_\theta\Big[\,\big(\,n\,(\hat{\theta}_n - \theta)\,\big)^2\Big] \geq \int_0^b \frac{1}{b}\,\mathbb{E}_\theta\Big[\,\big(\,n\,(\hat{\theta}_n - \theta)\,\big)^2\Big]\,d\theta$$

$$\geq \frac{1}{b}\int_0^b \mathbb{E}_\theta\Big[\,\big(\,n\,(\theta_n^*(b) - \theta)\,\big)^2\Big]\,d\theta \geq \frac{1}{b}\int_{\sqrt{b}}^{b-\sqrt{b}} \mathbb{E}_\theta\Big[\,\big(\,n\,(\theta_n^*(b) - \theta)\,\big)^2\Big]\,d\theta$$

$$\geq \frac{b - 2\sqrt{b}}{b}\inf_{\sqrt{b}\,\leq\,\theta\,\leq\, b-\sqrt{b}} \mathbb{E}_\theta\Big[\,\big(\,n\,(\theta_n^*(b) - \theta)\,\big)^2\Big].$$

The infimum is whatever close to 1 if b is sufficiently large. Thus, the limit as $b \to \infty$ of the right-hand side equals 1.

Chapter 5

Exercise 5.30 The Bayes estimator θ_n^* is the posterior mean,

$$\theta_n^* = \frac{(1/n)\sum_{\theta=1}^{n} \theta \exp\{L_n(\theta)\}}{(1/n)\sum_{\theta=1}^{n} \exp\{L_n(\theta)\}} = \frac{\sum_{\theta=1}^{n} \theta \exp\{L_n(\theta)\}}{\sum_{\theta=1}^{n} \exp\{L_n(\theta)\}}.$$

Applying Theorem 5.1 and some transformations, we get

$$\theta_n^* = \frac{\sum_{\theta=1}^{n} \theta \exp\{L_n(\theta) - L_n(\theta_0)\}}{\sum_{\theta=1}^{n} \exp\{L_n(\theta) - L_n(\theta_0)\}}$$

$$= \frac{\sum_{j\,:\,1\le j+\theta_0\le n} (j+\theta_0) \exp\{L_n(j+\theta_0) - L_n(\theta_0)\}}{\sum_{j\,:\,1\le j+\theta_0\le n} \exp\{L_n(j+\theta_0) - L_n(\theta_0)\}}$$

$$= \frac{\sum_{j\,:\,1\le j+\theta_0\le n} (j+\theta_0) \exp\{c\,W(j) - c^2\,|\,j\,|\,/\,2\}}{\sum_{j\,:\,1\le j+\theta_0\le n} \exp\{c\,W(j) - c^2\,|\,j\,|\,/\,2\}}$$

$$= \theta_0 + \frac{\sum_{j\,:\,1\le j+\theta_0\le n} j \exp\{c\,W(j) - c^2\,|\,j\,|\,/\,2\}}{\sum_{j\,:\,1\le j+\theta_0\le n} \exp\{c\,W(j) - c^2\,|\,j\,|\,/\,2\}}.$$

Exercise 5.31 We use the definition of $W(j)$ to notice that $W(j)$ has a $\mathcal{N}(0,\,|\,j\,|)$ distribution. Therefore,

$$\mathbb{E}_{\theta_0}\Big[\exp\big\{c\,W(j) - c^2\,|\,j\,|\,/\,2\big\}\Big] = \exp\{-\,c^2\,|\,j\,|\,/\,2\}\,\mathbb{E}_{\theta_0}\Big[\exp\big\{c\,W(j)\big\}\Big]$$

$$= \exp\big\{-\,c^2\,|\,j\,|\,/\,2 + c^2\,|\,j\,|\,/\,2\big\} = 1.$$

The expected value of the numerator in (5.3) is equal to

$$\mathbb{E}_{\theta_0}\Big[\sum_{j\in\mathbb{Z}} j \exp\big\{c\,W(j) - c^2\,|\,j\,|\,/\,2\big\}\Big] = \sum_{j\in\mathbb{Z}} j = \infty.$$

Likewise, the expectation of the denominator is infinite,

$$\mathbb{E}_{\theta_0}\Big[\sum_{j\in\mathbb{Z}} \exp\big\{c\,W(j) - c^2\,|\,j\,|\,/\,2\big\}\Big] = \sum_{j\in\mathbb{Z}} 1 = \infty.$$

Exercise 5.32 Note that

$$-\,K_{\pm} = \int_{-\infty}^{\infty}\Big[\ln\frac{p_0(x\pm\mu)}{p_0(x)}\Big]\,p_0(x)\,dx$$

$$= \int_{-\infty}^{\infty}\Big[\ln\Big(1 + \frac{p_0(x\pm\mu) - p_0(x)}{p_0(x)}\Big)\Big]\,p_0(x)\,dx$$

$$< \int_{-\infty}^{\infty} \Big[\frac{p_0(x \pm \mu) - p_0(x)}{p_0(x)} \Big] p_0(x)\, dx$$

$$\int_{-\infty}^{\infty} \big[p_0(x \pm \mu) - p_0(x) \big]\, dx = 1 - 1 = 0 \,.$$

Here we have applied the inequality $\ln(1+y) < y$, if $y \neq 0$, and the fact that probability densities $p_0(x \pm \mu)$ and $p_0(x)$ integrate to 1.

EXERCISE 5.33 Assume for simplicity that $\tilde{\theta}_n > \theta_0$. By the definition of the MLE, $\Delta L_n(\theta_0\,, \tilde{\theta}_n) = L_n(\tilde{\theta}_n) - L_n(\theta_0) \geq 0$. Also, by Theorem 5.14,

$$\Delta L_n(\theta_0\,, \tilde{\theta}_n) = W(\tilde{\theta}_n - \theta_0) - K_+ (\tilde{\theta}_n - \theta_0) = \sum_{i\,:\,\theta_0 < i \leq \tilde{\theta}_n} \varepsilon_i - K_+ (\tilde{\theta}_n - \theta_0) \,.$$

Therefore, the following inequalities take place

$$\mathbb{P}_{\theta_0} \big(\tilde{\theta}_n - \theta_0 = m \big) \leq \mathbb{P}_{\theta_0} \big(\tilde{\theta}_n - \theta_0 \geq m \big)$$

$$\leq \sum_{l=m}^{\infty} \mathbb{P}_{\theta_0} \Big(\Delta L_n(\theta_0, \theta_0 + l) \geq 0 \Big) = \sum_{l=m}^{\infty} \mathbb{P}_{\theta_0} \Big(\sum_{i=1}^{l} \varepsilon_i \geq K_+ \, l \Big)$$

$$\leq c_1 \sum_{l=m}^{\infty} l^{-(4+\delta)} \leq c_2 \, m^{-(3+\delta)} \,.$$

A similar argument treats the case $\tilde{\theta}_n < \theta_0$. Thus, there exists a positive constant c_3 such that

$$\mathbb{P}_{\theta_0} \big(|\tilde{\theta}_n - \theta_0| = m \big) \leq c_3 \, m^{-(3+\delta)} \,.$$

Consequently,

$$\mathbb{E}_{\theta_0} \Big[|\tilde{\theta}_n - \theta_0|^2 \Big] = \sum_{m=0}^{\infty} m^2 \, \mathbb{P}_{\theta_0} \big(|\tilde{\theta}_n - \theta_0| = m \big) \leq c_3 \sum_{m=0}^{\infty} m^2 \, m^{-(3+\delta)} < \infty.$$

EXERCISE 5.34 We estimate the true change point value by the maximum likelihood method. The log-likelihood function has the form

$$L(\theta) = \sum_{i=1}^{\theta} \Big[X_i \ln(0.4) + (1 - X_i) \ln(0.6) \Big] + \sum_{i=\theta+1}^{30} \Big[X_i \ln(0.7) + (1 - X_i) \ln(0.3) \Big].$$

Plugging in the concrete observations, we obtain the values of the log-likelihood function for different values of θ. They are summarized in the table below.

θ	$L(\theta)$	θ	$L(\theta)$	θ	$L(\theta)$
1	-21.87	11	-19.95	21	-20.53
2	-21.18	12	-20.51	22	-21.09
3	-21.74	13	-21.07	33	-21.65
4	-21.04	14	-20.37	24	-20.96
5	-21.60	25	-20.93	25	-21.52
6	-20.91	16	-20.24	26	-20.83
7	-20.22	17	-19.55	27	-21.39
8	-20.78	18	-20.11	28	-21.95
9	-21.36	19	-20.67	29	-22.51
10	-20.64	20	-19.97	30	-21.81

The log-likelihood function reaches its maximum -19.55 when $\theta = 17$.

EXERCISE 5.35 Consider a set $\mathfrak{X} \subseteq \mathbb{R}$ with the property that the probability of a random variable with the c.d.f. F_1 falling into that set is not equal to the probability of this event for a random variable with the c.d.f. F_2. Note that such a set necessarily exists, because otherwise, F_1 and F_2 would be identically equal. Ideally we would like the set $\mathfrak{X}$ to be as large as possible. That is, we want $\mathfrak{X}$ to be the largest set such that

$$\int_{\mathfrak{X}} dF_1(x) \neq \int_{\mathfrak{X}} dF_2(x)\,.$$

Replacing the original observations X_i by the indicators $Y_i = \mathbb{I}\big(X_i \in \mathfrak{X}\big)$, $i = 1, \dots, n$, we get a model of Bernoulli observations with the probability of a success $p_1 = \int_{\mathfrak{X}} dF_1(x)$ before the jump, and $p_2 = \int_{\mathfrak{X}} dF_2(x)$, afterwards. The method of maximum likelihood may be applied to find the MLE of the change point (see Exercise 5.34).

Chapter 6

Exercise 6.36 Take any event A in the σ-algebra $\mathcal{F}$. Denote by A^c its complement. By definition, A^c belongs to $\mathcal{F}$. Since an empty set can be written as the intersection of A and A^c, it is also $\mathcal{F}$ - measurable.

Exercise 6.37 (i) If $\tau = T$ for some positive integer T, then for any $t \geq 1$, the event $\{\tau = t\}$ is the whole probability space if $t = T$ and is empty if $t \neq T$. In either case, the event $\{\tau = t\} \in \mathcal{F}_t$. To see this, proceed as in the previous exercise. Take any event $A \in \mathcal{F}_t$. Then A^c belongs to $\mathcal{F}_t$ as well, and so do $A \cup A^c$ (the entire set) and $A \cap A^c$ (the empty set). Therefore, τ is a stopping time by definition.

(ii) If $\tau = \min\big\{\, i \,:\, X_i \in [a,b] \,\big\}$, then for any $t \geq 1$, we write

$$\big\{\tau = t\big\} \;=\; \bigcap_{i=1}^{t-1} \Big(\big\{\, X_i < a \,\big\} \cup \big\{\, X_i > b \,\big\}\Big) \bigcap \big\{\, a \leq X_t \leq b \,\big\}.$$

Each of these events belongs to $\mathcal{F}_t$, hence $\{\tau = t\}$ is $\mathcal{F}_t$ - measurable, and thus, τ is a stopping time.

(iii) Consider $\tau = \min(\tau_1, \tau_2)$. Then

$$\big\{\tau = t\big\} \;=\; \Big(\big\{\tau_1 > t\big\} \cap \big\{\tau_2 = t\big\}\Big) \bigcup \Big(\big\{\tau_2 > t\big\} \cap \big\{\tau_1 = t\big\}\Big).$$

As in the proof of Lemma 6.4, the events $\big\{\tau_1 > t\big\} = \big\{\tau_1 \leq t\big\}^c = \Big(\bigcup_{s=1}^t \big\{\tau_1 = s\big\}\Big)^c$, and $\big\{\tau_2 > t\big\} = \Big(\bigcup_{s=1}^t \big\{\tau_2 = s\big\}\Big)^c$ belong to $\mathcal{F}_t$. Events $\big\{\tau_1 = t\big\}$ and $\big\{\tau_2 = t\big\}$ are $\mathcal{F}_t$ - measurable by definition of a stopping time. Consequently, $\big\{\tau = t\big\} \in \mathcal{F}_t$, and τ is a stopping time.

As for $\tau = \max(\tau_1, \tau_2)$, we write

$$\big\{\tau = t\big\} \;=\; \Big(\big\{\tau_1 < t\big\} \cap \big\{\tau_2 = t\big\}\Big) \bigcup \Big(\big\{\tau_2 < t\big\} \cap \big\{\tau_1 = t\big\}\Big)$$

where each of these events is $\mathcal{F}_t$ - measurable. Thus, τ is a stopping time.

(iv) For $\tau = \tau_1 + s$, where τ_1 is a stopping time and s is a positive integer, we get

$$\big\{\tau = t\big\} \;=\; \big\{\tau_1 = t - s\big\}$$

which belongs to $\mathcal{F}_{t-s}$, and therefore, to $\mathcal{F}_t$. Thus, τ is a stopping time.

EXERCISE 6.38 (i) Let $\tau = \max\{\, i \,:\, X_i \in [a, b],\ 1 \le i \le n\}$. The event

$$\{\tau = t\} \;=\; \bigcap_{i=t+1}^{n} \Big(\{\, X_i < a \,\} \cup \{\, X_i > b \,\} \Big) \bigcap \{\, a \le X_t \le b \,\}\,.$$

All events for $i \ge t+1$ are not $\mathcal{F}_t$ - measurable since they depend on observations obtained after time t. Therefore, τ doesn't satisfy the definition of a stopping time. Intuitively, one has to collect all n observations to decide when was the last time an observation fell in a given interval.
(ii) Take $\tau = \tau_1 - s$ with a positive integer s and a given stopping time τ_1. We have

$$\big\{\tau = t\big\} \;=\; \big\{\tau_1 = t + s\big\} \in \mathcal{F}_{t+s} \not\subseteq \mathcal{F}_t\,.$$

Thus, this event is not $\mathcal{F}_t$ - measurable, and τ is not a stopping time. Intuitively, one cannot know s steps in advance when a stopping time τ_1 occurs.

EXERCISE 6.39 (i) Let $\tau = \min\{\, i \,:\, X_1^2 + \cdots + X_i^2 > H\,\}$. Then for any $t \ge 1$,

$$\big\{\tau = t\big\} \;=\; \Big(\bigcap_{i=1}^{t-1} \{\, X_1^2 + \cdots + X_i^2 \le H\} \Big) \bigcap \{\, X_1^2 + \cdots + X_t^2 > H\,\}\,.$$

All of these events are $\mathcal{F}_t$ - measurable, hence τ is a stopping time.

(ii) Note that $X_1^2 + \cdots + X_\tau^2 > H$ since we defined τ this way. Therefore, by Wald's identity (see Theorem 6.5),

$$H \;<\; \mathbb{E}\left[\, X_1^2 + \cdots + X_\tau^2\right] \;=\; \mathbb{E}[\, X_1^2\,]\, \mathbb{E}[\,\tau\,] \;=\; \sigma^2\, \mathbb{E}[\,\tau\,].$$

Thus, $\mathbb{E}[\,\tau\,] > H/\sigma^2$.

EXERCISE 6.40 Let $\mu = \mathbb{E}[\, X_1\,]$. Using Wald's first identity (see Theorem 6.5), we note that

$$\mathbb{E}[\, X_1 + \cdots + X_\tau - \mu\tau\,] \;=\; 0\,.$$

Therefore, we write

$$\mathbb{V}ar[\, X_1 + \cdots + X_\tau - \mu\tau\,] \;=\; \mathbb{E}\Big[\, \big(X_1 + \cdots + X_\tau - \mu\tau\big)^2\,\Big]$$

$$=\; \mathbb{E}\Big[\, \sum_{t=1}^{\infty} \big(X_1 + \cdots + X_t - \mu t\big)^2\, \mathbb{I}(\tau = t)\Big]$$

$$=\; \mathbb{E}\Big[(X_1 - \mu)^2\, \mathbb{I}(\tau \ge 1) + (X_2 - \mu)^2\, \mathbb{I}(\tau \ge 2) + \cdots + (X_t - \mu)^2\, \mathbb{I}(\tau \ge t) + \ldots\,\Big]$$

$$= \sum_{t=1}^{\infty} \mathbb{E}\Big[(X_t - \mu)^2 \, \mathbb{I}(\tau \geq t) \Big] .$$

The random event $\{\tau \geq t\}$ belongs to $\mathcal{F}_{t-1}$. Hence, $\mathbb{I}(\tau \geq t)$ and X_t are independent. Finally, we get

$$\mathbb{V}ar[\, X_1 + \cdots + X_\tau - \mu\tau \,] \, = \, \sum_{t=1}^{\infty} \mathbb{E}\big[\, (X_t - \mu)^2 \,\big] \, \mathbb{P}\big(\tau \geq t\big)$$

$$= \, \mathbb{V}ar[\, X_1 \,] \sum_{t=1}^{\infty} \mathbb{P}\big(\tau \geq t\big) \, = \, \mathbb{V}ar[\, X_1 \,] \, \mathbb{E}[\, \tau \,] .$$

Exercise 6.41 (i) Using Wald's first identity, we obtain

$$\mathbb{E}_\theta[\, \hat{\theta}_\tau \,] \, = \, \frac{1}{h} \, \mathbb{E}_\theta\big[\, X_1 + \cdots + X_\tau \,\big] \, = \, \frac{1}{h} \, \mathbb{E}_\theta[\, X_1 \,] \, \mathbb{E}_\theta[\, \tau \,] \, = \, \frac{1}{h} \, \theta \, h \, = \, \theta .$$

Thus, $\hat{\theta}_\tau$ is an unbiased estimator of θ.

(ii) First note the inequality derived from an elementary inequality $(x+y)^2 \leq 2(x^2 + y^2)$. For any random variables X and Y such that $\mathbb{E}[X] = \mu_X$ and $\mathbb{E}[Y] = \mu_Y$,

$$\mathbb{V}ar[X + Y] = \mathbb{E}\Big[\big((X - \mu_X) + (Y - \mu_Y)\big)^2 \Big]$$

$$\leq \, 2 \Big(\mathbb{E}\big[\, (X - \mu_X)^2 \,\big] \, + \, \mathbb{E}\big[\, (Y - \mu_Y)^2 \,\big] \Big) \, = \, 2 \Big(\mathbb{V}ar[\, X \,] \, + \, \mathbb{V}ar[\, Y \,] \Big) .$$

Applying this inequality, we arrive at

$$\mathbb{V}ar_\theta[\, \hat{\theta}_\tau \,] \, = \, \frac{1}{h^2} \, \mathbb{V}ar_\theta\big[\, X_1 + \cdots + X_\tau - \theta\,\tau + \theta\,\tau \,\big]$$

$$\leq \, \frac{2}{h^2} \Big(\mathbb{V}ar_\theta\big[X_1 + \cdots + X_\tau - \theta\,\tau\,\big] \, + \, \mathbb{V}ar_\theta[\, \theta\,\tau \,] \Big) .$$

Note that $\mathbb{E}_\theta[\, X_1 \,] = \theta$. Using this notation, we apply Wald's second identity from Exercise 6.40 to conclude that

$$\mathbb{V}ar_\theta[\, \hat{\theta}_\tau \,] \, \leq \, \frac{2}{h^2} \Big(\mathbb{V}ar_\theta[\, X_1 \,] \, \mathbb{E}_\theta[\, \tau \,] \, + \, \theta^2 \, \mathbb{V}ar_\theta[\, \tau \,] \Big) \, = \, \frac{2\sigma^2}{h} \, + \, \frac{2\theta^2 \, \mathbb{V}ar_\theta[\, \tau \,]}{h^2} .$$

Exercise 6.42 (i) Applying repeatedly the recursive equation of the autoregressive model (6.7), we obtain

$$X_i \, = \, \theta \, X_{i-1} + \varepsilon_i \, = \, \theta \, \big[\, \theta \, X_{i-2} + \varepsilon_{i-1} \,\big] + \varepsilon_i \, = \, \theta^2 \, X_{i-2} + \theta \, \varepsilon_{i-1} + \varepsilon_i$$

$$= \theta^2\left[\theta X_{i-3}+\varepsilon_{i-2}\right]+\theta\,\varepsilon_{i-1}+\varepsilon_i = \ldots = \theta^{i-1}[\theta X_0+\varepsilon_1]+\theta^{i-2}\varepsilon_2+\ldots+\theta\,\varepsilon_{i-1}+\varepsilon_i$$
$$= \theta^{i-1}\varepsilon_1 + \theta^{i-2}\varepsilon_2 + \ldots + \theta\,\varepsilon_{i-1} + \varepsilon_i$$

since $X_0 = 0$. Alternatively, we can write out the recursive equations (6.7),

$$\begin{aligned} X_1 &= \theta X_0 + \varepsilon_1 \\ X_2 &= \theta X_1 + \varepsilon_2 \\ &\ldots \\ X_{i-1} &= \theta X_{i-2} + \varepsilon_{i-1} \\ X_i &= \theta X_{i-1} + \varepsilon_i\,. \end{aligned}$$

Multiplying the first equation by θ^{i-1}, the second one by θ^{i-2}, and so on, and finally the equation number $i-1$ by θ, and adding up all the resulting identities, we get

$$X_i + \theta X_{i-1} + \ldots + \theta^{i-2}X_2 + \theta^{i-1}X_1$$
$$= \theta X_{i-1} + \ldots + \theta^{i-2}X_2 + \theta^{i-1}X_1 + \theta^{i-1}X_0$$
$$+\varepsilon_i + \theta\,\varepsilon_{i-1} + \ldots + \theta^{i-2}\varepsilon_2 + \theta^{i-1}\varepsilon_1\,.$$

Canceling the like terms and taking into account that $X_0 = 0$, we obtain

$$X_i = \varepsilon_i + \theta\,\varepsilon_{i-1} + \ldots + \theta^{i-2}\varepsilon_2 + \theta^{i-1}\varepsilon_1\,.$$

(ii) We use the representation of X_i from part (i). Since ε_i's are independent $\mathcal{N}(0,\sigma^2)$ random variables, the distribution of X_i is also normal with mean zero and variance

$$\mathbb{V}ar[X_i] = \mathbb{V}ar\left[\varepsilon_i + \theta\,\varepsilon_{i-1} + \ldots + \theta^{i-2}\varepsilon_2 + \theta^{i-1}\varepsilon_1\right]$$
$$= \mathbb{V}ar[\varepsilon_1]\left(1+\theta^2+\cdots+\theta^{2(i-1)}\right) = \sigma^2\,\frac{1-\theta^{2i}}{1-\theta^2}\,.$$

(iii) Since $|\,\theta\,| < 1$, the quantity θ^{2i} goes to zero as i increases, and therefore,

$$\lim_{i\to\infty}\mathbb{V}ar[X_i] = \lim_{i\to\infty}\sigma^2\,\frac{1-\theta^{2i}}{1-\theta^2} = \frac{\sigma^2}{1-\theta^2}\,.$$

(iv) The covariance between X_i and X_{i+j}, $j \geq 0$, is calculated as

$$\mathbb{C}ov[X_i, X_{i+j}] = \mathbb{E}\Big[\left(\varepsilon_i + \theta\,\varepsilon_{i-1} + \ldots + \theta^{i-2}\varepsilon_2 + \theta^{i-1}\varepsilon_1\right)\times$$
$$\times\left(\varepsilon_{i+j} + \theta\,\varepsilon_{i+j-1} + \ldots + \theta^j\varepsilon_i + \theta^{j+1}\varepsilon_{i-1} + \cdots + \theta^{i+j-2}\varepsilon_2 + \theta^{i+j-1}\varepsilon_1\right)\Big]$$
$$= \theta^j\,\mathbb{E}\Big[\left(\varepsilon_i + \theta\,\varepsilon_{i-1} + \ldots + \theta^{i-2}\varepsilon_2 + \theta^{i-1}\varepsilon_1\right)^2\Big]$$
$$= \theta^j\,\mathbb{V}ar[\varepsilon_1]\left(1 + \theta^2 + \ldots + \theta^{2(i-1)}\right) = \sigma^2\,\theta^j\,\frac{1-\theta^{2i}}{1-\theta^2}\,.$$

Chapter 7

Exercise 7.43 The system of normal equations (7.11) takes the form

$$\begin{cases} \hat{\theta}_0 \, n \, + \, \hat{\theta}_1 \, \sum_{i=1}^{n} x_i \; = \; \sum_{i=1}^{n} y_i \\ \hat{\theta}_0 \, \sum_{i=1}^{n} x_i \, + \, \hat{\theta}_1 \, \sum_{i=1}^{n} x_i^2 \; = \; \sum_{i=1}^{n} x_i \, y_i \end{cases}$$

with the solution

$$\hat{\theta}_1 = \frac{n \, \sum_{i=1}^{n} x_i \, y_i \, - \, \left(\sum_{i=1}^{n} x_i\right) \left(\sum_{i=1}^{n} y_i\right)}{n \, \sum_{i=1}^{n} x_i^2 \, - \, \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \, (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

and $\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}$ where $\bar{x} = \sum_{i=1}^{n} x_i / n$ and $\bar{y} = \sum_{i=1}^{n} y_i / n$.

Exercise 7.44 (a) Note that the vector of residuals $(r_1, \dots, r_n)'$ is orthogonal to the span-space $\mathcal{S}$, while $\mathbf{g}_0 = (1, \dots, 1)'$ belongs to this span-space. Thus, the dot product of these vectors must equal to zero, that is, $r_1 + \cdots + r_n = 0$.

Alternatively, as shown in the proof of Exercise 7.43, $\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}$, and therefore,

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (y_i - \hat{\theta}_0 - \hat{\theta}_1 \, x_i) = \sum_{i=1}^{n} (y_i - \bar{y} + \hat{\theta}_1 \bar{x} - \hat{\theta}_1 \, x_i)$$

$$= \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})}_{0} + \hat{\theta}_1 \underbrace{\sum_{i=1}^{n} (\bar{x} - x_i)}_{0} = 0.$$

(b) In a simple linear regression through the origin, the system of normal equations (7.11) is reduced to a single equation

$$\hat{\theta}_1 \sum_{i=1}^{n} x_i^2 \; = \; \sum_{i=1}^{n} x_i \, y_i,$$

hence, the estimate of the slope is

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n} x_i \, y_i}{\sum_{i=1}^{n} x_i^2}.$$

Consider, for instance, three observations $(0,0), (1,1)$, and $(2,1)$. We get $\hat{\theta}_1 = \sum_{i=1}^{3} x_i \, y_i / \sum_{i=1}^{3} x_i^2 = 0.6$ with the residuals $r_1 = 0$, $r_2 = 0.4$, and $r_3 = -0.2$. The sum of the residuals is equal to 0.2.

EXERCISE 7.45 By definition, the covariance matrix $\mathbf{D} = \sigma^2 \, (\mathbf{G}'\mathbf{G})^{-1}$. For the simple linear regression,

$$\mathbf{D} = \sigma^2 \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} = \frac{\sigma^2}{\det \mathbf{D}} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}.$$

By Lemmma 7.6,

$$\mathbb{V}ar_{\boldsymbol{\theta}}\big[\hat{f}_n(x) \,|\, \mathcal{X}\big] = \mathbf{D}_{00} + 2\mathbf{D}_{01}\, x + \mathbf{D}_{11}\, x^2 = \frac{\sigma^2}{\det \mathbf{D}} \Big(\sum_{i=1}^n x_i^2 - 2\big(\sum_{i=1}^n x_i\big)\, x + n\, x^2 \Big).$$

Differentiating with respect to x, we get

$$-2 \sum_{i=1}^n x_i + 2n\, x = 0\,.$$

Hence the minimum is attained at $x = \sum_{i=1}^n x_i / n = \bar{x}$.

EXERCISE 7.46 (i) We write

$$\mathbf{r} \,=\, \mathbf{y} - \hat{\mathbf{y}} \,=\, \mathbf{y} - \mathbf{G}\hat{\boldsymbol{\theta}} \,=\, \mathbf{y} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'\mathbf{y} \,=\, (\mathbf{I}_n - \mathbf{H})\mathbf{y}$$

where $\mathbf{H} = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'$. We see that the residual vector is a linear transformation of a normal vector $\mathbf{y}$, and therefore has a multivariate normal distribution. Its mean is equal to zero,

$$\mathbb{E}_{\boldsymbol{\theta}}[\,\mathbf{r}\,] \,=\, (\mathbf{I}_n - \mathbf{H})\, \mathbb{E}_{\boldsymbol{\theta}}[\,\mathbf{y}\,] \,=\, (\mathbf{I}_n - \mathbf{H})\, \mathbf{G}\boldsymbol{\theta}$$

$$=\, \mathbf{G}\boldsymbol{\theta} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'\mathbf{G}\,\boldsymbol{\theta} \,=\, \mathbf{G}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\theta} \,=\, \mathbf{0}\,.$$

Next, note that the matrix $\mathbf{I}_n - \mathbf{H}$ is symmetric and idempotent. Indeed,

$$(\mathbf{I}_n - \mathbf{H})' \,=\, \big(\mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'\,\big)' \,=\, \mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}' \,=\, \mathbf{I}_n - \mathbf{H}\,,$$

and

$$(\mathbf{I}_n - \mathbf{H})^2 \,=\, \big(\mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'\,\big)\big(\mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}'\,\big)$$
$$=\, \mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\,\mathbf{G}' \,=\, \mathbf{I}_n - \mathbf{H}\,.$$

Using these two properties, we conclude that

$$(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})' \,=\, (\mathbf{I}_n - \mathbf{H})\,.$$

Therefore, the covariance matrix of the residual vector is derived as follows,

$$\mathbb{E}_{\boldsymbol{\theta}}\big[\,\mathbf{r}\mathbf{r}'\,\big] \,=\, \mathbb{E}_{\boldsymbol{\theta}}\big[\,(\mathbf{I}_n - \mathbf{H})\mathbf{y}\mathbf{y}'(\mathbf{I}_n - \mathbf{H})'\,\big] \,=\, (\mathbf{I}_n - \mathbf{H})\mathbb{E}_{\boldsymbol{\theta}}\big[\,\mathbf{y}\mathbf{y}'\,\big](\mathbf{I}_n - \mathbf{H})'$$

$$= (\mathbf{I}_n - \mathbf{H})\,\sigma^2\,\mathbf{I}_n\,(\mathbf{I}_n - \mathbf{H})' \,=\, \sigma^2\,(\mathbf{I}_n - \mathbf{H})\,.$$

(ii) The vectors $\mathbf{r}$ and $\hat{\mathbf{y}} - \mathbf{G}\,\boldsymbol{\theta}$ are orthogonal since the vector of residuals is orthogonal to any vector that lies in the span-space $\mathcal{S}$. As shown in part (i), $\mathbf{r}$ has a multivariate normal distribution. By the definition of the linear regression model (7.7), the vector $\hat{\mathbf{y}} - \mathbf{G}\boldsymbol{\theta}$ is normally distributed as well. Therefore, being orthogonal and normal, these two vectors are independent.

EXERCISE 7.47 Denote by $\varphi(t)$ the moment generating function of the variable Y. Since X and Y are assumed independent, the moment generating functions of X, Y, and Z satisfy the identity

$$(1-2t)^{-n/2} \,=\, (1-2t)^{-m/2}\,\varphi(t)\,, \quad \text{for } \; t < 1/2\,.$$

Therefore, $\varphi(t) \,=\, (1-2t)^{-(n-m)/2}$, implying that Y has a chi-squared distribution with $n-m$ degrees of freedom.

EXERCISE 7.48 By the definition of a regular deterministic design,

$$\frac{1}{n} = \frac{i}{n} - \frac{i-1}{n} = F_X(x_i) \,-\, F_X(x_{i-1}) \,=\, p\,(x_i^*)\,(x_i \,-\, x_{i-1})$$

for an intermediate point $x_i^* \in (x_{i-1}, x_i)$. Therefore, we may write

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(x_i) \,=\, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{i-1})p(x_i^*)g(x_i) \,=\, \int_0^1 g(x)p(x)\,dx\,.$$

EXERCISE 7.49 Consider the matrix $\mathbf{D}_\infty^{-1}$ with the (l,m)-th entry $\sigma^2\int_0^1 x^l x^m\,dx$, where $l, m = 0, \dots, k$. To show that it is positive definite, we take a column-vector $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_k)'$ and write

$$\boldsymbol{\lambda}'\,\mathbf{D}_\infty^{-1}\,\boldsymbol{\lambda} \,=\, \sigma^2\sum_{i=0}^{k}\sum_{j=0}^{k}\lambda_i\,\lambda_j\int_0^1 x^i\,x^j\,dx \,=\, \sigma^2\int_0^1\Big(\sum_{i=0}^{k}\lambda_i\,x^i\Big)^2 dx\,,$$

which is equal to zero if and only if $\lambda_i = 0$ for all $i = 0, \dots, k$. Hence, $\mathbf{D}_\infty^{-1}$ is positive definite by definition, and thus invertible.

EXERCISE 7.50 By Lemma 7.6, for any design $\mathcal{X}$, the conditional expectation is equal to

$$\mathbb{E}_{\boldsymbol{\theta}}\Big[\,\big(\hat{f}_n(x) - f(x)\big)^2\,|\,\mathcal{X}\,\Big] \,=\, \sum_{l,m=0}^{k}\mathbf{D}_{l,m}\,g_l(x)\,g_m(x)\,.$$

The same equality is valid for the unconditional expectation, since $\mathcal{X}$ is a fixed non-random design. Using the fact that $n\,\mathbf{D} \to \mathbf{D}_\infty$ as $n \to \infty$, we obtain

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\theta}}\Big[\,\big(\sqrt{n}\,(\,\hat{f}_n(x) - f(x)\,)\,\big)^2\,\Big] \;=\; \lim_{n\to\infty} \sum_{l,\,m\,=\,0}^{k} n\,\mathbf{D}_{l,\,m}\, g_l(x)\, g_m(x)$$

$$= \sum_{l,\,m\,=\,0}^{k} (\mathbf{D}_\infty)_{l,\,m}\, g_l(x)\, g_m(x)\,.$$

Exercise 7.51 If all the design points belong to the interval $(1/2, 1)$, then the vector $\eth_0 = (1, \dots, 1)'$ and $\eth_1 = (1/2, \dots, 1/2)'$ are co-linear. The probability of this event is $1/2^n$. If at least one design point belongs to $(0, 1/2)$, then the system of normal equations has a unique solution.

Exercise 7.52 The Hoeffding inequality claims that if ξ_i's are zero-mean independent random variables and $|\xi_i| \le C$, then

$$\mathbb{P}\big(|xi_1 + \cdots + \xi_| > t\big) \le 2\exp\big\{-t^2/(2nC^2)\big\}\,.$$

We apply this inequality to $\xi_i = g_l(x_i)g_m(x_i) - \int_0^1 g_l(x)g_m(x)\,dx$ with $t = \delta n$ and $C = C_0^2$. The result of the lemma follows.

Exercise 7.53 By Theorem 7.5, the distribution of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is $(k+1)$-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$. We know that for regular random designs, $n\mathbf{D}$ goes to a deterministic limit $\mathbf{D}_\infty$, independent of the design. Thus, the unconditional covariance matrix (averaged over the distribution of the design points) goes to the same limiting matrix $\mathbf{D}_\infty$.

Exercise 7.54 Using the Cauchy-Schwarz inequality and Theorem 7.5, we obtain

$$\mathbb{E}_{\boldsymbol{\theta}}\Big[\|\,\hat{f}_n \;-\; f\,\|_2^2 \,|\, \mathcal{X}\,\Big] \;=\; \mathbb{E}_{\boldsymbol{\theta}}\Big[\int_0^1 \Big(\sum_{i\,=\,0}^{k}(\hat{\theta}_i - \theta_i)\, g_i(x)\,\Big)^2 dx \,|\, \mathcal{X}\,\Big]$$

$$\le \mathbb{E}_{\boldsymbol{\theta}}\Big[\sum_{i\,=\,0}^{k}(\hat{\theta}_i - \theta_i)^2 \,|\, \mathcal{X}\,\Big] \sum_{i\,=\,0}^{k}\int_0^1 \big(g_i(x)\big)^2\, dx \;=\; \sigma^2\, tr(\mathbf{D})\,\|\,\mathbf{g}\,\|_2^2\,.$$

Chapter 8

Exercise 8.55 (i) Consider the quadratic loss at a point

$$w\big(\,\hat{f}_n - f\,\big) \;=\; \big(\,\hat{f}_n(x) - f(x)\,\big)^2\,.$$

The risk that corresponds to this loss function (the mean squared error) satisfies

$$R_n(\hat{f}_n,\, f) \;=\; \mathbb{E}_f\big[\,w(\hat{f}_n - f)\,\big] \;=\; \mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x) - f(x)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x) \;-\; \mathbb{E}_f\big[\hat{f}_n(x)\big] \;+\; \mathbb{E}_f\big[\hat{f}_n(x)\big] \;-\; f(x)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x) \;-\; \mathbb{E}_f\big[\hat{f}_n(x)\big]\,\big)^2\,\Big] \;+\; \mathbb{E}_f\Big[\,\big(\,\mathbb{E}_f\big[\hat{f}_n(x)\big] \;-\; f(x)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\big[\,\xi_n^2(x)\,\big] \;+\; b_n^2(x) \;=\; \mathbb{E}_f\big[w(\xi_n)\big] \;+\; w(b_n)\,.$$

The cross term in the above disappears since

$$\mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x) \;-\; \mathbb{E}_f\big[\hat{f}_n(x)\big]\,\big)\,\big(\,\mathbb{E}_f\big[\hat{f}_n(x)\big] \;-\; f(x)\,\big)\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\hat{f}_n(x) \;-\; \mathbb{E}_f\big[\hat{f}_n(x)\big]\,\Big]\,\big(\,\mathbb{E}_f\big[\hat{f}_n(x)\big] \;-\; f(x)\,\big)$$

$$=\; \big(\,\mathbb{E}_f\big[\,\hat{f}_n(x)\,\big] \;-\; \mathbb{E}_f\big[\hat{f}_n(x)\big]\,\big)\,b_n(x) \;=\; 0\,.$$

(ii) Take the mean squared difference

$$w\big(\,\hat{f}_n - f\,\big) \;=\; \frac{1}{n}\sum_{i=1}^{n}\,\big(\,\hat{f}_n(x_i) \;-\; f(x_i)\,\big)^2\,.$$

The risk function (the discrete MISE) can be partitioned as follows.

$$R_n(\hat{f}_n,\, f) \;=\; \mathbb{E}_f\big[\,w(\hat{f}_n - f)\,\big] \;=\; \mathbb{E}_f\Big[\,\frac{1}{n}\sum_{i=1}^{n}\,\big(\,\hat{f}_n(x_i) - f(x_i)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\frac{1}{n}\sum_{i=1}^{n}\,\big(\,\hat{f}_n(x_i) \;-\; \mathbb{E}_f\big[\hat{f}_n(x_i)\big] \;+\; \mathbb{E}_f\big[\hat{f}_n(x_i)\big] \;-\; f(x_i)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\frac{1}{n}\sum_{i=1}^{n}\,\big(\,\hat{f}_n(x_i) - \mathbb{E}_f\big[\hat{f}_n(x_i)\big]\,\big)^2\,\Big] + \mathbb{E}_f\Big[\,\frac{1}{n}\sum_{i=1}^{n}\,\big(\,\mathbb{E}_f\big[\hat{f}_n(x_i)\big] - f(x_i)\,\big)^2\,\Big]$$

$$=\; \mathbb{E}_f\Big[\,\frac{1}{n}\sum_{i=1}^{n}\,\xi_n^2(x_i)\,\Big] \;+\; \frac{1}{n}\sum_{i=1}^{n}\,b_n^2(x_i) \;=\; \mathbb{E}_f\big[\,w(\xi_n)\,\big] \;+\; w(b_n)\,.$$

In the above, the cross term is equal to zero, because for any $i = 1, \dots, n$,

$$\mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x_i)\,-\,\mathbb{E}_f\,[\hat{f}_n(x_i)\,]\,\big)\,\big(\,\mathbb{E}_f\,[\hat{f}_n(x_i)\,]\,-\,f(x_i)\,\big)\,\Big]$$

$$=\,\mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x_i)\,-\,\mathbb{E}_f\,[\hat{f}_n(x_i)\,]\,\big)\,\Big]\,\big(\,\mathbb{E}_f\,[\hat{f}_n(x_i)\,]\,-\,f(x_i)\,\big)$$

$$=\,\big(\mathbb{E}_f\,[\,\hat{f}_n(x_i)\,]\,-\,\mathbb{E}_f\,[\,\hat{f}_n(x_i)\,]\,\big)\,b_n(x_i)\,=\,0\,.$$

Exercise 8.56 Take a linear estimator of f,

$$\hat{f}_n(x)\,=\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,y_i\,.$$

Its conditional bias, given the design $\mathcal{X}$, is computed as

$$b_n(x,\mathcal{X})\,=\,\mathbb{E}_f\,[\,\hat{f}_n(x)\,|\,\mathcal{X}\,]\,-\,f(x)\,=\,\mathbb{E}_f\,[\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,y_i\,|\,\mathcal{X}\,]\,-\,f(x)$$

$$=\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\mathbb{E}_f\,[\,y_i\,|\,\mathcal{X}\,]\,-\,f(x)\,=\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,f(x_i)\,-\,f(x)\,.$$

The conditional variance satisfies

$$\mathbb{E}_f\,[\,\xi_n^2(x,\mathcal{X})\,|\,\mathcal{X}\,]\,=\,\mathbb{E}_f\Big[\,\big(\,\hat{f}_n(x)\,-\,\mathbb{E}_f\,[\,\hat{f}_n(x)\,|\,\mathcal{X}\,]\,\big)^2\,|\,\mathcal{X}\,\Big]$$

$$=\,\mathbb{E}_f\Big[\,\hat{f}_n^2(x)\,|\,\mathcal{X}\,\Big]\,-\,2\,\big(\,\mathbb{E}_f\,[\,\hat{f}_n(x)\,|\,\mathcal{X}\,]\,\big)^2\,+\,\big(\,\mathbb{E}_f\,[\,\hat{f}_n(x)\,|\,\mathcal{X}\,]\,\big)^2$$

$$=\,\mathbb{E}_f\Big[\,\hat{f}_n^2(x)\,|\,\mathcal{X}\,\Big]\,-\,\big(\,\mathbb{E}_f\,[\,\hat{f}_n(x)\,|\,\mathcal{X}\,]\,\big)^2$$

$$=\,\mathbb{E}_f\Big[\,\Big(\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,y_i\Big)^2\,|\,\mathcal{X}\,\Big]\,-\,\Big(\,\mathbb{E}_f\Big[\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,y_i\,\,|\,\mathcal{X}\,\Big]\Big)^2$$

$$=\,\sum_{i=1}^{n}\upsilon_{n,\,i}^2(x)\,\mathbb{E}_f\,[\,y_i^2\,|\,\mathcal{X}\,]\,-\,\Big(\,\sum_{i=1}^{n}\upsilon_{n,\,i}(x)\,\mathbb{E}_f\,[\,y_i\,|\,\mathcal{X}\,]\Big)^2$$

Here the cross terms are negligible since for a given design, the responses are uncorrelated. Now we use the facts that $\mathbb{E}_f\,[\,y_i^2\,|\,\mathcal{X}\,]\,=\,\sigma^2$ and $\mathbb{E}_f\,[\,y_i\,|\,\mathcal{X}\,]\,=\,0$ to arrive at

$$\mathbb{E}_f\,[\,\xi_n^2(x,\mathcal{X})\,|\,\mathcal{X}\,]\,=\,\sigma^2\,\sum_{i=1}^{n}\upsilon_{n,\,i}^2(x)\,.$$

EXERCISE 8.57 (i) The integral of the uniform kernel is computed as

$$\int_{-\infty}^{\infty} K(u)\,du \;=\; \int_{-\infty}^{\infty} (1/2)\,\mathbb{I}\big(-1 \le u \le 1\big)\,du \;=\; \int_{-1}^{1} (1/2)\,du \;=\; 1\,.$$

(ii) For the triangular kernel, we compute

$$\int_{-\infty}^{\infty} K(u)\,du \;=\; \int_{-\infty}^{\infty} (\,1-|u|\,)\,\mathbb{I}\big(-1 \le u \le 1\big)\,du$$

$$=\; \int_{-1}^{0} (1+u)\,du \;+\; \int_{0}^{1} (1-u)\,du \;=\; 1/2+1/2 \;=1\,.$$

(iii) For the bi-square kernel, we have

$$\int_{-\infty}^{\infty} K(u)\,du \;=\; \int_{-\infty}^{\infty} (15/16)\,(1-u^2)^2\,\mathbb{I}\big(-1 \le u \le 1\big)\,du$$

$$=\; (15/16)\int_{-1}^{1} (1-u^2)^2\,du \;=\; (15/16)\int_{-1}^{1} (1-2u^2+u^4)\,du$$

$$=\; (15/16)\,\big(u-(2/3)u^3+(1/5)u^5\big)\Big|_{-1}^{1} \;=\; (15/16)\big(2-(2/3)(2)+(1/5)(2)\big)$$

$$=\; (15/16)(2-4/3+2/5) \;=\; (15/16)(30/15-20/15+6/15) = (15/16)(16/15) \;=\; 1\,.$$

(iv) For the Epanechnikov kernel,

$$\int_{-\infty}^{\infty} K(u)\,du \;=\; \int_{-\infty}^{\infty} (3/4)\,(1-u^2)\,\mathbb{I}\big(-1 \le u \le 1\big) \;=\; (3/4)\int_{-1}^{1} (1-u^2)\,du$$

$$=\; (3/4)\big(u-(1/3)u^3\big)\Big|_{-1}^{1} \;=\; (3/4)\big(2-(1/3)(2)\big) \;=\; (3/4)(2-2/3)$$

$$=\; (3/4)(6/3-2/3) \;=\; (3/4)(4/3) \;=\; 1\,.$$

EXERCISE 8.58 Fix a design $\mathcal{X}$. Consider the Nadaraya-Watson estimator

$$\hat{f}_n(x) \;=\; \sum_{i=1}^{n} \upsilon_{n,i}(x)\,y_i \;\text{ where }\; \upsilon_{n,i}(x) \;=\; K\Big(\frac{x_i-x}{h_n}\Big)\Big/\sum_{j=1}^{n} K\Big(\frac{x_j-x}{h_n}\Big).$$

Note that the weights sum up to one, $\sum_{i=1}^{n} \upsilon_{n,i}(x) \;=\; 1$.

(i) By (8.9), for any constant regression function $f(x) = \theta_0$, we have

$$b_n(x,\mathcal{X}) \;=\; \sum_{i=1}^{n} \upsilon_{n,i}(x) f(x_i) - f(x)$$

$$= \sum_{i=1}^{n} \upsilon_{n,\,i}(x)\,\theta_0 \;-\; \theta_0 \;=\; \theta_0\,\Big(\sum_{i=1}^{n} \upsilon_{n,\,i}(x) \;-\; 1\Big) \;=\; 0\,.$$

(ii) For any bounded Lipschitz regression function $f \in \Theta(1, L, L_1)$, the absolute value of the conditional bias is limited from above by

$$\big|\, b_n(x, \mathcal{X}) \,\big| \;=\; \Big|\, \sum_{i=1}^{n} \upsilon_{n,i}(x) f(x_i) - f(x) \Big|$$

$$\le \sum_{i=1}^{n} \upsilon_{n,\,i}(x)\,\big|\, f(x_i) - f(x)\,\big| \le \sum_{i=1}^{n} \upsilon_{n,\,i}(x)\, L\,\big|\, x_i - x\,\big|$$

$$\le \sum_{i=1}^{n} \upsilon_{n,\,i}(x)\, L\, h_n \;=\; L\, h_n\,.$$

EXERCISE 8.59 Consider a polynomial regression function of the order not exceeding $\beta - 1$,

$$f(x) = \theta_0 + \theta_1\, x + \cdots + \theta_m\, x^m\,, \quad m = 1, \ldots, \beta - 1\,.$$

The i-th observed response is $y_i = \theta_0 + \theta_1\, x_i + \cdots + \theta_m\, x_i^m + \varepsilon_i$ where the explanatory variable x_i has a $Uniform(0,1)$ distribution, and ε_i is a $\mathcal{N}(0, \sigma^2)$ random error independent of x_i, $i = 1, \ldots, n$.

Take a smoothing kernel estimator (8.16) of degree $\beta - 1$, that is, satisfying the normalization and orthogonality conditions (8.17). To show that it is an unbiased estimator of $f(x)$, we need to prove that for any $m = 0, \ldots, \beta - 1$,

$$\frac{1}{h_n}\,\mathbb{E}_f\Big[\, x_i^m\, K\Big(\,\frac{x_i - x}{h_n}\,\Big)\Big] \;=\; x^m\,, \;\; 0 < x < 1\,.$$

Recalling that the smoothing kernel $K(u)$ is non-zero only if $|u| \le 1$, we write

$$\frac{1}{h_n}\,\mathbb{E}_f\Big[\, x_i^m\, K\Big(\,\frac{x_i - x}{h_n}\,\Big)\Big] \;=\; \frac{1}{h_n}\int_0^1 x_i^m\, K\Big(\,\frac{x_i - x}{h_n}\,\Big)\, dx_i$$

$$= \frac{1}{h_n}\int_{x-h_n}^{x+h_n} x_i^m\, K\Big(\,\frac{x_i - x}{h_n}\,\Big)\, dx_i \;=\; \int_{-1}^{1} (h_n\, u + x)^m\, K(u)\, du$$

after a substitution $x_i \;=\; h_n\, u \;+\; x$. If $m = 0$,

$$\int_{-1}^{1} (h_n\, u + x)^m\, K(u)\, du \;=\; \int_{-1}^{1} K(u)\, du \;=\; 1\,,$$

by the normalization condition. If $m = 1, \dots, \beta - 1$,

$$\int_{-1}^{1} (h_n\, u + x)^m \, K(u)\, du \;=\; x^m \underbrace{\int_{-1}^{1} K(u)\, du}_{=1} \;+$$

$$+ \sum_{j=1}^{m} \binom{m}{j} h_n^j \, x^{m-j} \underbrace{\int_{-1}^{1} u^m \, K(u)\, du}_{=0} \;=\; x^m .$$

Therefore,

$$\mathbb{E}_f \Big[\frac{1}{nh_n} \sum_{i=1}^{n} y_i \, K\Big(\frac{x_i - x}{h_n} \Big) \Big]$$

$$= \mathbb{E}_f \Big[\frac{1}{nh_n} \sum_{i=1}^{n} \big(\theta_0 + \theta_1 \, x_i + \cdots + \theta_m \, x_i^m + \varepsilon_i \big) \, K\Big(\frac{x_i - x}{h_n} \Big) \Big]$$

$$= \theta_0 + \theta_1 \, x + \cdots + \theta_m \, x^m \;=\; f(x) .$$

Here we also used the facts that x_i and ε_i are independent, and that ε_i has mean zero.

Exercise 8.60 (i) To find the normalizing constant, integrate the kernel

$$\int_{-1}^{1} K(u)\, du \;=\; \int_{-1}^{1} C(\, 1 - |u|^3 \,)^3 \, du \;=\; 2\, C \int_{0}^{1} (1 - u^3)^3 \, du$$

$$= 2\, C \int_{0}^{1} (1 - 3u^3 + 3u^6 - u^9)\, du \;=\; 2\, C \big(u - \frac{3}{4} u^4 + \frac{3}{7} u^7 - \frac{1}{10} u^{10} \big) \Big|_0^1$$

$$= 2\, C \big(1 - \frac{3}{4} + \frac{3}{7} - \frac{1}{10} \big) \;=\; 2\, C \, \frac{81}{140} \;=\; \frac{81}{70} C \;=1 \;\;\Leftrightarrow\;\; C \;=\; \frac{70}{81} .$$

(ii) Note that the tri-cube kernel is symmetric (an even function). Therefore, it is orthogonal to the monomial x (an odd function), but not the monomial x^2 (an even function). Indeed,

$$\int_{-1}^{1} u(\, 1 - |u|^3 \,)^3 \, du \;=\; \int_{-1}^{0} u(1 + u^3)^3 \, du \;+\; \int_{0}^{1} u(1 - u^3)^3 \, du$$

$$= \; - \int_{0}^{1} u(1 - u^3)^3 \, du \;+\; \int_{0}^{1} u(1 - u^3)^3 \, du \;=\; 0 ,$$

whereas

$$\int_{-1}^{1} u^2(\, 1 - |u|^3 \,)^3 \, du \;=\; \int_{-1}^{0} u^2(1 + u^3)^3 \, du \;+\; \int_{0}^{1} u^2(1 - u^3)^3 \, du$$

$$= 2\int_0^1 u(1-u^3)^3\,du \neq 0\,.$$

Hence, the degree of the kernel is 1.

EXERCISE 8.61 (i) To prove that the normalization and orthogonal conditions hold for the kernel $K(u) = 4 - 6u,\ 0 \le u \le 1$, we write

$$\int_0^1 K(u)\,du = \int_0^1 (4-6u)\,du = (4u - 3u^2)\big|_0^1 = 4 - 3 = 1$$

and

$$\int_0^1 uK(u)\,du = \int_0^1 u(4-6u)\,du = (2u^2 - 2u^3)\big|_0^1 = 2 - 2 = 0.$$

(ii) Similarly, for the kernel $K(u) = 4 + 6u,\ -1 \le u \le 0$,

$$\int_{-1}^0 K(u)\,du = \int_{-1}^0 (4+6u)\,du = (4u + 3u^2)\big|_{-1}^0 = 4 - 3 = 1$$

and

$$\int_{-1}^0 uK(u)\,du = \int_{-1}^0 u(4+6u)\,du = (2u^2 + 2u^3)\big|_{-1}^0 = -2 + 2 = 0.$$

EXERCISE 8.62 (i) We will look for the family of smoothing kernels $K_\theta(u)$ in the class of linear functions with support $[-\theta, 1]$. Let

$$K_\theta(u) = A_\theta\, u + B_\theta\,, \quad -\theta \le u \le 1.$$

The constants A_θ and B_θ are functions of θ and can be found from the normalization and orthogonality conditions. They satisfy

$$\begin{cases} \int_{-\theta}^1 \big(A_\theta\, u + B_\theta \big)\,du = 1 \\ \int_{-\theta}^1 u\big(A_\theta\, u + B_\theta \big)\,du = 0\,. \end{cases}$$

The solution of this system is

$$A_\theta = -6\,\frac{1-\theta}{(1+\theta)^3} \quad \text{and} \quad B_\theta = 4\,\frac{1+\theta^3}{(1+\theta)^4}\,.$$

Therefore, the smoothing kernel has the form

$$K_\theta(u) = 4\,\frac{1+\theta^3}{\theta(1+\theta)^4} - 6u\,\frac{1-\theta}{(1+\theta)^3}\,, -\theta \le u \le 1\,.$$

Note that a linear kernel satisfying the above system of constaints is unique. Therefore, for $\theta = 0$, the kernel $K_\theta(u) = 4 - 6u, 0 \le u \le 1$, as is expected from Exercise 8.61 (i). If $\theta = 1$, then $K_\theta(u)$ turns into the uniform kernel $K_\theta(u) = 1/2,\ -1 \le u \le 1$.

The smoothing kernel estimator

$$\hat{f}_n(x) \;=\; \hat{f}_n(\theta h_n) \;=\; \frac{1}{nh_n} \sum_{i=1}^{n} y_i \, K_\theta\Big(\frac{x_i - \theta h_n}{h_n} \Big)$$

utilizes all the observations with the design points between 0 and $x + h_n$, since

$$\Big\{ -\theta \le \frac{x_i - \theta h_n}{h_n} \le 1 \Big\} = \big\{ 0 \le x_i \le \theta h_n + h_n \big\} \;=\; \big\{ 0 \le x_i \le x + h_n \big\}.$$

(ii) Take the smoothing kernel $K_\theta(u),\ -\theta \le u \le 1$, from part (i). Then the estimator that corresponds to the kernel $K_\theta(-u),\ -1 \le u \le \theta$, at the point $x = 1 - \theta h_n$, uses all the observations with the design points located between $x - h_n$ and 1. It is so, because

$$\Big\{ -1 \le \frac{x_i - x}{h_n} \le \theta \Big\} = \Big\{ -1 \le \frac{x_i - 1 + \theta h_n}{h_n} \le \theta \Big\}$$

$$= \big\{ 1 - \theta h_n - h_n \le x_i \le 1 \big\} \;=\; \big\{ x - h_n \le x_i \le 1 \big\}.$$

Chapter 9

EXERCISE 9.63 If h_n does not vanish as $n \to \infty$, the bias of the local polynomial estimator stays finite. If nh_n is finite, the number of observations N within the interval $[x - h_n, x + h_n]$ stays finite, and can be even zero. Then the system of normal equations (9.2) either does not have a solution or the variance of the estimates does not decrease as n grows.

EXERCISE 9.64 Using Proposition 9.4 and the Taylor expansion (8.14), we obtain

$$\hat{f}_n(0) = \sum_{m=0}^{\beta-1} (-1)^m \, \hat{\theta}_m = \Big(\sum_{m=0}^{\beta-1} (-1)^m \, \frac{f^{(m)}(0)}{m!} h_n^m + \rho(0, h_n) \Big) - \rho(0, h_n) +$$

$$+ \sum_{m=0}^{\beta-1} (-1)^m \big(b_m + \mathcal{N}_m \big) = f(0) - \rho\,(0, h_n) + \sum_{m=0}^{\beta-1} (-1)^m \, b_m + \sum_{m=0}^{\beta-1} (-1)^m \, \mathcal{N}_m \, .$$

Hence the absolute conditional bias of $\hat{f}_n(0)$ for a given design $\mathcal{X}$ admits the upper bound

$$\Big| \mathbb{E}_f \big[\, \hat{f}_n(0) - f(0) \, \big] \Big| \leq \big| \, \rho\,(0, h_n) \, \big| + \sum_{m=0}^{\beta-1} \big| \, b_m \, \big| \leq \frac{L h_n^\beta}{(\beta - 1)!} + \beta \, C_b \, h_n^\beta = O(h_n^\beta) \, .$$

Note that the random variables $\mathcal{N}_m$ can be correlated. That is why the conditional variance of $\hat{f}_n(0)$, given a design $\mathcal{X}$, may not be computed explicitly but only estimated from above by

$$\mathbb{V}ar_f \big[\, \hat{f}_n(0) \, \big| \, \mathcal{X} \, \big] = \mathbb{V}ar_f \Big[\, \sum_{m=0}^{\beta-1} (-1)^m \, \mathcal{N}_m \, \Big| \, \mathcal{X} \, \Big]$$

$$\leq \beta \sum_{m=0}^{\beta-1} \mathbb{V}ar_f \big[\mathcal{N}_m \, \big| \, \mathcal{X} \, \big] \leq \beta \, C_v / N = O(1/N) \, .$$

EXERCISE 9.65 Applying Proposition 9.4, we find that the bias of $m! \, \hat{\theta}_m / (h_n^*)^m$ has the magnitude $O\big(\, (h_n^*)^{\beta - m} \, \big)$, while the random term $m! \, \mathcal{N}_m / (h_n^*)^m$ has the variance $O\big(\, (h_n^*)^{-2m} \, (n \, h_n^*)^{-1} \, \big)$. These formulas guarantee the optimality of $h_n^* = n^{-1/(2\beta+1)}$. Indeed, for any m,

$$(h_n^*)^{2(\beta - m)} = (h_n^*)^{-2m} \, (n \, h_n^*)^{-1} \, .$$

So, the rate $(h_n^*)^{2(\beta-m)} = n^{-2(\beta-m)/(2\beta+1)}$ follows.

EXERCISE 9.66 We proceed by contradiction. Assume that the matrix $\mathbf{D}_\infty^{-1}$ is not invertible. Then there exists a set of numbers $\lambda_0, \dots, \lambda_{\beta-1}$, not all of which are zeros, such that the quadratic form defined by this matrix is equal to zero,

$$0 = \sum_{l,\,m=0}^{\beta-1} \big(\mathbf{D}_\infty^{-1}\big)_{l,\,m} \lambda_l \lambda_m = \frac{1}{2} \sum_{l,\,m=0}^{\beta-1} \lambda_l \lambda_m \int_{-1}^{1} u^{l+m}\, du$$

$$= \frac{1}{2} \int_{-1}^{1} \Big(\sum_{l=0}^{\beta-1} \lambda_l u^l \Big)^2 du\,.$$

On the other hand, the right-hand side is strictly positive, which is a contradiction, and thus, $\mathbf{D}_\infty^{-1}$ is invertible.

EXERCISE 9.67 (i) Let $\mathbb{E}[\,\cdot\,]$ and $\mathbb{V}ar[\,\cdot\,]$ denote the expected value and variance with respect to the distribution of the design points. Using the continuity of the design density $p(x)$, we obtain the explicit formulas

$$\mathbb{E}\Big[\frac{1}{nh_n^*}\sum_{i=1}^{n} \varphi^2\Big(\frac{x_i - x}{h_n^*}\Big)\Big] = \frac{1}{h_n^*}\int_0^1 \varphi^2\Big(\frac{t-x}{h_n^*}\Big) p(t)\, dt$$

$$= \int_0^1 \varphi^2(u) p(x + h_n u)\, du \to p(x) \| \varphi \|_2^2\,.$$

(ii) Applying the fact that $(h_n^*)^{4\beta} = 1/(n\, h_n^*)^2$ and the independence of the design points, we conclude that the variance is equal to

$$\mathbb{V}ar\Big[\sum_{i=1}^{n} f_1^2(x_i)\Big] = \sum_{i=1}^{n} \mathbb{V}ar\big[f_1^2(x_i)\big]$$

$$\le \sum_{i=1}^{n} \mathbb{E}\big[f_1^4(x_i)\big] = \frac{1}{(nh_n^*)^2} \sum_{i=1}^{n} \mathbb{E}\Big[\varphi^4\Big(\frac{x_i - x}{h_n^*}\Big)\Big]$$

$$= \frac{1}{nh_n^*}\int_{-1}^{1} \varphi^4(u) p(x + u h_n^*)\, du \le \frac{1}{nh_n^*} \max_{-1 \le u \le 1} \varphi^4(u)\,.$$

Since $nh_n^* \to \infty$, the variance of the random sum $\sum_{i=1}^{n} f_1^2(x_i)$ vanishes as $n \to \infty$.

(iii) From parts (i) and (ii), the random sum converges in probability to the positive constant $p(x)\| \varphi \|_2^2$. Thus, by the Markov inequality, for all large enough n,

$$\mathbb{P}\Big(\sum_{i=1}^{n} f_1^2(x_i) \le 2p(x)\| \varphi \|_2^2\Big) \ge 1/2\,.$$

EXERCISE 9.68 The proof for a random design $\mathcal{X}$ follows the lines of that in Theorem 9.16, conditionally on $\mathcal{X}$. It brings us directly to the analogue of inequalities (9.11) and (9.14),

$$\sup_{f \in \Theta(\beta)} \mathbb{E}_f\big(\hat{f}_n(x) - f(x)\big)^2 \geq \frac{1}{4}(h_n^*)^{2\beta}\varphi^2(0)\mathbb{E}\Big[1 - \Phi\Big(\frac{1}{2\sigma}\big[\sum_{i=1}^{n} f_1^2(x_i)\big]^{1/2}\Big)\Big].$$

Finally, we apply the result of part (iii) of Exercise 9.67, which claim that the latter expectation is strictly positive.

Chapter 10

EXERCISE 10.69 Applying Proposition 10.2, we obtain

$$\frac{d^m\,\hat{f}_n(x)}{d\,x^m} = \sum_{i=m}^{\beta-1} \frac{i!}{(i-m)!}\,\frac{1}{h_n^m}\Big(\frac{f^{(i)}(c_q)}{i!}\,h_n^i + b_{i,q} + \mathcal{N}_{i,q}\Big)\Big(\frac{x - c_q}{h_n}\Big)^{i-m}$$

$$= \sum_{i=m}^{\beta-1} \frac{f^{(i)}(c_q)}{(i-m)!}\,(x - c_q)^{i-m} + \frac{1}{h_n^m}\sum_{i=m}^{\beta-1}\frac{i!}{(i-m)!}\,b_{i,q}\Big(\frac{x - c_q}{h_n}\Big)^{i-m}$$

$$+ \frac{1}{h_n^m}\sum_{i=m}^{\beta-1}\frac{i!}{(i-m)!}\,\mathcal{N}_{i,q}\Big(\frac{x - c_q}{h_n}\Big)^{i-m}.$$

The first term on the right-hand side is the Taylor expansion around c_q of the m-th derivative of the regression function, which differs from $f^{(m)}(x)$ by no more than $O(h_n^{\beta-m})$. As in the proof of Theorem 10.3, the second bias term has the magnitude $O(h_n^{\beta-m})$, where the reduction in the rate is due to the extra factor h_n^{-m} in the front of the sum. Finally, the third term is a normal random variable which variance does not exceed $O\big(h_n^{-2m}\,(nh_n)^{-1}\big)$. Thus the balance equation takes the form

$$h_n^{2(\beta-m)} = \frac{1}{(h_n)^{2m}(nh_n)}.$$

Its solution is $h_n^* = n^{-1/(2\beta+1)}$, and the respective convergence rate is $(h_n^*)^{\beta-m}$.

EXERCISE 10.70 For any $y > 0$,

$$\mathbb{P}\Big(\mathcal{Z}^* \ge y\beta\sqrt{2\ln n}\Big) \le \mathbb{P}\Big(\bigcup_{q=1}^{Q}\bigcup_{m=,0}^{\beta-1}|Z_{m,q}| \ge y\sqrt{2\ln n}\Big)$$

$$\le Q\beta\mathbb{P}\big(|Z| \ge y\sqrt{2\ln n}\big) \quad \text{where } Z \sim \mathcal{N}(0,1)$$

$$\le Q\beta n^{-y^2} \quad \text{since } \mathbb{P}(|Z| \ge x) \le \exp\{-x^2/2\},\ x \ge 1.$$

If $n > 2$ and $y > 2$, then $Qn^{-y^2} \le 2^{-y}$, and hence

$$\mathbb{E}\Big[\frac{\mathcal{Z}^*}{\beta\sqrt{2\ln n}}\,\Big|\,\mathcal{X}\Big] = \int_0^\infty \mathbb{P}\Big(\frac{\mathcal{Z}^*}{\beta\sqrt{2\ln n}} \ge y\,\Big|\,\mathcal{X}\Big)\,dy$$

$$\le \int_0^2 dy + \beta\int_2^\infty 2^{-y}\,dy = 2 + \frac{\beta}{4\ln 2}.$$

Thus (10.11) holds with $C_z = \big(2 + \frac{\beta}{4\ln 2}\big)\beta\sqrt{2}$.

Exercise 10.71 Note that

$$\mathbb{P}\Big(\mathcal{Z}^* \,\geq\, y\,\sqrt{2\beta^2 \ln Q}\Big) \,\leq\, Q\beta Q^{-y^2} \,=\, \beta Q^{-(y^2-1)} \,\leq\, \beta 2^{-y}\,,$$

if $Q \geq 2$ and $y \geq 2$. The rest of the proof follows as in the solution to Exercise 10.70. Further, if we seek to equate the squared bias and the variance terms, the bandwidth would satisfy

$$h_n^{\beta} = \sqrt{(nh_n)^{-1}\ln Q}\,, \quad \text{where} \quad Q \,=\, 1/(2\,h_n)\,.$$

Omitting the constants in this identity, we arrive at the balance equation, which the optimal bandwidth solves,

$$h_n^{\beta} = \sqrt{-(nh_n)^{-1}\ln h_n}\,,$$

or, equivalently,

$$nh_n^{2\beta+1} \,=\, -\ln h_n\,.$$

To solve this equation, put

$$h_n \,=\, \Big(\frac{b_n \ln n}{(2\beta+1)n}\Big)^{1/(2\beta+1)}.$$

Then b_n satisfies the equation

$$b_n \,=\, 1 \,+\, \frac{\ln(2\beta+1) \,-\, \ln b_n \,-\, \ln\ln n}{\ln n}$$

with the asymptotics $b_n \to 1$ as $n \to \infty$.

Exercise 10.72 Consider the piecewise monomial functions given in (10.12),

$$\gamma_{m,\,q}(x) \,=\, \mathbb{I}(\,x \in B_q)\,\Big(\frac{x-c_q}{h_n}\Big)^m, \quad q=1,\dots,Q, \quad m=0,\dots,\beta-1\,. \quad (0.1)$$

The design matrix $\mathbf{\Gamma}$ in (10.16) has the columns

$$\boldsymbol{\gamma}_k = \big(\gamma_k(x_1),\dots,\gamma_k(x_n)\big)', \quad k=m+\beta(q-1), \quad q=1,\dots,Q\,, \quad m=0,\dots,\beta-1. \quad (0.2)$$

The matrix $\mathbf{\Gamma}'\mathbf{\Gamma}$ of the system of normal equations (10.17) is block-diagonal with Q blocks of dimension β each. Under Assumption 10.1, this matrix is invertible. Thus, the dimension of the span-space is $\beta\,Q \,=\, K$.

Exercise 10.73 If β is an even number, then

$$f^{(\beta)}(x) \,=\, \sum_{k=1}^{\infty}(-1)^{\beta/2}(2\pi k)^{\beta}\big[\,a_k\sqrt{2}\cos(2\pi kx) \,+\, b_k\sqrt{2}\sin(2\pi kx)\,\big]\,.$$

If β is an odd number, then

$$f^{(\beta)}(x) \;=\; \sum_{k=1}^{\infty}(-1)^{(\beta+1)/2}(2\pi k)^{\beta}\big[\, a_k\sqrt{2}\cos(2\pi kx) \;-\; b_k\sqrt{2}\sin(2\pi kx)\,\big]\,.$$

In either case,

$$\|\, f^{(\beta)} \,\|_2^2 \;=\; (2\pi)^{\beta}\sum_{k=1}^{\infty} k^{2\beta}\big[\, a_k^2 \;+\; b_k^2\,\big]\,.$$

Exercise 10.74 We will show only that

$$\sum_{i=1}^{n}\sin\big(2\pi mi/n\big) = 0\,.$$

To this end, we use the elementary trigonometric identity

$$2\sin\alpha\sin\beta = \cos(\alpha-\beta) \;-\; \cos(\alpha+\beta)$$

to conclude that

$$\sin\big(2\pi mi/n\big) = \frac{\cos\big(2\pi m(i-1/2)/n\big) \;-\; \cos\big(2\pi m(i+1/2)/n\big)}{2\sin\big(\pi m/n\big)}\,.$$

Thus, we get a telescoping sum

$$\sum_{i=1}^{n}\sin\big(2\pi mi/n\big) \;=\; \sum_{i=1}^{n}\Big[\,\frac{\cos\big(2\pi m(i-1/2)/n\big) \;-\; \cos\big(2\pi m(i+1/2)/n\big)}{2\sin\big(\pi m/n\big)}\Big]$$

$$= \frac{1}{2\sin\big(\pi m/n\big)}\Big[\,\cos\big(\pi m/n\big) \;-\; \cos\big(2\pi m(n+1/2)/n\big)\Big]$$

$$= \frac{1}{2\sin\big(\pi m/n\big)}\Big[\,\cos\big(\pi m/n\big) \;-\; \cos\big(2\pi m+\pi m/n\big)\Big]$$

$$= \frac{1}{2\sin\big(\pi m/n\big)}\Big[\,\cos\big(\pi m/n\big) \;-\; \cos\big(\pi m/n\big)\Big] \;=\; 0\,.$$

Chapter 11

Exercise 11.75 The standard B-spline of order 2 can be computed as

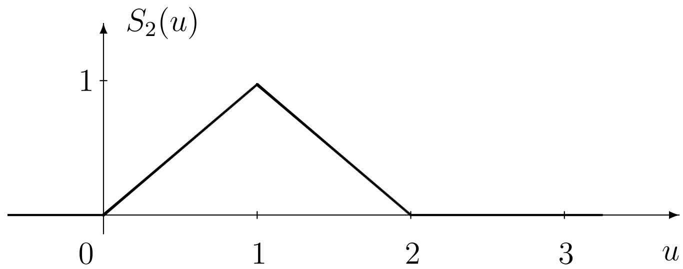
$$S_2(u) = \int_{-\infty}^{\infty} \mathbb{I}_{[0,1)}(z)\,\mathbb{I}_{[0,1)}(u-z)\,dz = \begin{cases} \displaystyle\int_0^u dz = u\,, & \text{if } 0 \le u < 1, \\ \displaystyle\int_{u-1}^1 dz = 2-u\,, & \text{if } 1 \le u < 2. \end{cases}$$

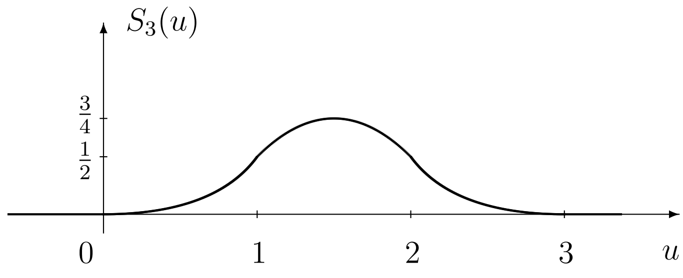
The standard B-spline of order 3 has the form

$$S_3(u) = \int_{-\infty}^{\infty} S_2(z)\,\mathbb{I}_{[0,1)}(u-z)\,dz$$

$$= \begin{cases} \displaystyle\int_0^u z\,dz = \frac{1}{2}\,u^2\,, & \text{if } 0 \le u < 1, \\ \displaystyle\int_{u-1}^1 z\,dz + \int_1^u (2-z)\,dz = -u^2 + 3u - \frac{3}{2}\,, & \text{if } 1 \le u < 2, \\ \displaystyle\int_{u-1}^2 (2-z)\,dz = \frac{1}{2}\,(3-u)^2\,, & \text{if } 2 \le u < 3. \end{cases}$$

Both splines $S_2(u)$ and $S_3(u)$ are depicted in the figure below.





EXERCISE 11.76 For $k = 0$, (11.6) is a tautology. Assume that the statement is true for some $k \geq 0$. Then, applying (11.2), we obtain that

$$S_m^{(k+1)}(u) = \left(S_m^{(k)}(u)\right)' = \sum_{j=0}^{k} (-1)^j \binom{k}{j} S'_{m-k}(u-j)$$

$$= \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left[S_{m-k-1}(u-j) - S_{m-k-1}(u-j-1) \right]$$

$$= \binom{k}{0} S_{m-k-1}(u) + (-1)^1 \left[\binom{k}{1} + \binom{k}{0} \right] S_{m-k-1}(u-1)$$

$$+ \ldots + (-1)^k \left[\binom{k}{k} + \binom{k}{k-1} \right] S_{m-k-1}(u-k) - (-1)^k \binom{k}{k} S_{m-k-1}(u-k-1)$$

$$= \sum_{j=0}^{k+1} (-1)^j \binom{k+1}{j} S_{m-(k+1)}(u-j).$$

Here we used the elementary formulas

$$\binom{k}{j} + \binom{k}{j-1} = \binom{k+1}{j}, \quad \binom{k}{0} = \binom{k+1}{0} = 1,$$

and

$$-(-1)^k \binom{k}{k} = (-1)^{k+1} \binom{k+1}{k+1}.$$

EXERCISE 11.77 Applying Lemma 11.2, we obtain that

$$LS^{(m-1)}(u) = \sum_{i=0}^{m-2} a_i S_m^{(m-1)}(u-i) = \sum_{i=0}^{m-2} a_i \sum_{l=0}^{m-1} (-1)^l \binom{m-1}{l} \mathbb{I}_{[0,1)}(u-i-l).$$

If $u \in [j, j+1)$, then the only non-trivial contribution into the latter sum comes from i and l such that $i + l = j$. In view of the restriction, $0 \leq j \leq m-2$, the double sum in the last formula turns into

$$\lambda_j = \sum_{i=0}^{j} a_i (-1)^{j-i} \binom{m-1}{j-i}.$$

EXERCISE 11.78 If we differentiate j times the function

$$P_k(u) = \frac{(u-k)^{m-1}}{(m-1)!}, \quad u \geq k,$$

we find that

$$P_k^{(j)}(u) \;=\; (u-k)^{m-1-j}\,\frac{(m-1)\,(m-2)\,\ldots\,(m-j)}{(m-1)!} \;=\; \frac{(u-k)^{m-j-1}}{(m-j-1)!}\,.$$

Hence

$$\nu_j \;=\; LP^{(j)}(m-1) \;=\; \sum_{k=0}^{m-2} b_k\,\frac{(m-k-1)^{m-j-1}}{(m-j-1)!}\,.$$

EXERCISE 11.79 The matrix $\mathbf{M}$ has the explicit form,

$$\mathbf{M} \;=\; \begin{bmatrix} \frac{(m-1)^{m-1}}{(m-1)!} & \frac{(m-2)^{m-1}}{(m-1)!} & \cdots & \frac{(1)^{m-1}}{(m-1)!} \\ \frac{(m-1)^{m-2}}{(m-2)!} & \frac{(m-2)^{m-2}}{(m-2)!} & \cdots & \frac{(1)^{m-2}}{(m-2)!} \\ & & \cdots & \\ \frac{(m-1)^{1}}{1!} & \frac{(m-2)^{1}}{1!} & \cdots & \frac{(1)^{1}}{1!} \end{bmatrix}$$

so that its determinant

$$\det\mathbf{M} \;=\; \Big(\prod_{k=1}^{m-1} k!\Big)^{-1} \det\mathbf{V}_{m-1} \;\neq\; 0$$

where $\mathbf{V}_{m-1}$ is the $(m-1)\times(m-1)$ Vandermonde matrix with the elements $x_1 = 1, \ldots, x_{m-1} = m-1$.

EXERCISE 11.80 In view of Lemma 11.4, the proof repeats the proof of Lemma 11.8. The polynomial $g(u) = 1 - u^2$ in the interval $[2, 3)$ has the representation

$$g(u) \;=\; b_0\,P_0(u) + b_1\,P_1(u) + b_2\,P_2(u) \;=\; (-1)\,\frac{u^2}{2!} + (-2)\,\frac{(u-1)^2}{2!} + \frac{(u-2)^2}{2!}$$

with $b_0 = -1$, $b_1 = -2$, and $b_2 = 1$.

EXERCISE 11.81 Note that the derivative of the order $(\beta - j - 1)$ of $f^{(j)}$ is $f^{(\beta-1)}$ which is the Lipschitz function with the Lipschitz constant L by the definition of $\Theta(\beta, L, L_1)$. Thus, what is left to show is that all the derivatives $f^{(1)}, \ldots, f^{(\beta-1)}$ are bounded in their absolute values by some constant L_2. By Lemma 10.2, any function $f \in \Theta(\beta, L, L_1)$ admits the Taylor approximation

$$f(x) \;=\; \sum_{m=0}^{\beta-1} \frac{f^{(m)}(c)}{m!}(x-c)^m \;+\; \rho(x,c)\,, \quad 0 \le x,\, c \le 1\,,$$

with the remainder term $\rho(x,c)$ such that

$$|\,\rho(x,c)\,| \;\leq\; \frac{L\,|x-c|^{\beta}}{(\beta-1)!} \;\leq\; C_{\rho} \;\;\text{where}\;\; C_{\rho} \;=\; \frac{L}{(\beta-1)!}$$

That is why, if $f \in \Theta(\beta, L, L_1)$, then at any point $x = c$, the inequality holds

$$\Big|\,\sum_{m=0}^{\beta-1} \frac{f^{(m)}(c)}{m!}\,(x-c)^m\,\Big| \;\leq\; \big|\,f(x)\,\big| \,+\, \big|\,\rho(x,c)\,\big| \;\leq\; L_1 \,+\, C_{\rho} \;=\; L_0\,.$$

So, it suffices to show that if a polynomial $g(x) \;=\; \sum_{m=0}^{\beta-1} b_m\,(x-c)^m$ is bounded, $|\,g(x)\,| \;=\; \big|\,\sum_{m=0}^{\beta-1} b_m\,(x-c)^m\,\big| \;\leq\; L_0$, for all $x, c \in [0,\,1]$, then

$$\max\,\big[\,b_0, \dots, b_{\beta-1}\,\big] \;\leq\; L_2 \tag{0.3}$$

with a constant L_2 independent of $c \in [0,\,1]$. Assume for definiteness that $0 \;\leq\; c \;\leq\; 1/2$, and choose the points $c \;<\; x_0 \;<\; \cdots \;<\; x_{\beta-1}$ so that $t_i \;=\; x_i \,-\, c \;=\; (i+1)\alpha$, $i \;=\; 0, \dots, \beta-1$. A positive constant α is such that $\alpha\,\beta \;<\; 1/2$, which yields $0 \;\leq\; t_i \;\leq\; 1$. Put $g_i \;=\; g(x_i)$. The coefficients $b_0, \dots, b_{\beta-1}$ of polynomial $g(x)$ satisfy the system of linear equations

$$b_0 \,+\, b_1 t_i \,+\, b_2 t_i^2 \,+\, \dots \,+\, b_{\beta-1} t_i^{\beta-1} \;=\; g_i\,,\;\; i \;=\; 0, \dots, \beta-1\,.$$

The determinant of the system's matrix is the Vandermonde determinant, that is, it is non-zero and independent of c. The right-hand side elements of this system are bounded by L_0. Thus, the upper bound (0.3) follows. Similar considerations are true for $1/2 \;\leq\; c \;\leq\; 1$.

Chapter 12

EXERCISE 12.82 We have n design points in Q bins. That is why, for any design, there exist at least $Q/2$ bins with at most $2n/Q$ design points. Indeed, otherwise we would have strictly more than $(Q/2)(2n/Q) = n$ points. Denote the set of the indices of these bins by $\mathcal{M}$. By definition, $|\mathcal{M}| \geq Q/2$. In each such bin B_q, the respective variance is bounded by

$$\sigma_{q,n}^2 = \sum_{x_i \in B_q} f_q^2(x_i) \leq \sum_{x_i \in B_q} \left(h_n^*\right)^{2\beta} \varphi^2\left(\frac{x_i - c_q}{h_n^*}\right)$$

$$\leq \|\varphi\|_\infty^2 \left(h_n^*\right)^{2\beta}(2n/Q) = 4n\|\varphi\|_\infty^2 \left(h_n^*\right)^{2\beta+1} = 4\|\varphi\|_\infty^2 \ln n$$

which can be made less than $2\alpha \ln Q$ if we choose $\|\varphi\|_\infty$ sufficiently small.

EXERCISE 12.83 Select the test function defined by (12.3). Substitute M in the proof of Lemma 12.11 by Q, to obtain

$$\sup_{f \in \Theta(\beta)} \mathbb{E}_f\left[\psi_n^{-1}\|\hat{f}_n - f\|_\infty\right] \geq d_0\psi_n^{-1} \max_{1 \leq q \leq Q} \mathbb{E}_{f_q}\left[\mathbb{E}_{f_q}\left[\mathbb{I}(\mathcal{D}_q) \,|\, \mathcal{X}\right]\right]$$

$$\geq d_0\psi_n^{-1}\mathbb{E}^{(\mathcal{X})}\left[\frac{1}{2}\mathbb{P}_0\left(\mathcal{D}_0 \,|\, \mathcal{X}\right) + \frac{1}{2Q}\sum_{q=1}^{Q} \mathbb{P}_q\left(\mathcal{D}_q \,|\, \mathcal{X}\right)\right]$$

where $\mathbb{E}^{(\mathcal{X})}[\,\cdot\,]$ denotes the expectation taken over the distribution of the random design.

Note that $d_0\psi_n^{-1} = (1/2)\|\varphi\|_\infty$. Due to (12.22), with probability 1, for any random design $\mathcal{X}$, there exists a set $\mathcal{M}(\mathcal{X})$ such that

$$\frac{1}{2}\mathbb{P}_0\left(\mathcal{D}_0 \,|\, \mathcal{X}\right) + \frac{1}{2Q}\sum_{q=1}^{Q} \mathbb{P}\left(\mathcal{D}_q \,|\, \mathcal{X}\right) \geq \frac{|\mathcal{M}|}{4Q} \geq \frac{Q/2}{4Q} = \frac{1}{8}.$$

Combining these bounds, we get that

$$\sup_{f \in \Theta(\beta)} \mathbb{E}_f\left[\psi_n^{-1}\|\hat{f}_n - f\|_\infty\right] \geq (1/16)\|\varphi\|_\infty.$$

EXERCISE 12.84 The log-likelihood function is equal to

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - f(x_i, \boldsymbol{\omega}')\right)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - f(x_i, \boldsymbol{\omega}'')\right)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\big(y_i - f(x_i,\,\boldsymbol{\omega}'')\big)\big(f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')\big)$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')\big)^2$$

$$= \sum_{i=1}^{n}\Big(\frac{\varepsilon_i}{\sigma}\Big)\Big(\frac{f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')}{\sigma}\Big) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')\big)^2$$

so that (12.24) holds with

$$\sigma_n^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\big(f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')\big)^2$$

and

$$\mathcal{N}_n = \frac{1}{\sigma_n}\sum_{i=1}^{n}\Big(\frac{\varepsilon_i}{\sigma}\Big)\Big(\frac{f(x_i,\,\boldsymbol{\omega}') - f(x_i,\,\boldsymbol{\omega}'')}{\sigma}\Big).$$

EXERCISE 12.85 By definition,

$$\mathbb{E}\big[\exp\big\{z\xi_q'\big\}\big] = \frac{1}{2}e^{z/2} + \frac{1}{2}e^{-z/2} = \sum_{k=0}^{\infty}\frac{1}{(2k)!}\Big(\frac{z}{2}\Big)^{2k}$$

$$= \sum_{k=0}^{\infty}\frac{1}{k!}\frac{1}{(k+1)\dots(k+k)}\Big(\frac{z^2}{4}\Big)^k \le \sum_{k=0}^{\infty}\frac{1}{k!\,2^k}\Big(\frac{z^2}{4}\Big)^k = \sum_{k=0}^{\infty}\frac{1}{k!}\Big(\frac{z^2}{8}\Big)^k = e^{z^2/8}.$$

EXERCISE 12.86 Consider the case $\beta = 1$. The bandwidth $h_n^* = n^{-1/3}$, and the number of the bins $Q = 1/(2h_n^*) = (1/2)n^{1/3}$. Let $N = n/Q = 2n^{2/3}$ denote the number of design points in every bin. We assume that N is an integer. In the bin B_q, $1 \le q \le Q$, the estimator has the form

$$f_n^* = \bar{y}_q = \sum_{i/n\,\in\,B_q} y_i/N = \bar{f}_q + \xi_q/\sqrt{N}$$

with $\bar{f}_q = \sum_{i/n\,\in\,B_q} f(x_i)/N$, and independent $\mathcal{N}(0,\,\sigma^2)$-random variables $\xi_q = \sum_{i/n\,\in\,B_q}\big(y_i - f(x_i)\big)/\sqrt{N} = \sum_{i/n\,\in\,B_q}\varepsilon_i/\sqrt{N}$.

Put $\bar{f}_n(x) = \bar{f}_q$ if $x \in B_q$. From the Lipschitz condition on f it follows that $\|\bar{f}_n - f\|_2^2 \le Cn^{-2/3}$ with some positive constant C independent of n. Next,

$$\|f_n^* - f\|_2^2 \le 2\|\bar{f}_n - f\|_2^2 + 2\|f_n^* - \bar{f}_n\|_2^2$$

$$= 2\|\bar{f}_n - f\|_2^2 + \frac{2}{QN}\sum_{q=1}^{Q} \xi_q^2 = 2\|\bar{f}_n - f\|_2^2 + \frac{2}{n}\sum_{q=1}^{Q} \xi_q^2 ,$$

so that

$$n^{2/3}\|f_n^* - f\|_2^2 \le 2C + 2\frac{n^{2/3}}{n}\sum_{q=1}^{Q} \xi_q^2 = 2C + 2n^{-1/3}\sum_{q=1}^{Q} \xi_q^2 .$$

By the Law of Large Numbers,

$$2n^{-1/3}\sum_{q=1}^{Q} \xi_q^2 = \frac{1}{Q}\sum_{q=1}^{Q} \xi_q^2 \to \sigma^2$$

almost surely as $n \to \infty$. Hence for any constant c such that $c^2 > 2C + \sigma^2$, the inequality holds $n^{1/3}\|f_n^* - f\|_2 \le c$ with probability whatever close to 1 as $n \to \infty$. Thus, there is no p_0 that satisfies

$$\mathbb{P}_f\big(\|\hat{f}_n - f\|_2 \ge cn^{-1/3} \,\big|\, \mathcal{X}\big) \ge p_0 .$$

Chapter 13

Exercise 13.87 The expected value $\mathbb{E}_f[\hat{\Psi}_n] = n^{-1}\sum_{i=1}^{n} w(i/n)f(i/n)$. Since w and f are the Lipschitz functions, their product is also Lipschitz with some constant L_0 so that

$$|b_n| = \big|\,\mathbb{E}_f[\hat{\Psi}_n] - \Psi(f)\,\big| = \Big|\,\mathbb{E}_f[\hat{\Psi}_n] - \int_0^1 w(x)f(x)\,dx\,\Big| \le L_0/n\,.$$

Next, $\hat{\Psi}_n - \mathbb{E}_f[\hat{\Psi}_n] = n^{-1}\sum_{i=1}^{n} w(i/n)\varepsilon_i$, hence the variance of $\hat{\Psi}_n$ equals to

$$\frac{\sigma^2}{n^2}\sum_{i=1}^{n} w^2(i/n) = \frac{\sigma^2}{n}\Big(\int_0^1 w^2(x)\,dx + O\big(n^{-1}\big)\Big).$$

Exercise 13.88 Note that $\Psi(1) = e^{-1}\int_0^1 e^t f(t)\,dt$, thus the estimator (13.4) takes the form

$$\hat{\Psi}_n = n^{-1}\sum_{i=1}^{n} \exp\big\{(i-n)/n\big\}\,y_i\,.$$

By Proposition 13.2, the bias of this estimator has the magnitude $O(n^{-1})$, and its variance is

$$\mathbb{V}ar[\hat{\Psi}_n] = \frac{\sigma^2}{n}\int_0^1 e^{2(t-1)}\,dt + O(n^{-2}) = \frac{\sigma^2}{2n}\big(1-e^{-2}\big) + O(n^{-2})\,, \text{ as } n\to\infty\,.$$

Exercise 13.89 Take any $f_0 \in \Theta(\beta, L, L_1)$, and put $\Delta f = f - f_0$. Note that

$$f^4 = f_0^4 + 4f_0^3(\Delta f) + 6f_0^2(\Delta f)^2 + 4f_0(\Delta f)^3 + (\Delta f)^4\,.$$

Hence

$$\Psi(f) = \Psi(f_0) + \int_0^1 w(x,\,f_0)f(x)\,dx + \rho(f,\,f_0)$$

with a Lipschitz weight function $w(x,\,f_0) = 4f_0^3(x)$, and the remainder term

$$\rho(f_0,\,f) = \int_0^1 \Big(6f_0^2(\Delta f)^2 + 4f_0(\Delta f)^3 + (\Delta f)^4\Big)\,dx\,.$$

Since f_0 and f belong to the set $\Theta(\beta, L, L_1)$, they are bounded by L_1, and, thus, $|\Delta f| \le 2L_1$. Consequently, the remainder term satisfies the condition

$$|\rho(f_0,\,f)| \le \big(6L_1^2 + 4L_1(2L_1) + (2L_1)^2\big)\,\|f-f_0\|_2^2$$
$$= 18L_1^2\|f-f_0\|_2^2 = C_\rho\|f-f_0\|_2^2 \text{ with } C_\rho = 18L_1^2\,.$$

EXERCISE 13.90 From (13.12), we have to verify is that

$$\mathbb{E}_f\Big[\,\big(\sqrt{n}\rho(f,\,f_n^*)\big)^2\,\Big] \to 0 \ \text{ as } \ n \to \infty\,.$$

Under the assumption on the remainder term, this expectation is bounded from above by

$$\mathbb{E}_f\Big[\,\big(\sqrt{n}C_\rho\|\,f_n^* - f\|_2^2\big)^2\,\Big] \;= nC_\rho^2\mathbb{E}_f\Big[\,\Big(\int_0^1 (f_n^*(x) - f(x))^2\,dx\Big)^2\,\Big]$$

$$\le\; nC_\rho^2\mathbb{E}_f\Big[\,\int_0^1 (f_n^*(x) - f(x))^4\,dx\,\Big] \to 0 \ \text{ as } \ n \to \infty\,.$$

EXERCISE 13.91 The expected value of the sample mean is equal to

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) \;=\; \sum_{i=1}^{n} f(x_i)p(x_i)(x_i - x_{i-1})\big(np(x_i)(x_i - x_{i-1})\big)^{-1}$$

$$=\; \int_0^1 f(x)p(x)\big(1 \,+\, o_n(1)\big)\,dx\,,$$

because, as shown in the proof of Lemma9.8, $np(x_i)(x_i - x_{i-1}) \;\to\; 1$ uniformly in $i = 1, \dots, n$. Hence

$$\hat{\Psi}_n \;=\; (y_1 + \cdots + y_n)/n \;\sim\; \mathcal{N}\Big(\,\int_0^1 f(x)\,p(x)\,dx\,,\,\sigma^2/n\,\Big)\,.$$

To prove the efficiency, consider the family of the constant regression functions $f_\theta(x) \;=\; \theta\,,\ \theta \in \mathbb{R}\,.$ The corresponding functional is equal to

$$\Psi(f_\theta) \;=\; \int_0^1 f_\theta(x)p(x)\,dx \;=\; \theta\int_0^1 p(x)\,dx \;=\; \theta\,.$$

Thus, we have a parametric model of observations $y_i \;=\; \theta \,+\, \varepsilon_i$ with the efficient sample mean.

Chapter 14

Exercise 14.92 The number of monomials equals to the number of non-negative integer solutions of the equation $z_1 + \cdots + z_d = i$. Indeed, we can interpret z_j as the power of the j-th variable in the monomial, $j = 1, \dots, d$. Consider all the strings of the length $d + (i-1)$ filled with i ones and $d-1$ zeros. For example, if $d = 4$ and $i = 6$, one possible such string is 100110111. Now count the number of ones between every two consecutive zeros. In our example, they are $z_1 = 1,\ z_2 = 0, z_3 = 2$, and $z_4 = 3$. Each string corresponds to a solution of the equation $z_1 + \cdots + z_d = i$. Clearly, there are as many solutions of this equation as many strings with the described property. The latter number is the number of combinations of i objects from a set of $i+d-1$ objects.

Exercise 14.93 As defined in (14.9),

$$\hat{f}_0 = \frac{1}{n} \sum_{i,j=1}^{m} \tilde{y}_{ij} = \frac{1}{m^2} \sum_{i,j=1}^{m} \left[f_0 + f_1(i/m) + f_2(j/m) + \tilde{\varepsilon}_{ij} \right]$$

$$= f_0 + \frac{1}{m} \sum_{i=1}^{m} f_1(i/m) + \frac{1}{m} \sum_{j=1}^{m} f_2(j/m) + \frac{1}{m} \tilde{\varepsilon}$$

where

$$\tilde{\varepsilon} = \frac{1}{m} \sum_{i,j=1}^{m} \tilde{\varepsilon}_{ij} \sim \mathcal{N}(0, \sigma^2) .$$

Put

$$z_i = \frac{1}{m} \sum_{j=1}^{m} \left(y_{ij} - \hat{f}_0 \right) = \frac{1}{m} \sum_{j=1}^{m} \left[f_0 + f_1(i/m) + f_2(j/m) - \hat{f}_0 \right] + \frac{1}{m} \sum_{j=1}^{m} \varepsilon_{ij}$$

$$= f_1(i/m) + \delta_n + \frac{1}{\sqrt{m}} \bar{\varepsilon}_i - \frac{1}{m} \tilde{\varepsilon} \ \text{ with } \ \delta_n = -\frac{1}{m} \sum_{i=1}^{m} f_1(i/m) = O(1/m) .$$

The random error $\bar{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2)$ is independent of $\tilde{\varepsilon}$. The rest follows as in the proof of Proposition 14.5 with the only difference that in this case the variance of the stochastic term is bounded by $C_v N^{-1} \left(\sigma^2/m + \sigma^2/m^2 \right)$.

Exercise 14.94 Define an *anisotropic bin*, a rectangle with the sides h_1 and h_2 along the respective coordinates. Choose the sides so that $h_1^{\beta_1} = h_2^{\beta_2}$. As our estimator take the local polynomial estimator from the observations in the selected bin. The bias of this estimator has the magnitude $O(h_1^{\beta_1}) =$

$O(h_2^{\beta_2})$, while the variance is reciprocal to the number of design points in the bin, that is, $O\big((nh_1h_2)^{-1}\big)$. Under our choice of the bandwidths, we have that $h_2 = h_1^{\beta_1/\beta_2}$. The balance equation takes the form

$$h_1^{2\beta_1} = (nh_1h_2)^{-1} \ \text{ or, equivalently, } \ \big(h_1^{\beta_1}\big)^{2+1/\tilde{\beta}} = n^{-1} \ .$$

The magnitude of the bias term defines the rate of convergence which is equal to $h_1^{\beta_1} = n^{-\tilde{\beta}/(2\tilde{\beta}+1)}$.

Chapter 15

EXERCISE 15.95 Choose the bandwidths $h_{\beta_1} = (n/\ln n)^{-1/(2\beta_1+1)}$ and $h_{\beta_2} = n^{-1/(2\beta_2+1)}$. Let $\hat{f}_{\beta_1}$ and $\hat{f}_{\beta_2}$ be the local polynomial estimators of $f(x_0)$ with the chosen bandwidths.

Define $\tilde{f}_n = \hat{f}_{\beta_1}$, if the difference of the estimators $|\hat{f}_{\beta_1} - \hat{f}_{\beta_2}| \geq C\,(h_{\beta_1})^{\beta_1}$, and $\tilde{f}_n = \hat{f}_{\beta_2}$, otherwise. A sufficiently large constant C is chosen below.

As in Sections 15.2 and 15.3, we care about the risk when the adaptive estimator does not match the true smoothness parameter. If $f \in \Theta(\beta_1)$ and $\tilde{f}_n = \hat{f}_{\beta_2}$, then the difference $|\hat{f}_{\beta_1} - \hat{f}_{\beta_2}|$ does not exceed $C\,(h_{\beta_1})^{\beta_1} = C\,\psi_n(f)$, and the upper bound follows similarly to (15.11).

If $f \in \Theta(\beta_2)$, while $\tilde{f}_n = \hat{f}_{\beta_1}$, then the performance of the risk is controlled by the probabilities of large deviations $\mathbb{P}_f\big(\,|\hat{f}_{\beta_1} - \hat{f}_{\beta_2}| \geq C\,(h_{\beta_1})^{\beta_1}\,\big)$. Note that each estimator has a bias which does not exceed $C_b\,(h_{\beta_1})^{\beta_1}$. If the constant C is chosen so that $C \geq 2C_b + 2C_0$ for some large positive C_0, then the random event of interest can happen only if the stochastic term of at least one estimator exceeds $C_0(h_{\beta_1})^{\beta_1}$. The stochastic terms are zero-mean normal with the variances bounded by $C_v(h_{\beta_1})^{2\beta_1}$ and $C_v(h_{\beta_2})^{2\beta_2}$, respectively. The probabilities of the large deviations decrease faster that any power of n if C_0 is large enough.

EXERCISE 15.96 From (15.7), we have

$$\|f^*_{n,\beta_1} - f\|^2_\infty \;\leq\; 2\,A_b^2\,(h^*_{n,\beta_1})^{2\beta} \;+\; 2\,A_v^2\,\big(\,n\,h^*_{n,\beta_1}\,\big)^{-1}(\mathcal{Z}^*_{\beta_1})^2\,.$$

Hence

$$(h^*_{n,\beta_1})^{-2\beta_1}\mathbb{E}_f\big[\,\|f^*_{n,\beta_1} - f\|^2_\infty\,\big] \;\leq\; 2\,A_b^2 \;+\; 2\,A_v^2\,\mathbb{E}_f\big[\,(\mathcal{Z}^*_{\beta_1})^2\,\big]\,.$$

In view of (15.8), the latter expectation is finite.

Chapter 16

Exercise 16.97 Note that by our assumption,

$$\alpha = \mathbb{P}_0\big(\Delta_n^* = 1\big) \,\geq\, \mathbb{P}_0\big(\Delta_n = 1\big)\,.$$

It is equivalent to

$$\mathbb{P}_0\big(\Delta_n^* = 1,\, \Delta_n = 1\big) \,+\, \mathbb{P}_0\big(\Delta_n^* = 1\,,\Delta_n = 0\big)$$
$$\geq\, \mathbb{P}_0\big(\Delta_n^* = 1,\, \Delta_n = 1\big) \,+\, \mathbb{P}_0\big(\Delta_n^* = 0,\, \Delta_n = 1\big)\,,$$

which implies that

$$\mathbb{P}_0\big(\Delta_n^* = 0,\, \Delta_n = 1\big) \,\leq\, \mathbb{P}_0\big(\Delta_n^* = 1,\, \Delta_n = 0\big)\,.$$

Next, the probabilities of type II error for Δ_n^* and Δ_n are respectively equal to

$$\mathbb{P}_{\theta_1}(\Delta_n^* = 0) \,=\, \mathbb{P}_{\theta_1}(\Delta_n^* = 0,\, \Delta_n = 0) \,+\, \mathbb{P}_{\theta_1}(\Delta_n^* = 0,\, \Delta_n = 1)\,,$$

and

$$\mathbb{P}_{\theta_1}(\Delta_n = 0) \,=\, \mathbb{P}_{\theta_1}(\Delta_n^* = 0,\, \Delta_n = 0) + \mathbb{P}_{\theta_1}(\Delta_n^* = 1,\, \Delta_n = 0)\,.$$

Hence, to prove that $\mathbb{P}_{\theta_1}(\Delta_n = 0) \geq \mathbb{P}_{\theta_1}(\Delta_n^* = 0)$, it suffices to show that

$$\mathbb{P}_{\theta_1}(\Delta_n^* = 0,\, \Delta_n = 1) \,\leq\, \mathbb{P}_{\theta_1}(\Delta_n^* = 1,\, \Delta_n = 0)\,.$$

From the definition of the likelihood ratio Λ_n, and since $\Delta_n^* = \mathbb{I}(L_n \geq c)$, we obtain

$$\mathbb{P}_{\theta_1}(\Delta_n^* = 0,\, \Delta_n = 1) \,=\, \mathbb{E}_0\Big[\, e^{L_n}\, \mathbb{I}\big(\Delta_n^* = 0,\, \Delta_n = 1\big)\,\Big]$$
$$\leq\, e^c\, \mathbb{P}_0\big(\Delta_n^* = 0,\, \Delta_n = 1\big) \,\leq\, e^c\, \mathbb{P}_0\big(\Delta_n^* = 1,\, \Delta_n = 0\big)$$
$$\leq\, \mathbb{E}_0\Big[\, e^{L_n}\, \mathbb{I}\big(\Delta_n^* = 1,\, \Delta_n = 0\big)\,\Big] \,=\, \mathbb{P}_{\theta_1}(\Delta_n^* = 1,\, \Delta_n = 0)\,.$$