Chapter 2

Fundamentals of Statistics

Exercise 1 (#2.9). Consider the family of double exponential distributions: $\mathcal{P} = \left\{2^{-1}e^{-|t-\mu|} : \ -\infty < \mu < \infty\right\}$. Show that $\mathcal{P}$ is not an exponential family.

Solution. Assume that $\mathcal{P}$ is an exponential family. Then there exist p-dimensional Borel functions $T(X)$ and $\eta(\mu)$ ($p \geq 1$) and one-dimensional Borel functions $h(X)$ and $\xi(\mu)$ such that

$$2^{-1}\exp\{-|t-\mu|\} = \exp\{[\eta(\mu)]^\tau T(t) - \xi(\mu)\}h(t)$$

for any t and μ. Let $X = (X_1, ..., X_n)$ be a random sample from $P \in \mathcal{P}$ (i.e., $X_1, ..., X_n$ are independent and identically distributed with $P \in \mathcal{P}$), where $n > p$, $T_n(X) = \sum_{i=1}^n T(X_i)$, and $h_n(X) = \Pi_{i=1}^n h(X_i)$. Then the joint Lebesgue density of X is

$$2^{-n}\exp\left\{-\sum_{i=1}^n |x_i - \mu|\right\} = \exp\left\{[\eta(\mu)]^\tau T_n(x) - n\xi(\mu)\right\}h_n(x)$$

for any $x = (x_1, ..., x_n)$ and μ, which implies that

$$\sum_{i=1}^n |x_i| - \sum_{i=1}^n |x_i - \mu| = [\tilde{\eta}(\mu)]^\tau T_n(x) - n\tilde{\xi}(\mu)$$

for any x and μ, where $\tilde{\eta}(\mu) = \eta(\mu) - \eta(0)$ and $\tilde{\xi}(\mu) = \xi(\mu) - \xi(0)$. Define $\psi_\mu(x) = \sum_{i=1}^n |x_i| - \sum_{i=1}^n |x_i - \mu|$. We conclude that if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ such that $T_n(x) = T_n(y)$, then $\psi_\mu(x) = \psi_\mu(y)$ for all μ, which implies that vector of the ordered x_i's is the same as the vector of the ordered y_i's.

On the other hand, we may choose real numbers $\mu_1, ..., \mu_p$ such that $\tilde{\eta}(\mu_i)$, $i = 1, ..., p$, are linearly independent vectors. Since

$$\psi_{\mu_i}(x) = [\tilde{\eta}(\mu_i)]^\tau T_n(x) - n\tilde{\xi}(\mu_i), \quad i = 1, ..., p,$$

for any x, $T_n(x)$ is then a function of the p functions $\psi_{\mu_i}(x)$, $i = 1, ..., p$. Since $n > p$, it can be shown that there exist x and y in $\mathcal{R}^n$ such that $\psi_{\mu_i}(x) = \psi_{\mu_i}(y)$, $i = 1, ..., p$, (which implies $T_n(x) = T_n(y)$), but the vector of ordered x_i's is not the same as the vector of ordered y_i's. This contradicts the previous conclusion. Hence, $\mathcal{P}$ is not an exponential family. ∎

Exercise 2 (#2.13). A discrete random variable X with

$$P(X = x) = \gamma(x)\theta^x/c(\theta), \quad x = 0, 1, 2, ...,$$

where $\gamma(x) \geq 0$, $\theta > 0$, and $c(\theta) = \sum_{x=0}^{\infty} \gamma(x)\theta^x$, is called a random variable with a power series distribution. Show that
(i) $\{\gamma(x)\theta^x/c(\theta) : \theta > 0\}$ is an exponential family;
(ii) if $X_1, ..., X_n$ are independent and identically distributed with a power series distribution $\gamma(x)\theta^x/c(\theta)$, then $\sum_{i=1}^n X_i$ has the power series distribution $\gamma_n(x)\theta^x/[c(\theta)]^n$, where $\gamma_n(x)$ is the coefficient of θ^x in the power series expansion of $[c(\theta)]^n$.
Solution. (i) Note that

$$\gamma(x)\theta^x/c(\theta) = \exp\{x \log\theta - \log(c(\theta))\}\gamma(x).$$

Thus, $\{\gamma(x)\theta^x/c(\theta) : \theta > 0\}$ is an exponential family.
(ii) From part (i), we know that the natural parameter $\eta = \log\theta$, and also $\zeta(\eta) = \log(c(e^\eta))$. From the properties of exponential families (e.g., Theorem 2.1 in Shao, 2003), the moment generating function of X is $\psi_X(t) = e^{\zeta(\eta+t)}/e^{\zeta(\eta)} = c(\theta e^t)/c(\theta)$. The moment generating function of $\sum_{i=1}^n X_i$ is $[c(\theta e^t)]^n/[c(\theta)]^n$, which is the moment generating function of the power series distribution $\gamma_n(x)\theta^x/[c(\theta)]^n$. ∎

Exercise 3 (#2.17). Let X be a random variable having the gamma distribution with shape parameter α and scale parameter γ, where α is know and γ is unknown. let $Y = \sigma \log X$. Show that
(i) if $\sigma > 0$ is unknown, then the distribution of Y is in a location-scale family;
(ii) if $\sigma > 0$ is known, then the distribution of Y is in an exponential family.
Solution. (i) The Lebesgue density of X is

$$\frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x).$$

Applying the result in the note of Exercise 17 in Chapter 1, the Lebesgue density for $Y = \sigma \log X$ is

$$\frac{1}{\Gamma(\alpha)\sigma} e^{\alpha(y-\sigma\log\gamma)/\sigma} \exp\left\{-e^{(y-\sigma\log\gamma)/\sigma}\right\}.$$

It belongs to a location-scale family with location parameter $\eta = \sigma\log\gamma$ and scale parameter σ.

(ii) When σ is known, we rewrite the density of Y as

$$\frac{1}{\sigma\Gamma(\alpha)}\exp\{\alpha y/\sigma\}\exp\left\{-\frac{e^{y/\sigma}}{\gamma}-\alpha\log\gamma\right\}.$$

Therefore, the distribution of Y is from an exponential family. ∎

Exercise 4. Let $(X_1,...,X_n)$ be a random sample from $N(0,1)$. Show that $X_i^2/\sum_{j=1}^n X_j^2$ and $\sum_{j=1}^n X_j^2$ are independent, $i=1,...,n$.
Solution. Note that $X_1^2,...,X_n^2$ are independent and have the chi-square distribution χ_1^2. Hence their joint Lebesgue density is

$$\frac{ce^{-(y_1+\cdots+y_n)/2}}{\sqrt{y_1\cdots y_n}},\quad y_j>0,$$

where c is a constant. Let $U=\sum_{j=1}^n X_j^2$ and $V_i=X_i^2/U$, $i=1,...,n$. Then $X_i^2=UV_i$ and $\sum_{j=1}^n V_j=1$. The Lebesgue density for $(U,V_1,...,V_{n-1})$ is

$$\frac{ce^{-u/2}v_nu^{n-1}}{\sqrt{u^nv_1\cdots v_n}}=cu^{n/2-1}e^{-u/2}\sqrt{\frac{1-v_1\cdots v_{n-1}}{v_1\cdots v_{n-1}}},\quad u>0,v_j>0.$$

Hence U and $(V_1/U,...,V_{n-1}/U)$ are independent. Since $V_n=1-(V_1+\cdots+V_{n-1})$, we conclude that U and V_n/U are independent.

An alternative solution can be obtained by using Basu's theorem (e.g., Theorem 2.4 in Shao, 2003). ∎

Exercise 5. Let $X=(X_1,...,X_n)$ be a random n-vector having the multivariate normal distribution $N_n(\mu J,D)$, where J is the n-vector of 1's,

$$D=\sigma^2\begin{pmatrix}1&\rho&\cdots&\rho\\ \rho&1&\cdots&\rho\\ \cdots&\cdots&\cdots&\cdots\\ \rho&\rho&\cdots&1\end{pmatrix},$$

and $|\rho|<1$. Show that $\bar{X}=n^{-1}\sum_{i=1}^n X_i$ and $W=\sum_{i=1}^n(X_i-\bar{X})^2$ are independent, $\bar{X}$ has the normal distribution $N\left(\mu,\frac{1+(n-1)\rho}{n}\sigma^2\right)$, and $W/[(1-\rho)\sigma^2]$ has the chi-square distribution χ_{n-1}^2.
Solution. Define

$$A=\begin{pmatrix}\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\cdots&\frac{1}{\sqrt{n}}\\ \frac{1}{\sqrt{2\cdot1}}&\frac{-1}{\sqrt{2\cdot1}}&0&0&\cdots&0\\ \frac{1}{\sqrt{3\cdot2}}&\frac{1}{\sqrt{3\cdot2}}&\frac{-2}{\sqrt{3\cdot2}}&0&\cdots&0\\ \cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ \frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\cdots&\frac{-(n-1)}{\sqrt{n(n-1)}}\end{pmatrix}.$$

Then $AA^\tau = I$ (the identity matrix) and

$$ADA^\tau = \sigma^2 \begin{pmatrix} 1+(n-1)\rho & 0 & \cdots & 0 \\ 0 & 1-\rho & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1-\rho \end{pmatrix}.$$

Let $Y = AX$. Then Y is normally distributed with $E(Y) = AE(X) = (\sqrt{n}\mu, 0, ..., 0)$ and $\text{Var}(Y) = ADA^\tau$, i.e., the components of Y are independent. Let Y_i be the ith component of Y. Then, $Y_1 = \sqrt{n}\bar{X}$ and $\sum_{i=1}^n Y_i^2 = Y^\tau Y = X^\tau A^\tau AX = X^\tau X = \sum_{i=1}^n X_i^2$. Hence $\bar{X} = Y_1/\sqrt{n}$ and $W = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2$. Since Y_i's are independent, $\bar{X}$ and W are independent.

Since Y_1 has distribution $N(\sqrt{n}\mu, [1+(n-1)\rho]\sigma^2)$, $\bar{X} = Y_1/\sqrt{n}$ has distribution $N\left(\mu, \frac{1+(n-1)\rho}{n}\sigma^2\right)$. Since $Y_2, ..., Y_n$ are independent and identically distributed as $N(0, (1-\rho)\sigma^2)$, $W/[(1-\rho)\sigma^2] = \sum_{i=2}^n Y_i^2/[(1-\rho)\sigma^2]$ has the χ^2_{n-1} distribution. ∎

Exercise 6. Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $[0, 1]$ and let $R = X_{(n)} - X_{(1)}$, where $X_{(i)}$ is the ith order statistic. Derive the Lebesgue density of R and show that the limiting distribution of $2n(1-R)$ is the chi-square distribution χ^2_4.

Solution. The joint Lebesgue density of $X_{(1)}$ and $X_{(n)}$ is

$$f(x,y) = \begin{cases} n(n-1)(y-x)^{n-2} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(see, e.g., Example 2.9 in Shao, 2003). Then, the joint Lebesgue density of R and $X_{(n)}$ is

$$g(x,y) = \begin{cases} n(n-1)x^{n-2} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

and, when $0 < x < 1$, the Lebesgue density of R is

$$\int g(x,y)dy = \int_x^1 n(n-1)y^{n-2}ds = n(n-1)x^{n-2}(1-x)$$

for $0 < x < 1$. Consequently, the Lebesgue density of $2n(1-R)$ is

$$h_n(x) = \begin{cases} \frac{n-1}{4n}x\left(1 - \frac{x}{2n}\right)^{n-2} & 0 < x < 2n \\ 0 & \text{otherwise.} \end{cases}$$

Since $\lim_n \left(1 - \frac{x}{2n}\right)^{n-2} = e^{-x/2}$, $\lim_n h_n(x) = 4^{-1}xe^{x/2}I_{(0,\infty)}(x)$, which is the Lebesgue density of the χ^2_4 distribution. By Scheffé's theorem (e.g., Proposition 1.18 in Shao, 2003), the limiting distribution of $2n(1-R)$ is the χ^2_4 distribution. ∎

Exercise 7. Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution with Lebesgue density $\theta^{-1}e^{(a-x)/\theta}I_{(0,\infty)}(x)$, where $a \in \mathcal{R}$ and $\theta > 0$ are parameters. Let $X_{(1)} \le \cdots \le X_{(n)}$ be order statistics, $X_{(0)} = 0$, and $Z_i = X_{(i)} - X_{(i-1)}$, $i = 1, ..., n$. Show that
(i) $Z_1, ..., Z_n$ are independent and $2(n-i+1)Z_i/\theta$ has the χ^2_2 distribution;
(ii) $2[\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} - na]/\theta$ has the χ^2_{2r} distribution, $r = 1, ..., n$;
(iii) $X_{(1)}$ and Y are independent and $(X_{(1)} - a)/Y$ has the Lebesgue density $n\left(1 + \frac{nt}{n-1}\right)^{-n} I_{(0,\infty)}(t)$, where $Y = (n-1)^{-1}\sum_{i=1}^n (X_i - X_{(1)})$.

Solution. If we can prove the result for the case of $a = 0$ and $\theta = 1$, then the result for the general case follows by considering the transformation $(X_i - a)/\theta$, $i = 1, ..., n$. Hence, we assume that $a = 0$ and $\theta = 1$.
(i) The joint Lebesgue density of $X_{(1)}, ..., X_{(n)}$ is

$$f(x_1, ..., x_n) = \begin{cases} n!e^{-x_1 - \cdots - x_n} & 0 < x_1 < \cdots < x_n \\ 0 & \text{otherwise.} \end{cases}$$

Then the joint Lebesgue density of Z_i, $i = 1, ..., n$, is

$$g(x_1, ..., x_n) = \begin{cases} n!e^{-nx_1 - \cdots - (n-i+1)x_i - \cdots - x_n} & x_i > 0,\ i = 1, ..., n, \\ 0 & \text{otherwise.} \end{cases}$$

Hence $Z_1, ..., Z_n$ are independent and, for each i, the Lebesgue density of $2Z_i$ is $(n-i+1)e^{-(n-i+1)x_i}I_{(0,\infty)}(x_i)$. Then the density of $2(n-i+1)Z_i$ is $2^{-1}e^{-x_i/2}I_{(0,\infty)}(x_i)$, which is the density of the χ^2_2 distribution.
(ii) For $r = 1, ..., n$,

$$\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} = \sum_{i=1}^r (n-i+1)Z_i.$$

From (i), $Z_1, ..., Z_n$ are independent and $2(n-i+1)Z_i$ has the χ^2_2 distribution. Hence $2\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)}$ has the χ^2_{2r} distribution for any r.
(iii) Note that

$$Y = \frac{1}{n-1}\sum_{i=2}^n (X_{(i)} - X_{(1)}) = \frac{1}{n-1}\sum_{i=2}^n (n-i+1)Z_i.$$

From the result in (i), Y and $X_{(1)}$ are independent and $2(n-1)Y$ has the $\chi^2_{2(n-1)}$ distribution. Hence the Lebesgue density of Y is $f_Y(y) = \frac{(n-1)^n}{(n-1)!}y^{n-2}e^{-(n-1)y}I_{(0,\infty)}(y)$. Note that the Lebesgue density of $X_{(1)}$ is

$f_{X_{(1)}}(x) = ne^{-nx}I_{(0,\infty)}(x)$. Hence, for $t > 0$, the density of the ratio $X_{(1)}/Y$ is (e.g., Example 1.15 in Shao, 2003)

$$\begin{aligned} f(t) &= \int |x| f_Y(x) f_{X_{(1)}}(tx)dx \\ &= \int_0^\infty \frac{n(n-1)^n}{(n-1)!} x^{n-1} e^{-(n+nt-1)x} dx \\ &= n\left(1 + \frac{nt}{n-1}\right)^{-n} \int_0^\infty \frac{(n+nt-1)^n}{(n-1)!} x^{n-1} e^{-(n+nt-1)x} dx \\ &= n\left(1 + \frac{nt}{n-1}\right)^{-n}. \ \blacksquare \end{aligned}$$

Exercise 8 (#2.19). Let $(X_1, ..., X_n)$ be a random sample from the gamma distribution with shape parameter α and scale parameter γ_x and let $(Y_1, ..., Y_n)$ be a random sample from the gamma distribution with shape parameter α and scale parameter γ_y. Assume that X_i's and Y_i's are independent. Derive the distribution of the statistic $\bar{X}/\bar{Y}$, where $\bar{X}$ and $\bar{Y}$ are the sample means based on X_i's and Y_i's, respectively.

Solution. From the property of the gamma distribution, $n\bar{X}$ has the gamma distribution with shape parameter $n\alpha$ and scale parameter γ_x and $n\bar{Y}$ has the gamma distribution with shape parameter $n\alpha$ and scale parameter γ_y. Since $\bar{X}$ and $\bar{Y}$ are independent, the Lebesgue density of the ratio $\bar{X}/\bar{Y}$ is, for $t > 0$,

$$\begin{aligned} f(t) &= \frac{1}{[\Gamma(n\alpha)]^2(\gamma_x\gamma_y)^{n\alpha}} \int_0^\infty (tx)^{n\alpha-1} e^{-tx/\gamma_x} x^{n\alpha} e^{-x/\gamma_y} dx \\ &= \frac{\Gamma(2n\alpha)t^{n\alpha-1}}{[\Gamma(n\alpha)]^2(\gamma_x\gamma_y)^{n\alpha}} \left(\frac{t}{\gamma_x} + \frac{1}{\gamma_y}\right)^{-2n\alpha}. \ \blacksquare \end{aligned}$$

Exercise 9 (#2.22). Let (Y_i, Z_i), $i = 1, ..., n$, be independent and identically distributed random 2-vectors. The sample correlation coefficient is defined to be

$$T = \frac{1}{(n-1)S_Y S_Z} \sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z}),$$

where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$, $S_Y^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$, and $S_Z^2 = (n-1)^{-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$.

(i) Assume that $E|Y_i|^4 < \infty$ and $E|Z_i|^4 < \infty$. Show that

$$\sqrt{n}(T - \rho) \to_d N(0, c^2),$$

where ρ is the correlation coefficient between Y_1 and Z_1 and c is a constant. Identify c in terms of moments of (Y_1, Z_1).
(ii) Assume that Y_i and Z_i are independently distributed as $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Show that T has the Lebesgue density

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)}(1-t^2)^{(n-4)/2}I_{(-1,1)}(t).$$

(iii) Under the conditions of part (ii), show that the result in (i) is the same as that obtained by applying Scheffé's theorem to the density of $\sqrt{n}T$.
Solution. (i) Consider first the special case of $EY_1 = EZ_1 = 0$ and $\mathrm{Var}(Y_1) = \mathrm{Var}(Z_1) = 1$. Let $W_i = \left(Y_i, Z_i, Y_i^2, Z_i^2, Y_iZ_i\right)$ and $\bar{W} = n^{-1}\sum_{i=1}^n W_i$. Since $W_1, ..., W_n$ are independent and identically distributed and $\mathrm{Var}(W_1)$ is finite under the assumption of $E|Y_1|^4 < \infty$ and $E|Z_1|^4 < \infty$, by the central limit theorem, $\sqrt{n}(\bar{W} - \theta) \to_d N_5(0, \Sigma)$, where $\theta = (0, 0, 1, 1, \rho)$ and

$$\Sigma = \begin{pmatrix} 1 & \rho & E(Y_1^3) & E(Y_1Z_1^2) & E(Y_1^2Z_1) \\ \rho & 1 & E(Y_1^2Z_1) & E(Z_1^3) & E(Y_1Z_1^2) \\ E(Y_1^3) & E(Y_1^2Z_1) & E(Y_1^4) - 1 & E(Y_1^2Z_1^2) - 1 & E(Y_1^3Z_1) - \rho \\ E(Y_1Z_1^2) & E(Z_1^3) & E(Y_1^2Z_1^2) - 1 & E(Z_1^4) - 1 & E(Y_1Z_1^3) - \rho \\ E(Y_1^2Z_1) & E(Y_1Z_1^2) & E(Y_1^3Z_1) - \rho & E(Y_1Z_1^3) - \rho & E(Y_1^2Z_1^2) - \rho^2 \end{pmatrix}.$$

Define

$$h(x_1, x_2, x_3, x_4, x_5) = \frac{x_5 - x_1x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}}.$$

Then $T = h(\bar{W})$ and $\rho = h(\theta)$. By the δ-method (e.g., Theorem 1.12 in Shao, 2003), $\sqrt{n}[h(\bar{W}) - h(\theta)] \to_d N(0, c^2)$, where $c^2 = \xi^\tau\Sigma\xi$ and $\xi = \frac{\partial h(w)}{\partial w}\big|_{w=\theta} = (0, 0, -\rho/2, -\rho/2, 1)$. Hence

$$\begin{aligned} c^2 &= \rho^2[E(Y_1^4) + E(Z_1^4) + 2E(Y_1^2Z_1^2)]/4 \\ &\quad - \rho[E(Y_1^3Z_1) + E(Y_1Z_1^3)] + E(Y_1^2Z_1^2). \end{aligned}$$

The result for the general case can be obtained by considering the transformation $(Y_i - EY_i)/\sqrt{\mathrm{Var}(Y_i)}$ and $(Z_i - EZ_i)/\sqrt{\mathrm{Var}(Z_i)}$. The value of c^2 is then given by the previous expression with Y_1 and Z_1 replaced by $(Y_1 - EY_1)/\sqrt{\mathrm{Var}(Y_1)}$ and $(Z_1 - EZ_1)/\sqrt{\mathrm{Var}(Z_1)}$, respectively.
(ii) We only need to consider the case of $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. Let $Y = (Y_1, ..., Y_n)$, $Z = (Z_1, ..., Z_n)$, and A_Z be the n-vector whose ith component is $(Z_i - \bar{Z})/(\sqrt{n-1}S_Z)$. Note that

$$(n-1)S_Y^2 - (A_Z^\tau Y)^2 = Y^\tau B_Z Y$$

with $B_Z = I_n - n^{-1}JJ^\tau - A_ZA_Z^\tau$, where I_n is the identity matrix of order n and J is the n-vector of 1's. Since $A_Z^\tau A_Z = 1$ and $J^\tau A_Z = 0$, $B_ZA_Z = 0$,

$B_Z^2 = B_Z$ and $\text{tr}(B_Z) = n - 2$. Consequently, when Z is considered to be a fixed vector, $Y^\tau B_Z Y$ and $A_Z^\tau Y$ are independent, $A_Z^\tau Y$ is distributed as $N(0,1)$, $Y^\tau B_Z Y$ has the χ^2_{n-2} distribution, and $\sqrt{n-2}A_Z^\tau Y/\sqrt{Y^\tau B_Z Y}$ has the t-distribution t_{n-2}. Since $T = A_Z^\tau Y/(\sqrt{n-1}S_Y)$,

$$
\begin{aligned}
P(T \leq t) &= E[P(T \leq t|Z)] \\
&= E\left[P\left(\frac{A_Z^\tau Y}{\sqrt{Y^\tau B_Z Y + (A_Z^\tau Y)^2}} \leq t \middle| Z\right)\right] \\
&= E\left[P\left(\frac{A_Z^\tau Y}{\sqrt{Y^\tau B_Z Y}} \leq \frac{t}{\sqrt{1-t^2}} \middle| Z\right)\right] \\
&= E\left[P\left(t_{n-2} \leq \frac{t\sqrt{n-2}}{\sqrt{1-t^2}}\right)\right] \\
&= \frac{\Gamma(\frac{n-1}{2})}{\sqrt{(n-2)\pi}\Gamma(\frac{n-2}{2})}\int_0^{\frac{t\sqrt{n-2}}{\sqrt{1-t^2}}}\left(1+\frac{x^2}{n-2}\right)^{-(n-1)/2} dx,
\end{aligned}
$$

where t_{n-2} denotes a random variable having the t-distribution t_{n-2} and the third equality follows from the fact that $\frac{a}{\sqrt{a^2+b^2}} \leq t$ if and only if $\frac{a}{\sqrt{b^2}} \leq \frac{t}{\sqrt{1-t^2}}$ for real numbers a and b and $t \in (0,1)$. Thus, T has Lebesgue density

$$
\frac{d}{dt}P(T \leq t) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n-2}{2})}(1-t^2)^{(n-4)/2} I_{(-1,1)}(t).
$$

(iii) Under the conditions of part (ii), $\rho = 0$ and, from the result in (i), $c = 1$ and $\sqrt{n}T \to_d N(0,1)$. From the result in (ii), $\sqrt{n}T$ has Lebesgue density

$$
\frac{\Gamma(\frac{n-1}{2})}{\sqrt{n\pi}\Gamma(\frac{n-2}{2})}\left(1-\frac{t^2}{n}\right)^{(n-4)/2} I_{(-\sqrt{n},\sqrt{n})}(t) \to \frac{1}{\sqrt{2\pi}}e^{-t^2/2}
$$

by Stirling's formula. By Scheffé's Theorem, $\sqrt{n}T \to_d N(0,1)$. ∎

Exercise 10 (#2.23). Let $X_1, ..., X_n$ be independent and identically distributed random variables with $EX_1^4 < \infty$, $T = (Y, Z)$, and $T_1 = Y/\sqrt{Z}$, where $Y = n^{-1}\sum_{i=1}^n |X_i|$ and $Z = n^{-1}\sum_{i=1}^n X_i^2$.
(i) Show that $\sqrt{n}(T-\theta) \to_d N_2(0,\Sigma)$ and $\sqrt{n}(T_1-\vartheta) \to_d N(0,c^2)$. Identify θ, Σ, ϑ, and c^2 in terms of moments of X_1.
(ii) Repeat (i) when X_1 has the normal distribution $N(0,\sigma^2)$.
(iii) Repeat (i) when X_1 has Lebesgue density $(2\sigma)^{-1}e^{-|x|/\sigma}$.
Solution. (i) Define $\theta_j = E|X_1|^j$, $j = 1,2,3,4$, and $W_i = (|X_i|, X_i^2)$, $i = 1, ..., n$. Then $T = n^{-1}\sum_{i=1}^n W_i$. Let $\theta = EW_1 = (\theta_1, \theta_2)$. By the central limit theorem, $\sqrt{n}(T-\theta) \to_d N_2(0,\Sigma)$, where

$$
\Sigma = \begin{pmatrix} \text{Var}(|X_1|) & \text{Cov}(|X_1|, X_1^2) \\ \text{Cov}(|X_1|, X_1^2) & \text{Var}(X_1^2) \end{pmatrix} = \begin{pmatrix} \theta_2 - \theta_1^2 & \theta_3 - \theta_1\theta_2 \\ \theta_3 - \theta_1\theta_2 & \theta_4 - \theta_2^2 \end{pmatrix}.
$$

Let $g(y,z) = y/\sqrt{z}$. Then $T_1 = g(T)$, $g(\theta) = \theta_1/\sqrt{\theta_2}$, $\frac{\partial g}{\partial y}\big|_{(y,z)=\theta} = 1/\sqrt{\theta_2}$, and $\frac{\partial g}{\partial z}\big|_{(y,z)=\theta} = -\theta_1/(2\theta_2^{3/2})$. Then, by the δ-method, $\sqrt{n}(T_1 - \vartheta) \to_d N(0, c^2)$ with $\vartheta = \theta_1/\sqrt{\theta_2}$ and

$$c^2 = 1 + \frac{\theta_1^2\theta_4}{4\theta_2^3} - \frac{\theta_1\theta_3}{\theta_2^2} - \frac{\theta_1^2}{4\theta_3}.$$

(ii) We only need to calculate θ_j. When X_1 is distributed as $N(0,\sigma^2)$, a direct calculation shows that $\theta_1 = \sqrt{2}\sigma/\sqrt{\pi}$, $\theta_2 = \sigma^2$, $\theta_3 = 2\sqrt{2}\sigma^3/\sqrt{\pi}$, and $\theta_4 = 3\sigma^4$.
(iii) Note that $|X_1|$ has the exponential distribution with Lebesgue density $\sigma^{-1}e^{-x/\sigma}I_{(0,\infty)}(x)$. Hence, $\theta_j = \sigma^j j!$. ∎

Exercise 11 (#2.25). Let X be a sample from $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of distributions on the Borel σ-field on $\mathcal{R}^n$. Show that if $T(X)$ is a sufficient statistic for $P \in \mathcal{P}$ and $T = \psi(S)$, where ψ is measurable and $S(X)$ is another statistic, then $S(X)$ is sufficient for $P \in \mathcal{P}$.
Solution. Assume first that all P in $\mathcal{P}$ are dominated by a σ-finite measure ν. Then, by the factorization theorem (e.g., Theorem 2.2 in Shao, 2003),

$$\frac{dP}{d\nu}(x) = g_P(T(x))h(x),$$

where h is a Borel function of x (not depending on P) and $g_P(t)$ is a Borel function of t. If $T = \psi(S)$, then

$$\frac{dP}{d\nu}(x) = g_P(\psi(S(x)))h(x)$$

and, by the factorization theorem again, $S(X)$ is sufficient for $P \in \mathcal{P}$.

Consider the general case. Suppose that $S(X)$ is not sufficient for $P \in \mathcal{P}$. By definition, there exist at least two measures $P_1 \in \mathcal{P}$ and $P_2 \in \mathcal{P}$ such that the conditional distributions of X given $S(X)$ under P_1 and P_2 are different. Let $\mathcal{P}_0 = \{P_1, P_2\}$, which is a sub-family of $\mathcal{P}$. Since $T(X)$ is sufficient for $P \in \mathcal{P}$, it is also sufficient for $P \in \mathcal{P}_0$. Since all P in $\mathcal{P}_0$ are dominated by the measure $P_1 + P_2$, by the previously proved result, $S(X)$ is sufficient for $P \in \mathcal{P}_0$. Hence, the conditional distributions of X given $S(X)$ under P_1 and P_2 are the same. This contradiction proves that $S(X)$ is sufficient for $P \in \mathcal{P}$. ∎

Exercise 12. Let $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$, where f_θ's are probability densities, $f_\theta(x) > 0$ for all $x \in \mathcal{R}$ and, for any $\theta \in \Theta$, $f_\theta(x)$ is continuous in x. Let X_1 and X_2 be independent and identically distributed as f_θ. Show that if $X_1 + X_2$ is sufficient for θ, then $\mathcal{P}$ is an exponential family indexed by θ.
Solution. The joint density of X_1 and X_2 is $f_\theta(x_1)f_\theta(x_2)$. By the factorization theorem, there exist functions $g_\theta(t)$ and $h(x_1, x_2)$ such that

$$f_\theta(x_1)f_\theta(x_2) = g_\theta(x_1 + x_2)h(x_1, x_2).$$

Then

$$\log f_\theta(x_1) + \log f_\theta(x_2) = g(x_1 + x_2, \theta) + h_1(x_1, x_2),$$

where $g(t,\theta) = \log g_\theta(t)$ and $h_1(x_1, x_2) = \log h(x_1, x_2)$. Let $\theta_0 \in \Theta$ and $r(x,\theta) = \log f_\theta(x) - \log f_{\theta_0}(x)$ and $q(x,\theta) = g(x,\theta) - g(x,\theta_0)$. Then

$$\begin{aligned} q(x_1 + x_2, \theta) &= \log f_\theta(x_1) + \log f_\theta(x_2) + h_1(x_1, x_2) \\ &\quad - \log f_{\theta_0}(x_1) - \log f_{\theta_0}(x_2) - h_1(x_1, x_2) \\ &= r(x_1, \theta) + r(x_2, \theta). \end{aligned}$$

Consequently,

$$r(x_1 + x_2, \theta) + r(0, \theta) = q(x_1 + x_2, \theta) = r(x_1, \theta) + r(x_2, \theta)$$

for any x_1, x_2, and θ. Let $s(x,\theta) = r(x,\theta) - r(0,\theta)$. Then

$$s(x_1, \theta) + s(x_2, \theta) = s(x_1 + x_2, \theta)$$

for any x_1, x_2, and θ. Hence,

$$s(n, \theta) = ns(1, \theta) \quad n = 0, \pm 1, \pm 2, ...$$

For any rational number $\frac{n}{m}$ (n and m are integers and $m \neq 0$),

$$s\left(\tfrac{n}{m}, \theta\right) = ns\left(\tfrac{1}{m}, \theta\right) = \tfrac{m}{m} ns\left(\tfrac{1}{m}, \theta\right) = \tfrac{n}{m} s\left(\tfrac{m}{m}, \theta\right) = \tfrac{n}{m} s\left(1, \theta\right).$$

Hence $s(x,\theta) = xs(1,\theta)$ for any rational x. From the continuity of f_θ, we conclude that $s(x,\theta) = xs(1,\theta)$ for any $x \in \mathcal{R}$, i.e.,

$$r(x, \theta) = s(1, \theta)x + r(0, \theta)$$

any $x \in \mathcal{R}$. Then, for any x and θ,

$$\begin{aligned} f_\theta(x) &= \exp\{r(x,\theta) + \log f_{\theta_0}(x)\} \\ &= \exp\{s(1,\theta)x + r(0,\theta) + \log f_{\theta_0}(x)\} \\ &= \exp\{\eta(\theta)x - \xi(\theta)\}h(x), \end{aligned}$$

where $\eta(\theta) = s(1,\theta)$, $\xi(\theta) = -r(0,\theta)$, and $h(x) = f_{\theta_0}(x)$. This shows that $\mathcal{P}$ is an exponential family indexed by θ. ∎

Exercise 13 (#2.30). Let X and Y be two random variables such that Y has the binomial distribution with size N and probability π and, given $Y = y$, X has the binomial distribution with size y and probability p.
(i) Suppose that $p \in (0,1)$ and $\pi \in (0,1)$ are unknown and N is known. Show that (X, Y) is minimal sufficient for (p, π).
(ii) Suppose that π and N are known and $p \in (0,1)$ is unknown. Show

whether X is sufficient for p and whether Y is sufficient for p.

Solution. (i) Let $A = \{(x,y) : x = 0,1,...,y,\ y = 0,1,...,N\}$. The joint probability density of (X,Y) with respect to the counting measure is

$$\binom{N}{y}\pi^y(1-\pi)^{N-y}\binom{y}{x}p^x(1-p)^{y-x}I_A$$
$$= \exp\left\{x\log\frac{p}{1-p} + y\log\frac{\pi(1-p)}{1-\pi} + N\log(1-\pi)\right\}\binom{N}{y}\binom{y}{x}I_A.$$

Hence, (X,Y) has a distribution from an exponential family of full rank ($0 < p < 1$ and $0 < \pi < 1$). This implies that (X,Y) is minimal sufficient for (p,π).

(ii) The joint probability density of (X,Y) can be written as

$$\exp\left\{x\log\frac{p}{1-p} + y\log(1-p)\right\}\pi^y(1-\pi)^{N-y}\binom{N}{y}\binom{y}{x}I_A.$$

This is from an exponential family not of full rank. Let $p_0 = \frac{1}{2}$, $p_1 = \frac{1}{3}$, $p_2 = \frac{2}{3}$, and $\eta(p) = (\log\frac{p}{1-p}, \log(1-p))$. Then, two vectors in $\mathcal{R}^2$, $\eta(p_1) - \eta(p_0) = (-\log 2, 2\log 2 - \log 3)$ and $\eta(p_2) - \eta(p_0) = (\log 2, \log 2 - \log 3)$, are linearly independent. By the properties of exponential families (e.g., Example 2.14 in Shao, 2003), (X,Y) is minimal sufficient for p. Thus, neither X nor Y is sufficient for p. ∎

Exercise 14 (#2.34). Let $X_1,...,X_n$ be independent and identically distributed random variables having the Lebesgue density

$$\exp\left\{-\left(\tfrac{x-\mu}{\sigma}\right)^4 - \xi(\theta)\right\},$$

where $\theta = (\mu,\sigma) \in \Theta = \mathcal{R}\times(0,\infty)$. Show that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is an exponential family, where P_θ is the joint distribution of $X_1,...,X_n$, and that the statistic $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i^3, \sum_{i=1}^n X_i^4\right)$ is minimal sufficient for $\theta \in \Theta$.

Solution. Let $T(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i^3, \sum_{i=3}^n x_i^4)$ for any $x = (x_1,...,x_n)$ and let $\eta(\theta) = \sigma^{-4}(-4\mu^3, 6\mu^2, -4\mu, 1)$. The joint density of $(X_1,...,X_n)$ is

$$f_\theta(x) = \exp\left\{[\eta(\theta)]^\tau T(x) - n\mu^4/\sigma^4 - n\xi(\theta)\right\},$$

which belongs to an exponential family. For any two sample points $x = (x_1,...,x_n)$ and $y = (y_1,...,y_n)$,

$$\frac{f_\theta(x)}{f_\theta(y)} = \exp\left\{-\frac{1}{\sigma^4}\left[\left(\sum_{i=1}^n x_i^4 - \sum_{i=1}^n y_i^4\right) - 4\mu\left(\sum_{i=1}^n x_i^3 - \sum_{i=1}^n y_i^3\right)\right.\right.$$
$$\left.\left. +\ 6\mu^2\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) - 4\mu^3\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)\right]\right\},$$

which is free of parameter (μ, σ) if and only if $T(x) = T(y)$. By Theorem 2.3(iii) in Shao (2003), $T(X)$ is minimal sufficient for θ. ∎

Exercise 15 (#2.35). Let $(X_1, ..., X_n)$ be a random sample of random variables having the Lebesgue density $f_\theta(x) = (2\theta)^{-1}\left[I_{(0,\theta)}(x) + I_{(2\theta,3\theta)}(x)\right]$. Find a minimal sufficient statistic for $\theta \in (0, \infty)$.
Solution. We use the idea of Theorem 2.3(i)-(ii) in Shao (2003). Let $\Theta_r = \{\theta_1, \theta_2, ...\}$ be the set of positive rational numbers, $\mathcal{P}_0 = \{g_\theta : \theta \in \Theta_r\}$, and $\mathcal{P} = \{g_\theta : \theta > 0\}$, where $g_\theta(x) = \prod_{i=1}^n f_\theta(x_i)$ for $x = (x_1, ..., x_n)$. Then $\mathcal{P}_0 \subset \mathcal{P}$ and a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$ (i.e., if an event A satisfying $P(A) = 0$ for all $P \in \mathcal{P}_0$, then $P(A) = 0$ for all $P \in \mathcal{P}$). Let $\{c_i\}$ be a sequence of positive numbers satisfying $\sum_{i=1}^\infty c_i = 1$ and $g_\infty(x) = \sum_{i=1}^\infty c_i g_{\theta_i}(x)$. Define $T = (T_1, T_2, ...)$ with $T_i(x) = g_{\theta_i}(x)/g_\infty(x)$. By Theorem 2.3(ii) in Shao (2003), T is minimal sufficient for $\theta \in \Theta_0$ (or $P \in \mathcal{P}_0$). For any $\theta > 0$, there is a sequence $\{\theta_{i_k}\} \subset \{\theta_i\}$ such that $\lim_k \theta_{i_k} = \theta$. Then

$$g_\theta(x) = \lim_k g_{\theta_{i_k}}(x) = \lim_k T_{i_k}(x) g_\infty(x)$$

holds for all $x \in C$ with $P(C) = 1$ for all $P \in \mathcal{P}$. By the factorization theorem, T is sufficient for $\theta > 0$ (or $P \in \mathcal{P}$). By Theorem 2.3(i) in Shao (2003), T is minimal sufficient for $\theta > 0$. ∎

Exercise 16 (#2.36). Let $(X_1, ..., X_n)$ be a random sample of random variables having the Cauchy distribution with location parameter μ and scale parameter σ, where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown parameters. Show that the vector of order statistics is minimal sufficient for (μ, σ).
Solution. The joint Lebesgue density of $(X_1, ..., X_n)$ is

$$f_{\mu,\sigma}(x) = \frac{\sigma^n}{\pi^n} \prod_{i=1}^n \frac{1}{\sigma^2 + (x_i - \mu)^2}, \quad x = (x_1, ..., x_n).$$

For any $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, suppose that

$$\frac{f_{\mu,\sigma}(x)}{f_{\mu,\sigma}(y)} = \psi(x, y)$$

holds for any μ and σ, where ψ does not depend on (μ, σ). Let $\sigma = 1$. Then we must have

$$\prod_{i=1}^n \left[1 + (y_i - \mu)^2\right] = \psi(x, y) \prod_{i=1}^n \left[1 + (x_i - \mu)^2\right]$$

for all μ. Both sides of the above identity can be viewed as polynomials of degree $2n$ in μ. Comparison of the coefficients to the highest terms gives

$\psi(x, y) = 1$. Thus,

$$\prod_{i=1}^{n} \left[1 + (y_i - \mu)^2\right] = \prod_{i=1}^{n} \left[1 + (x_i - \mu)^2\right]$$

for all μ. As a polynomial of μ, the left-hand side of the above identity has $2n$ complex roots $x_i \pm \sqrt{-1}$, $i = 1, ..., n$, while the right-hand side of the above identity has $2n$ complex roots $y_i \pm \sqrt{-1}$, $i = 1, ..., n$. By the unique factorization theorem for the entire functions in complex analysis, we conclude that the two sets of roots must agree. This means that the ordered values of x_i's are the same as the ordered values of y_i's. By Theorem 2.3(iii) in Shao (2003), the order statistics of $X_1, ..., X_n$ is minimal sufficient for (μ, σ). ∎

Exercise 17 (#2.40). Let $(X_1, ..., X_n)$, $n \geq 2$, be a random sample from a distribution having Lebesgue density $f_{\theta,j}$, where $\theta > 0$, $j = 1, 2$, $f_{\theta,1}$ is the density of $N(0, \theta^2)$, and $f_{\theta,2}(x) = (2\theta)^{-1} e^{-|x|/\theta}$. Show that $T = (T_1, T_2)$ is minimal sufficient for (θ, j), where $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n |X_i|$.
Solution A. Let P be the joint distribution of $X_1, ..., X_n$. By the factorization theorem, T is sufficient for (θ, j). Let $\mathcal{P} = \{P : \theta > 0, j = 1, 2\}$, $\mathcal{P}_1 = \{P : \theta > 0, j = 1\}$, and $\mathcal{P}_2 = \{P : \theta > 0, j = 2\}$. Let S be a statistic sufficient for $P \in \mathcal{P}$. Then S is sufficient for $P \in \mathcal{P}_j$, $j = 1, 2$. Note that $\mathcal{P}_1$ is an exponential family with T_1 as a minimal sufficient statistic. Hence, there exists a Borel function ψ_1 such that $T_1 = \psi_1(S)$ a.s. $\mathcal{P}_1$. Since all densities in $\mathcal{P}$ are dominated by those in $\mathcal{P}_1$, we conclude that $T_1 = \psi_1(S)$ a.s. $\mathcal{P}$. Similarly, $\mathcal{P}_2$ is an exponential family with T_2 as a minimal sufficient statistic and, thus, there exists a Borel function ψ_2 such that $T_2 = \psi_2(S)$ a.s. $\mathcal{P}$. This proves that $T = (\psi_1(S), \psi_2(S))$ a.s. $\mathcal{P}$. Hence T is minimal sufficient for (θ, j).
Solution B. Let P be the joint distribution of $X_1, ..., X_n$. The Lebesgue density of P can be written as

$$\exp\left\{-\frac{I_{\{1\}}(j)}{2\theta^2} T_1 - \frac{I_{\{2\}}(j)}{\theta} T_2\right\} \left[\frac{I_{\{1\}}(j)}{(2\pi\theta^2)^{n/2}} + \frac{I_{\{2\}}(j)}{(2\theta)^n}\right].$$

Hence $\mathcal{P} = \{P : \theta > 0, j = 1, 2\}$ is an exponential family. Let

$$\eta(\theta, j) = -\left(\frac{I_{\{1\}}(j)}{2\theta^2}, \frac{I_{\{2\}}(j)}{\theta}\right).$$

Note that $\eta(1, 1) = (-\frac{1}{2}, 0)$, $\eta(2^{-1/2}, 1) = (-1, 0)$, and $\eta(1, 2) = (0, -1)$. Then, $\eta(2^{-1/2}, 1) - \eta(1, 1) = (-\frac{1}{2}, 0)$ and $\eta(1, 2) - \eta(1, 1) = (\frac{1}{2}, -1)$ are two linearly independent vectors in $\mathcal{R}^2$. Hence $T = (T_1, T_2)$ is minimal sufficient for (θ, j) (e.g., Example 2.14 in Shao, 2003). ∎

Exercise 18 (#2.41). Let $(X_1,...,X_n)$, $n\geq 2$, be a random sample from a distribution with discrete probability density $f_{\theta,j}$, where $\theta\in(0,1)$, $j=1,2$, $f_{\theta,1}$ is the Poisson distribution with mean θ, and $f_{\theta,2}$ is the binomial distribution with size 1 and probability θ.
(i) Show that $T=\sum_{i=1}^n X_i$ is not sufficient for (θ,j).
(ii) Find a two-dimensional minimal sufficient statistic for (θ,j).
Solution. (i) To show that T is not sufficient for (θ,j), it suffices to show that, for some $x\leq t$, $P(X_n=x|T=t)$ for $j=1$ is different from $P(X_n=x|T=t)$ for $j=2$. When $j=1$,

$$P(X_n=x|T=t)=\binom{t}{x}\frac{(n-1)^{t-x}}{n^t}>0,$$

whereas when $j=2$, $P(X_n=x|T=t)=0$ as long as $x>1$.
(ii) Let $g_{\theta,j}$ be the joint probability density of $X_1,...,X_n$. Let $\mathcal{P}_0=\{g_{\frac{1}{4},1},g_{\frac{1}{2},1},g_{\frac{1}{2},2}\}$. Then, a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$. By Theorem 2.3(ii) in Shao (2003), the two-dimensional statistic

$$S=\left(\frac{g_{\frac{1}{2},1}}{g_{\frac{1}{4},1}},\frac{g_{\frac{1}{2},2}}{g_{\frac{1}{4},1}}\right)=\left(e^{n/4}2^{-T},e^{n/2}W2^{T-n}\right)$$

is minimal sufficient for the family $\mathcal{P}_0$, where

$$W=\begin{cases}1 & X_i=0 \text{ or } 1, i=1,...,n,\\ 0 & \text{otherwise.}\end{cases}$$

Since there is a one-to-one transformation between S and (T,W), we conclude that (T,W) is minimal sufficient for the family $\mathcal{P}_0$. For any $x=(x_1,...,x_n)$, the joint density of $X_1,...,X_n$ is

$$e^{n\theta I_{\{1\}}(j)}(1-\theta)^{nI_{\{2\}}(j)}W^{I_{\{2\}}(j)}e^{T[I_{\{1\}}(j)\log\theta+I_{\{2\}}(j)\log\frac{\theta}{(1-\theta)}]}\prod_{i=1}^n\frac{1}{x_i!}.$$

Hence, by the factorization theorem, (T,W) is sufficient for (θ,j). By Theorem 2.3(i) in Shao (2003), (T,W) is minimal sufficient for (θ,j). ∎

Exercise 19 (#2.44). Let $(X_1,...,X_n)$ be a random sample from a distribution on $\mathcal{R}$ having the Lebesgue density $\theta^{-1}e^{-(x-\theta)/\theta}I_{(\theta,\infty)}(x)$, where $\theta>0$ is an unknown parameter.
(i) Find a statistic that is minimal sufficient for θ.
(ii) Show whether the minimal sufficient statistic in (i) is complete.
Solution. (i) Let $T(x)=\sum_{i=1}^n x_i$ and $W(x)=\min_{1\leq i\leq n}x_i$, where $x=(x_1,...,x_n)$. The joint density of $X=(X_1,...,X_n)$ is

$$f_\theta(x)=\frac{e^n}{\theta^n}e^{-T(x)/\theta}I_{(\theta,\infty)}(W(x)).$$

For $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$,

$$\frac{f_\theta(x)}{f_\theta(y)} = e^{[T(y)-T(x)]/\theta} \frac{I_{(\theta,\infty)}(W(x))}{I_{(\theta,\infty)}(W(y))}$$

is free of θ if and only if $T(x) = T(y)$ and $W(x) = W(y)$. Hence, the two-dimensional statistic $(T(X), W(X))$ is minimal sufficient for θ.
(ii) A direct calculation shows that, for any θ, $E[T(X)] = 2n\theta$ and $E[W(X)] = (1 + n^{-1})\theta$. Hence $E[(2n)^{-1}T - (1 + n^{-1})^{-1}W(X)] = 0$ for any θ and $(2n)^{-1}T - (1 + n^{-1})^{-1}W(X)$ is not a constant. Thus, (T, W) is not complete. ∎

Exercise 20 (#2.48). Let T be a complete (or boundedly complete) and sufficient statistic. Suppose that there is a minimal sufficient statistic S. Show that T is minimal sufficient and S is complete (or boundedly complete).
Solution. We prove the case when T is complete. The case in which T is boundedly complete is similar. Since S is minimal sufficient and T is sufficient, there exists a Borel function h such that $S = h(T)$ a.s. Since h cannot be a constant function and T is complete, we conclude that S is complete. Consider $T - E(T|S) = T - E[T|h(T)]$, which is a Borel function of T and hence can be denoted as $g(T)$. Note that $E[g(T)] = 0$. By the completeness of T, $g(T) = 0$ a.s., that is, $T = E(T|S)$ a.s. This means that T is also a function of S and, therefore, T is minimal sufficient. ∎

Exercise 21 (#2.53). Let X be a discrete random variable with probability density

$$f_\theta(x) = \begin{cases} \theta & x = 0 \\ (1-\theta)^2\theta^{x-1} & x = 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta \in (0, 1)$. Show that X is boundedly complete, but not complete.
Solution. Consider any Borel function $h(x)$ such that

$$E[h(X)] = h(0)\theta + \sum_{x=1}^{\infty} h(x)(1-\theta)^2\theta^{x-1} = 0$$

for any $\theta \in (0, 1)$. Rewriting the left-hand side of the above equation in the ascending order of the powers of θ, we obtain that

$$h(1) + \sum_{x=1}^{\infty} [h(x-1) - 2h(x) + h(x+1)]\,\theta^x = 0$$

for any $\theta \in (0, 1)$. Comparing the coefficients of both sides, we obtain that $h(1) = 0$ and $h(x-1) - h(x) = h(x) - h(x+1)$. Therefore, $h(x) = (1-x)h(0)$

for $x = 1, 2, ...$ This function is bounded if and only if $h(0) = 0$. If $h(x)$ is assumed to be bounded, then $h(0) = 0$ and, hence, $h(x) \equiv 0$. This means that X is boundedly complete. For $h(x) = 1 - x$, $E[h(X)] = 0$ for any θ but $h(X) \neq 0$. Therefore, X is not complete. ∎

Exercise 22. Let X be a discrete random variable with

$$P_\theta(X = x) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, ..., \min\{\theta, n\}, n - x \leq N - \theta,$$

where n and N are positive integers, $N \geq n$, and $\theta = 0, 1, ..., N$. Show that X is complete.
Solution. Let $g(x)$ be a function of $x \in \{0, 1, ..., n\}$. Assume $E_\theta[g(X)] = 0$ for any θ, where E_θ is the expectation with respect to P_θ. When $\theta = 0$, $P_0(X = x) = 1$ if $x = 0$ and $E_0[g(X)] = g(0)$. Thus, $g(0) = 0$. When $\theta = 1$, $P_1(X \geq 2) = 0$ and

$$E_1[g(X)] = g(0)P_1(X = 0) + g(1)P_1(X = 1) = g(1)\frac{\binom{N-1}{n-1}}{\binom{N}{n}}.$$

Since $E_1[g(X)] = 0$, we obtain that $g(1) = 0$. Similarly, we can show that $g(2) = \cdots = g(n) = 0$. Hence X is complete. ∎

Exercise 23. Let X be a random variable having the uniform distribution on the interval $(\theta, \theta + 1)$, $\theta \in \mathcal{R}$. Show that X is not complete.
Solution. Consider $g(X) = \cos(2\pi X)$. Then $g(X) \neq 0$ but

$$E[g(X)] = \int_\theta^{\theta+1} \cos(2\pi x)dx = \frac{\sin(2\pi(\theta + 1)) - \sin(2\pi\theta)}{2\pi} = 0$$

for any θ. Hence X is not complete. ∎

Exercise 24 (#2.57). Let $(X_1, ..., X_n)$ be a random sample from the $N(\theta, \theta^2)$ distribution, where $\theta > 0$ is a parameter. Find a minimal sufficient statistic for θ and show whether it is complete.
Solution. The joint Lebesgue density of $X_1, ..., X_n$ is

$$\frac{1}{(2\pi\theta^2)^n} \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 + \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{2}\right\}.$$

Let

$$\eta(\theta) = \left(-\frac{1}{2\theta^2}, \frac{1}{\theta}\right).$$

Then $\eta(\frac{1}{2}) - \eta(1) = (-\frac{3}{2}, 1)$ and $\eta(\frac{1}{\sqrt{2}}) - \eta(1) = (-\frac{1}{2}, \sqrt{2})$ are linearly independent vectors in $\mathcal{R}^2$. Hence $T = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ is minimal sufficient for θ. Note that

$$E\left(\sum_{i=1}^{n} X_i^2\right) = nEX_1^2 = 2n\theta^2$$

and

$$E\left(\sum_{i=1}^{n} X_i\right)^2 = n\theta^2 + (n\theta)^2 = (n+n^2)\theta^2.$$

Let $h(t_1,t_2) = \frac{1}{2n}t_1 - \frac{1}{n(n+1)}t_2^2$. Then $h(t_1,t_2) \neq 0$ but $E[h(T)] = 0$ for any θ. Hence T is not complete. ∎

Exercise 25 (#2.56). Suppose that $(X_1,Y_1),...,(X_n,Y_n)$ are independent and identically distributed random 2-vectors and X_i and Y_i are independently distributed as $N(\mu,\sigma_X^2)$ and $N(\mu,\sigma_Y^2)$, respectively, with $\theta = (\mu,\sigma_X^2,\sigma_Y^2) \in \mathcal{R} \times (0,\infty) \times (0,\infty)$. Let $\bar{X}$ and S_X^2 be the sample mean and variance for X_i's and $\bar{Y}$ and S_Y^2 be the sample mean and variance for Y_i's. Show that $T = (\bar{X},\bar{Y},S_X^2,S_Y^2)$ is minimal sufficient for θ but T is not boundedly complete.
Solution. Let

$$\eta = \left(-\frac{1}{2\sigma_X^2}, \frac{\mu}{\sigma_X^2}, -\frac{1}{2\sigma_Y^2}, \frac{\mu}{\sigma_Y^2}\right)$$

and

$$S = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} Y_i^2, \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i\right).$$

Then the joint Lebesgue density of $(X_1,Y_1),...,(X_n,Y_n)$ is

$$\frac{1}{(2\pi)^n}\exp\left\{\eta^\tau S - \frac{n\mu^2}{2\sigma_X^2} - \frac{n\mu^2}{2\sigma_Y^2} - n\log(\sigma_X\sigma_Y)\right\}.$$

Since the parameter space $\{\eta : \mu \in \mathcal{R}, \sigma_X^2 > 0, \sigma_Y^2 > 0\}$ is a three-dimensional curved hyper-surface in $\mathcal{R}^4$, we conclude that S is minimal sufficient. Note that there is a one-to-one correspondence between T and S. Hence T is also minimal sufficient.

To show that T is not boundedly complete, consider $h(T) = I_{\{\bar{X}>\bar{Y}\}} - \frac{1}{2}$. Then $|h(T)| \leq 0.5$ and $E[h(T)] = 0$ for any η, but $h(T) \neq 0$. Hence T is not boundedly complete. ∎

Exercise 26 (#2.58). Suppose that $(X_1,Y_1),...,(X_n,Y_n)$ are independent and identically distributed random 2-vectors having the normal distribution with $EX_1 = EY_1 = 0$, $\mathrm{Var}(X_1) = \mathrm{Var}(Y_1) = 1$, and $\mathrm{Cov}(X_1,Y_1) = \theta \in (-1,1)$.
(i) Find a minimal sufficient statistic for θ.
(ii) Show whether the minimal sufficient statistic in (i) is complete or not.

(iii) Prove that $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n Y_i^2$ are both ancillary but (T_1, T_2) is not ancillary.
Solution. (i) The joint Lebesgue density of $(X_1, Y_1), ..., (X_n, Y_n)$ is

$$\left(\frac{1}{2\pi\sqrt{1-\theta^2}}\right)^n \exp\left\{-\frac{1}{1-\theta^2}\sum_{i=1}^n (x_i^2 + y_i^2) + \frac{2\theta}{1-\theta^2}\sum_{i=1}^n x_i y_i\right\}.$$

Let

$$\eta = \left(-\frac{1}{1-\theta^2}, \frac{2\theta}{1-\theta^2}\right).$$

The parameter space $\{\eta : -1 < \theta < 1\}$ is a curve in $\mathcal{R}^2$. Therefore, $(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i)$ is minimal sufficient.
(ii) Note that $E[\sum_{i=1}^n (X_i^2 + Y_i^2)] - 2n = 0$, but $\sum_{i=1}^n (X_i^2 + Y_i^2) - 2n \neq 0$. Therefore, the minimal sufficient statistic is not complete.
(iii) Both T_1 and T_2 have the chi-square distribution χ_n^2, which does not depend on θ. Hence both T_1 and T_2 are ancillary. Note that

$$\begin{aligned}
E(T_1 T_2) &= E\left(\sum_{i=1}^n X_i^2\right)\left(\sum_{j=1}^n Y_j^2\right) \\
&= E\left(\sum_{i=1}^n X_i^2 Y_i^2\right) + E\left(\sum_{i\neq j} X_i^2 Y_j^2\right) \\
&= nE(X_1^2 Y_1^2) + n(n-1)E(X_1^2)E(Y_1^2) \\
&= n(1+2\theta^2) + 2n(n-1),
\end{aligned}$$

which depends on θ. Therefore the distribution of (T_1, T_2) depends on θ and (T_1, T_2) is not ancillary. ∎

Exercise 27 (#2.59). Let $(X_1, ..., X_n)$, $n > 2$, be a random sample from the exponential distribution on (a, ∞) with scale parameter θ. Show that
(i) $\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ are independent for any (a, θ), where $X_{(j)}$ is the jth order statistic;
(ii) $Z_i = (X_{(n)} - X_{(i)})/(X_{(n)} - X_{(n-1)})$, $i = 1, ..., n-2$, are independent of $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$.
Solution: (i) Let θ be arbitrarily fixed. Since the joint density of $X_1, ..., X_n$ is

$$\theta^{-n} e^{na/\theta} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^n x_i\right\} I_{(a,\infty)}(x_{(1)}),$$

where only a is considered as an unknown parameter, we conclude that $X_{(1)}$ is sufficient for a. Note that $\frac{n}{\theta} e^{-n(x-a)/\theta} I_{(a,\infty)}(x)$ is the Lebesgue density for $X_{(1)}$. For any Borel function g,

$$E[g(X_{(1)})] = \frac{n}{\theta}\int_a^\infty g(x)e^{-n(x-a)/\theta}dx = 0$$

for any a is equivalent to

$$\int_a^\infty g(x)e^{-nx/\theta}dx = 0$$

for any a, which implies $g(x) = 0$ a.e. with respect to Lebesgue measure. Hence, for any fixed θ, $X_{(1)}$ is sufficient and complete for a. The Lebesgue density of $X_i - a$ is $\theta^{-1}e^{-x/\theta}I_{(0,\infty)}(x)$, which does not depend on a. Therefore, for any fixed θ, $\sum_{i=1}^n (X_i - X_{(1)}) = \sum_{i=1}^n [(X_i - a) - (X_{(1)} - a)]$ is ancillary. By Basu's theorem (e.g., Theorem 2.4 in Shao, 2003), $\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ are independent for any fixed θ. Since θ is arbitrary, we conclude that $\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ are independent for any (a, θ).
(ii) From Example 5.14 in Lehmann (1983, p. 47), $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$ is sufficient and complete for (a, θ). Note that $(X_i - a)/\theta$ has Lebesgue density $e^{-x}I_{(0,\infty)}(x)$, which does not depend on (a, θ). Since

$$Z_i = \frac{X_{(n)} - X_{(i)}}{X_{(n)} - X_{(n-1)}} = \frac{\frac{X_{(n)}-a}{\theta} - \frac{X_{(i)}-a}{\theta}}{\frac{X_{(n)}-a}{\theta} - \frac{X_{(n-1)}-a}{\theta}},$$

the statistic $(Z_1, ..., Z_{n-2})$ is ancillary. By Basu's Theorem, $(Z_1, ..., Z_{n-2})$ is independent of $\left(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)})\right)$. ∎

Exercise 28 (#2.61). Let $(X_1, ..., X_n)$, $n > 2$, be a random sample of random variables having the uniform distribution on the interval $[a, b]$, where $-\infty < a < b < \infty$. Show that $Z_i = (X_{(i)} - X_{(1)})/(X_{(n)} - X_{(1)})$, $i = 2, ..., n-1$, are independent of $(X_{(1)}, X_{(n)})$ for any a and b, where $X_{(j)}$ is the jth order statistic.
Solution. Note that $(X_i - a)/(b - a)$ has the uniform distribution on the interval $[0, 1]$, which does not depend on any (a, b). Since

$$Z_i = \frac{X_{(i)} - X_{(1)}}{X_{(n)} - X_{(1)}} = \frac{\frac{X_{(i)}-a}{b-a} - \frac{X_{(1)}-a}{b-a}}{\frac{X_{(n)}-a}{b-a} - \frac{X_{(1)}-a}{b-a}},$$

the statistic $(Z_2, ..., Z_{n-1})$ is ancillary. By Basu's Theorem, the result follows if $(X_{(1)}, X_{(n)})$ is sufficient and complete for (a, b). The joint Lebesgue density of $X_1, ..., X_n$ is $(b-a)^{-n}I_{\{a<x_{(1)}<x_{(n)}<b\}}$. By the factorization theorem, $(X_{(1)}, X_{(n)})$ is sufficient for (a, b). The joint Lebesgue density of $(X_{(1)}, X_{(n)})$ is

$$\frac{n(n-1)}{(b-a)^n}(y-x)^{n-2}I_{\{a<x<y<b\}}.$$

For any Borel function $g(x,y)$, $E[g(X_{(1)},X_{(n)})] = 0$ for any $a < b$ implies that

$$\int_{a<x<y<b} g(x,y)(y-x)^{n-2}dxdy = 0$$

for any $a < b$. Hence $g(x,y)(y-x)^{n-2} = 0$ a.e. m^2, where m^2 is the Lebesgue measure on $\mathcal{R}^2$. Since $(y-x)^{n-2} \neq 0$ a.e. m^2, we conclude that $g(x,y) = 0$ a.e. m^2. Hence, $(X_{(1)},X_{(n)})$ is complete. ∎

Exercise 29 (#2.62). Let $(X_1,...,X_n)$, $n > 2$, be a random sample from a distribution P on $\mathcal{R}$ with $EX_1^2 < \infty$, $\bar{X}$ be the sample mean, $X_{(j)}$ be the jth order statistic, and $T = (X_{(1)} + X_{(n)})/2$. Consider the estimation of a parameter $\theta \in \mathcal{R}$ under the squared error loss.
(i) Show that $\bar{X}$ is better than T if $P = N(\theta,\sigma^2)$, $\theta \in \mathcal{R}$, $\sigma > 0$.
(ii) Show that T is better than $\bar{X}$ if P is the uniform distribution on the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathcal{R}$.
(iii) Find a family $\mathcal{P}$ for which neither $\bar{X}$ nor T is better than the other.
Solution. (i) Since $\bar{X}$ is complete and sufficient for θ and $T-\bar{X}$ is ancillary to θ, by Basu's theorem, $T-\bar{X}$ and $\bar{X}$ are independent. Then

$$R_T(\theta) = E[(T-\bar{X}) + (\bar{X}-\theta)]^2 = E(T-\bar{X})^2 + R_{\bar{X}}(\theta) > R_{\bar{X}}(\theta),$$

where the last inequality follows from the fact that $T \neq \bar{X}$ a.s. Therefore $\bar{X}$ is better.
(ii) Let $W = \frac{X_{(1)}-\theta+X_{(n)}-\theta}{2}$. Then the Lebesgue density of W is

$$f(w) = \begin{cases} n2^{n-1}\left(w+\frac{1}{2}\right)^{n-1} & -\frac{1}{2} < w < 0 \\ n2^{n-1}\left(\frac{1}{2}-w\right)^{n-1} & 0 < w < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $ET = EW + \theta = \theta$ and

$$R_T(\theta) = \text{Var}(T) = \text{Var}(W) = \frac{1}{2(n+1)(n+2)}.$$

On the other hand,

$$R_{\bar{X}}(\theta) = \text{Var}(\bar{X}) = \frac{\text{Var}(X_1)}{n} = \frac{1}{12n}.$$

Hence, when $n > 2$, $R_T(\theta) < R_{\bar{X}}(\theta)$.
(iii) Consider the family $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ is the family in part (i) and $\mathcal{P}_2$ is the family in part (ii). When $P \in \mathcal{P}_1$, $\bar{X}$ is better than T. When $P \in \mathcal{P}_2$, T is better than $\bar{X}$. Therefore, neither of them is better than the other for $P \in \mathcal{P}$. ∎

Exercise 30 (#2.64). Let $(X_1,...,X_n)$ be a random sample of binary random variables with $P(X_i = 1) = \theta \in (0,1)$. Consider estimating θ with

the squared error loss. Calculate the risks of the following estimators:
(i) the nonrandomized estimators $\bar{X}$ (the sample mean) and

$$T_0(X) = \begin{cases} 0 & \text{if more than half of } X_i\text{'s are } 0 \\ 1 & \text{if more than half of } X_i\text{'s are } 1 \\ \frac{1}{2} & \text{if exactly half of } X_i\text{'s are } 0; \end{cases}$$

(ii) the randomized estimators

$$T_1(X) = \begin{cases} \bar{X} & \text{with probability } \frac{1}{2} \\ T_0 & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$T_2(X) = \begin{cases} \bar{X} & \text{with probability } \bar{X} \\ \frac{1}{2} & \text{with probability } 1 - \bar{X}. \end{cases}$$

Solution. (i) Note that

$$\begin{aligned} R_{T_0}(\theta) &= E(T_0 - \theta)^2 \\ &= \theta^2 P(\bar{X} < 0.5) + (1-\theta)^2 P(\bar{X} > 0.5) + (0.5-\theta)^2 P(\bar{X} = 0.5). \end{aligned}$$

When $n = 2k$,

$$P(\bar{X} < 0.5) = \sum_{j=1}^{k-1} \binom{2k}{j} \theta^j (1-\theta)^{2k-j},$$

$$P(\bar{X} > 0.5) = \sum_{j=k+1}^{2k} \binom{2k}{j} \theta^j (1-\theta)^{2k-j},$$

and

$$P(\bar{X} = 0.5) = \binom{2k}{k} \theta^k (1-\theta)^k.$$

When $n = 2k+1$,

$$P(\bar{X} < 0.5) = \sum_{j=0}^{k} \binom{2k+1}{j} \theta^j (1-\theta)^{2k+1-j},$$

$$P(\bar{X} > 0.5) = \sum_{j=k+1}^{2k+1} \binom{2k+1}{j} \theta^j (1-\theta)^{2k+1-j},$$

and $P(\bar{X} = 0.5) = 0$.
(ii) A direct calculation shows that

$$\begin{aligned} R_{T_1}(\theta) &= E(T_1 - \theta)^2 \\ &= \frac{1}{2} E(\bar{X} - \theta)^2 + \frac{1}{2} E(T_0 - \theta)^2 \\ &= \frac{\theta(1-\theta)}{2n} + \frac{1}{2} R_{T_0}(\theta), \end{aligned}$$

where $R_{T_0}(\theta)$ is given in part (i), and

$$\begin{aligned}
R_{T_2}(\theta) &= E(T_2 - \theta)^2 \\
&= E\left[\bar{X}(\bar{X}-\theta)^2 + \left(\frac{1}{2}-\theta\right)^2 (1-\bar{X})\right] \\
&= E(\bar{X}-\theta)^3 + \theta E(\bar{X}-\theta)^2 + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n E(X_i-\theta)(X_j-\theta)(X_k-\theta) \\
&\quad + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{E(X_1-\theta)^3}{n^2} + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{\theta(1-\theta)^3 - \theta^3(1-\theta)}{n^2} + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta),
\end{aligned}$$

where the fourth equality follows from $E(\bar{X}-\theta)^2 = \mathrm{Var}(\bar{X}) = \theta(1-\theta)/n$ and the fifth equality follows from the fact that $E(X_i-\theta)(X_j-\theta)(X_k-\theta) \neq 0$ if and only if $i = j = k$. ∎

Exercise 31 (#2.66). Consider the estimation of an unknown parameter $\theta \geq 0$ under the squared error loss. Show that if T and U are two estimators such that $T \leq U$ and $R_T(P) < R_U(P)$, then $R_{T_+}(P) < R_{U_+}(P)$, where $R_T(P)$ is the risk of an estimator T and T_+ denotes the positive part of T.
Solution. Note that $T = T_+ - T_-$, where $T_- = \max\{-T, 0\}$ is the negative part of T, and $T_+T_- = 0$. Then

$$\begin{aligned}
R_T(P) &= E(T-\theta)^2 \\
&= E(T_+ - T_- - \theta)^2 \\
&= E(T_+ - \theta)^2 + E(T_-^2) + 2\theta E(T_-) - 2E(T_+T_-) \\
&= R_{T_+}(P) + E(T_-^2) + 2\theta E(T_-).
\end{aligned}$$

Similarly,

$$R_U(P) = R_{U_+}(P) + E(U_-^2) + 2\theta E(U_-).$$

Since $T \leq U$, $T_- \geq U_-$. Also, $\theta \geq 0$. Hence,

$$E(T_-^2) + 2\theta E(T_-) \geq E(U_-^2) + 2\theta E(U_-).$$

Since $R_T(P) < R_U(P)$, we must have $R_{T_+}(P) < R_{U_+}(P)$. ∎

Exercise 32. Consider the estimation of an unknown parameter $\theta \in \mathcal{R}$ under the squared error loss. Show that if T and U are two estimators such

that $P(\theta - t < T < \theta + t) \geq P(\theta - t < U < \theta + t)$ for any $t > 0$, then $R_T(P) \leq R_U(P)$.
Solution. From the condition,

$$P((T-\theta)^2 > s) \leq P((U-\theta)^2 > s)$$

for any $s > 0$. Hence,

$$\begin{aligned} R_T(P) &= E(T-\theta)^2 \\ &= \int_0^\infty P((T-\theta)^2 > s)ds \\ &\leq \int_0^\infty P((U-\theta)^2 > s)ds \\ &= E(U-\theta)^2 \\ &= R_U(P). \quad \blacksquare \end{aligned}$$

Exercise 33 (#2.67). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on $(0, \infty)$ with scale parameter $\theta \in (0, \infty)$. Consider the hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is a fixed constant. Obtain the risk function (in terms of θ) of the test rule $T_c = I_{(c,\infty)}(\bar{X})$ under the 0-1 loss, where $\bar{X}$ is the sample mean and $c > 0$ is a constant.
Solution. Let $L(\theta, a)$ be the loss function. Then $L(\theta, 1) = 0$ when $\theta > \theta_0$, $L(\theta, 1) = 1$ when $\theta \leq \theta_0$, $L(\theta, 0) = 0$ when $\theta \leq \theta_0$, and $L(\theta, 0) = 1$ when $\theta > \theta_0$. Hence,

$$\begin{aligned} R_{T_c}(\theta) &= E[L(\theta, I_{(c,\infty)}(\bar{X}))] \\ &= E\left[L(\theta,1)I_{(c,\infty)}(\bar{X}) + L(\theta,0)I_{(0,c]}(\bar{X})\right] \\ &= L(\theta,1)P(\bar{X} > c) + L(\theta,0)P(\bar{X} \leq c) \\ &= \begin{cases} P(\bar{X} > c) & \theta \leq \theta_0 \\ P(\bar{X} \leq c) & \theta > \theta_0. \end{cases} \end{aligned}$$

Since $n\bar{X}$ has the gamma distribution with shape parameter n and scale parameter θ,

$$P(\bar{X} > c) = \frac{1}{\theta^n (n-1)!} \int_{nc}^\infty x^{n-1} e^{-x/\theta} dx. \quad \blacksquare$$

Exercise 34 (#2.71). Consider an estimation problem with a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and the squared error loss. If $\theta_0 \in \Theta$ satisfies that $P_\theta \ll P_{\theta_0}$ for any $\theta \in \Theta$, show that the estimator $T \equiv \theta_0$ is admissible.

Solution. Note that the risk $R_T(\theta) = 0$ when $\theta = \theta_0$. Suppose that U is an estimator of θ and $R_U(\theta) = E(U-\theta)^2 \leq R_T(\theta)$ for all θ. Then $R_U(\theta_0) = 0$, i.e., $E(U-\theta_0)^2 = 0$ under P_{θ_0}. Therefore, $U = \theta_0$ a.s. P_{θ_0}. Since $P_\theta \ll P_{\theta_0}$ for any θ, we conclude that $U = \theta_0$ a.s. $\mathcal{P}$. Hence $U = T$ a.s. $\mathcal{P}$. Thus, T is admissible. ∎

Exercise 35 (#2.73). Let $(X_1, ..., X_n)$ be a random sample of random variables with $EX_1^2 < \infty$. Consider estimating $\mu = EX_1$ under the squared error loss. Show that
(i) any estimator of the form $a\bar{X} + b$ is inadmissible, where $\bar{X}$ is the sample mean, a and b are constants, and $a > 1$;
(ii) any estimator of the form $\bar{X} + b$ is inadmissible, where $b \neq 0$ is a constant.
Solution. (i) Note that

$$\begin{aligned} R_{a\bar{X}+b}(P) &= E(a\bar{X} + b - \mu)^2 \\ &= a^2\mathrm{Var}(\bar{X}) + (a\mu + b - \mu)^2 \\ &\geq a^2\mathrm{Var}(\bar{X}) \\ &= a^2 R_{\bar{X}}(P) \\ &> R_{\bar{X}}(P) \end{aligned}$$

when $a > 1$. Hence $\bar{X}$ is better than $a\bar{X} + b$ with $a > 1$.
(ii) For $b \neq 0$,

$$R_{\bar{X}+b}(P) = E(\bar{X} + b - \mu)^2 = \mathrm{Var}(\bar{X}) + b^2 > \mathrm{Var}(\bar{X}) = R_{\bar{X}}(P).$$

Hence $\bar{X}$ is better than $\bar{X} + b$ with $b \neq 0$. ∎

Exercise 36 (#2.74). Consider an estimation problem with $\vartheta \in [c, d] \subset \mathcal{R}$, where c and d are known. Suppose that the action space contains $[c, d]$ and the loss function is $L(|\vartheta - a|)$, where $L(\cdot)$ is an increasing function on $[0, \infty)$. Show that any decision rule T with $P(T(X) \notin [c, d]) > 0$ for some $P \in \mathcal{P}$ is inadmissible.
Solution. Consider the decision rule

$$T_1 = cI_{(-\infty,c)}(T) + TI_{[c,d]}(T) + dI_{(d,\infty)}(T).$$

Then $|T_1 - \vartheta| \leq |T - \vartheta|$ and, since L is an increasing function,

$$R_{T_1}(P) = E[L(|T_1 - \vartheta|)] \leq E[L(|T - \vartheta|)] = R_T(P)$$

for any $P \in \mathcal{P}$. Since

$$P(|T_1(X) - \vartheta| < |T(X) - \vartheta|) = P(T(X) \notin [a, b]) > 0$$

holds for some $P_* \in \mathcal{P}$,

$$R_{T_1}(P_*) < R_T(P_*).$$

Hence T_1 is better than T and T is inadmissible. ∎

Exercise 37 (#2.75). Let X be a sample from $P \in \mathcal{P}$, $\delta_0(X)$ be a nonrandomized rule in a decision problem with $\mathcal{R}^k$ as the action space, and T be a sufficient statistic for $P \in \mathcal{P}$. Show that if $E[I_A(\delta_0(X))|T]$ is a nonrandomized rule, i.e., $E[I_A(\delta_0(X))|T] = I_A(h(T))$ for any Borel $A \subset \mathcal{R}^k$, where h is a Borel function, then $\delta_0(X) = h(T(X))$ a.s. P.
Solution. From the assumption,

$$E\left[\sum_{i=1}^n c_i I_{A_i}(\delta_0(X))\bigg|T\right] = \sum_{i=1}^n c_i I_{A_i}(h(T))$$

for any positive integer n, constants $c_1, ..., c_n$, and Borel sets $A_1, ..., A_n$. Using the results in Exercise 39 of Chapter 1, we conclude that for any bounded continuous function f, $E[f(\delta_0(X))|T] = f(h(T))$ a.s. P. Then, by the result in Exercise 45 of Chapter 1, $\delta_0(X) = h(T)$ a.s. P. ∎

Exercise 38 (#2.76). Let X be a sample from $P \in \mathcal{P}$, $\delta_0(X)$ be a decision rule (which may be randomized) in a problem with $\mathcal{R}^k$ as the action space, and T be a sufficient statistic for $P \in \mathcal{P}$. For any Borel $A \subset \mathcal{R}^k$, define

$$\delta_1(T, A) = E[\delta_0(X, A)|T].$$

Let $L(P, a)$ be a loss function. Show that

$$\int_{\mathcal{R}^k} L(P, a)d\delta_1(X, a) = E\left[\int_{\mathcal{R}^k} L(P, a)d\delta_0(X, a)\bigg|T\right] \quad \text{a.s. } P.$$

Solution. If L is a simple function (a linear combination of indicator functions), then the result follows from the definition of δ_1. For nonnegative L, it is the limit of a sequence of nonnegative increasing simple functions. Then the result follows from the result for simple L and the monotone convergence theorem for conditional expectations (Exercise 38 in Chapter 1). ∎

Exercise 39 (#2.80). Let $X_1, ..., X_n$ be random variables with a finite common mean $\mu = EX_i$ and finite variances. Consider the estimation of μ under the squared error loss.
(i) Show that there is no optimal rule in $\Im$ if $\Im$ contains all possible estimators.
(ii) Find an optimal rule in

$$\Im_2 = \left\{\sum_{i=1}^n c_i X_i : c_i \in \mathcal{R}, \sum_{i=1}^n c_i = 1\right\}$$

if $\text{Var}(X_i) = \sigma^2/a_i$ with an unknown σ^2 and known a_i, $i = 1, ..., n$.
(iii) Find an optimal rule in $\Im_2$ if $X_1, ..., X_n$ are identically distributed but

are correlated with correlation coefficient ρ.

Solution. (i) Suppose that there exists an optimal rule T^*. Let P_1 and P_2 be two possible distributions of $X = (X_1, ..., X_n)$ such that $\mu = \mu_j$ under P_j and $\mu_1 \neq \mu_2$. Let $R_T(P)$ be the risk of T. For $T_1(X) \equiv \mu_1$, $R_{T_1}(P_1) = 0$. Since T^* is better than T_1, $R_{T^*}(P_1) \leq R_{T_1}(P_1) = 0$ and, hence, $T^* \equiv \mu_1$ a.s. P_1. Let $\bar{P} = (P_1 + P_2)/2$. If X has distribution $\bar{P}$, then $\mu = (\mu_1 + \mu_2)/2$. Let $T_0(X) \equiv (\mu_1 + \mu_2)/2$. Then $R_{T_0}(\bar{P}) = 0$. Since T^* is better than T_0, $R_{T^*}(\bar{P}) = 0$ and, hence, $T^* \equiv (\mu_1 + \mu_2)/2$ a.s. $\bar{P}$, which implies that $T^* \equiv (\mu_1 + \mu_2)/2$ a.s. P_1 since $P_1 \ll \bar{P}$. This is impossible since $\mu_1 \neq (\mu_1 + \mu_2)/2$.

(ii) Let $T = \sum_{i=1}^n c_i X_i$ and $T^* = \sum_{i=1}^n a_i X_i / \sum_{i=1}^n a_i$. Then

$$
\begin{aligned}
R_{T^*}(P) &= \mathrm{Var}(T^*) \\
&= \mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\
&= \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i) \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\
&= \sum_{i=1}^n a_i \sigma^2 \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\
&= \sigma^2 \Big/ \left(\sum_{i=1}^n a_i\right).
\end{aligned}
$$

By the Cauchy-Schwarz inequality,

$$
\left(\sum_{i=1}^n a_i\right)\left(\sum_{i=1}^n \frac{c_i^2}{a_i}\right) \geq \left(\sum_{i=1}^n c_i\right)^2 = 1.
$$

Hence,

$$
R_{T^*}(P) \leq \sigma^2 \sum_{i=1}^n \frac{c_i^2}{a_i} = \mathrm{Var}\left(\sum_{i=1}^n c_i X_i\right) = \mathrm{Var}(T) = R_T(P).
$$

Therefore T^* is optimal.

(iii) For any $T = \sum_{i=1}^n c_i X_i$,

$$
\begin{aligned}
R_T(P) &= \mathrm{Var}(T) \\
&= \sum_{i=1}^n c_i^2 \sigma^2 + \sum_{i \neq j} c_i c_j \rho \sigma^2 \\
&= \sum_{i=1}^n c_i^2 \sigma^2 + \rho\sigma^2 \left[\left(\sum_{i=1}^n c_i\right)^2 - \sum_{i=1}^n c_i^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma^2 \left[(1-\rho) \sum_{i=1}^n c_i^2 + \rho \right] \\
&\geq \sigma^2 \left[(1-\rho) \left(\sum_{i=1}^n c_i \right)^2 \Big/ n + \rho \right] \\
&= \sigma^2 \left[(1-\rho)/n + \rho \right] \\
&= \mathrm{Var}(\bar{X}),
\end{aligned}
$$

where the last equality follows from the Cauchy–Schwarz inequality

$$
\left(\sum_{i=1}^n c_i \right)^2 \leq n \sum_{i=1}^n c_i^2 .
$$

Hence, the sample mean $\bar{X}$ is optimal. ∎

Exercise 40 (#2.83). Let X be a discrete random variable with

$$
P(X=-1) = p, \quad P(X=k) = (1-p)^2 p^k, \quad k = 0, 1, 2, ...,
$$

where $p \in (0,1)$ is unknown. Show that
(i) $U(X)$ is an unbiased estimator of 0 if and only if $U(k) = ak$ for all $k = -1, 0, 1, 2, ...$ and some a;
(ii) $T_0(X) = I_{\{0\}}(X)$ is unbiased for $(1-p)^2$ and, under the squared error loss, T_0 is an optimal rule in $\Im$, where $\Im$ is the class of all unbiased estimators of $(1-p)^2$;
(iii) $T_0(X) = I_{\{-1\}}(X)$ is unbiased for p and, under the squared error loss, there is no optimal rule in $\Im$, where $\Im$ is the class of all unbiased estimators of p.
Solution. (i) If $U(X)$ is unbiased for 0, then

$$
\begin{aligned}
E[U(X)] &= U(-1)p + \sum_{k=0}^\infty U(k)(1-p)^2 p^k \\
&= \sum_{k=0}^\infty U(k)p^k - 2\sum_{k=0}^\infty U(k)p^{k+1} + U(-1)p + \sum_{k=0}^\infty U(k)p^{k+2} \\
&= U(0) + \sum_{k=-1}^\infty U(k+2)p^{k+2} - 2\sum_{k=-1}^\infty U(k+1)p^{k+2} \\
&\quad + \sum_{k=-1}^\infty U(k)p^{k+2} \\
&= \sum_{k=-1}^\infty [U(k) - 2U(k+1) + U(k+2)]p^{k+2} \\
&= 0
\end{aligned}
$$

for all p, which implies $U(0)=0$ and $U(k)-2U(k+1)+U(k+2)=0$ for $k=-1,0,1,2,...$, or equivalently, $U(k)=ak$, where $a=U(1)$.
(ii) Since
$$E[T_0(X)]=P(X=0)=(1-p)^2,$$
T_0 is unbiased. Let T be another unbiased estimator of $(1-p)^2$. Then $T(X)-T_0(X)$ is unbiased for 0 and, by the result in (i), $T(X)=T_0(X)+aX$ for some a. Then,
$$\begin{aligned}
R_T(p) &= E[T_0(X)+aX-(1-p)^2]^2 \\
&= E(T_0+aX)^2+(1-p)^4-2(1-p)^2E[T_0(X)+aX] \\
&= E(T_0+aX)^2-(1-p)^4 \\
&= a^2P(X=-1)+P(X=0)+a^2\sum_{k=1}^{\infty}k^2P(X=k)-(1-p)^4 \\
&\geq P(X=0)-(1-p)^4 \\
&= \mathrm{Var}(T_0).
\end{aligned}$$
Hence T_0 is a optimal rule in $\Im$.
(iii) Since
$$E[T_0(X)]=P(X=-1)=p,$$
T_0 is unbiased. Let T be another unbiased estimator of p. Then $T(X)=T_0(X)+aX$ for some a and
$$\begin{aligned}
R_T(p) &= E(T_0+aX)^2-p^2 \\
&= (1-a)^2p+a^2\sum_{k=0}^{\infty}k^2(1-p)p^2-p^2,
\end{aligned}$$
which is a quadratic function in a with minimum
$$a=\left[1+(1-p)\sum_{k=1}^{\infty}k^2p^{k-1}\right]^{-1}$$
depending on p. Therefore, there is no optimal rule in $\Im$. ∎

Exercise 41. Let X be a random sample from a population and θ be an unknown parameter. Suppose that there are $k+1$ estimators of θ, $T_1,...,T_{k+1}$, such that $ET_i=\theta+\sum_{j=1}^{k}c_{i,j}b_j(\theta)$, $i=1,...,k+1$, where $c_{i,j}$'s are constants and $b_j(\theta)$ are functions of θ. Suppose that the determinant
$$C=\begin{vmatrix} 1 & 1 & \cdots & 1 \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix}\neq 0.$$

Show that

$$T^* = \frac{1}{C}\begin{vmatrix} T_1 & T_2 & \cdots & T_{k+1} \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix}$$

is an unbiased estimator of θ.

Solution. From the properties of a determinant,

$$\begin{aligned} ET^* &= \frac{1}{C}\begin{vmatrix} ET_1 & ET_2 & \cdots & ET_{k+1} \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \frac{1}{C}\begin{vmatrix} \theta + \sum_{j=1}^k c_{1,j}b_j(\theta) & \cdots & \theta + \sum_{j=1}^k c_{k+1,j}b_j(\theta) \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \frac{\theta}{C}\begin{vmatrix} 1 & \cdots & 1 \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &\quad + \frac{1}{C}\begin{vmatrix} \sum_{j=1}^k c_{1,j}b_j(\theta) & \cdots & \sum_{j=1}^k c_{k+1,j}b_j(\theta) \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \theta, \end{aligned}$$

where the last equality follows from the fact that the last determinant is 0 because its first row is a linear combination of its other k rows. ∎

Exercise 42 (#2.84). Let X be a random variable having the binomial distribution with size n and probability $p \in (0,1)$. Show that there is no unbiased estimator of p^{-1}.

Solution. Suppose that $T(X)$ is an unbiased estimator of p^{-1}. Then

$$E[T(X)] = \sum_{k=0}^n \binom{n}{k} T(k)p^k(1-p)^{n-k} = \frac{1}{p}$$

for all p. However,

$$\sum_{k=0}^n \binom{n}{k} T(k)p^k(1-p)^{n-k} \le \sum_{k=0}^n \binom{n}{k} T(k) < \infty$$

for any p but p^{-1} diverges to ∞ as $p \to 0$. This is impossible. Hence, there is no unbiased estimator of p^{-1}. ∎

Exercise 43 (#2.85). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\theta, 1)$, where $\theta = 0$ or 1. Consider the estimation of θ with action space $\{0, 1\}$, i.e., the range of any estimator is $\{0, 1\}$.
(i) Show that there does not exist any unbiased estimator of θ.
(ii) Find an estimator $\hat{\theta}$ of θ that is approximately unbiased, that is, $\lim_n E(\hat{\theta}) = \theta$.
Solution. (i) Since the action space is $\{0, 1\}$, any randomized estimator $\hat{\theta}$ can be written as $T(X)$, where T is Borel, $0 \le T(X) \le 1$, and

$$\hat{\theta} = \begin{cases} 1 & \text{with probability } T(X) \\ 0 & \text{with probability } 1 - T(X). \end{cases}$$

Then $E(\hat{\theta}) = E[T(X)]$. If $\hat{\theta}$ is unbiased, then $E[T(X)] = \theta$ for $\theta = 0, 1$. This implies that, when $\theta = 0$, $T(X) = 0$ a.e. Lebesgue measure, whereas when $\theta = 1$, $T(X) = 1$ a.e. Lebesgue measure. This is impossible. Hence there does not exist any unbiased estimator of θ.
(ii) Consider $\hat{\theta} = I_{(n^{-1/4}, \infty)}(|\bar{X}|)$, where $\bar{X}$ is the sample mean. Since $\bar{X}$ is distributed as $N(\theta, n^{-1})$,

$$E(\hat{\theta}) = P(|\bar{X}| > n^{-1/4}) = 1 - \Phi\left(n^{1/4} - \theta\sqrt{n}\right) + \Phi\left(-n^{1/4} - \theta\sqrt{n}\right),$$

where Φ is the cumulative distribution function of $N(0, 1)$. Hence, when $\theta = 0$, $\lim_n E(\hat{\theta}) = 1 - \Phi(\infty) + \Phi(-\infty) = 0$ and, when $\theta = 1$, $\lim_n E(\hat{\theta}) = 1 - \Phi(-\infty) + \Phi(-\infty) = 1$. ∎

Exercise 44 (#2.92(c)). Let X be a sample from P_θ, where $\theta \in \Theta \subset \mathcal{R}$. Consider the estimation of θ under the absolute error loss function $|a - \theta|$. Let Π be a given distribution on Θ with finite mean. Find a Bayes rule.
Solution. Let $P_{\theta|X}$ be the posterior distribution of θ and P_X be the marginal of X. By Fubini's theorem,

$$\int \int |\theta| dP_{\theta|X} dP_X = \int \int |\theta| dP_\theta d\Pi = \int |\theta| d\Pi < \infty.$$

Hence, for almost all X, $\int |\theta| dP_{\theta|X} < \infty$. From Exercise 11 in Chapter 1, if m_X is a median of $P_{\theta|X}$, then

$$\int |\theta - m_X| dP_{\theta|X} \le \int |\theta - a| dP_{\theta|X} \quad \text{for almost all } X$$

holds for any a. Hence, $E|\theta - m_X| \le E|\theta - T(X)|$ for any other estimator $T(X)$. This shows that m_X is a Bayes rule. ∎

Exercise 45 (#2.93). Let X be a sample having a probability density $f_j(x)$ with respect to a σ-finite measure ν, where j is unknown and $j \in \{1, ..., J\}$ with a known integer $J \geq 2$. Consider a decision problem in which the action space is $\{1, ..., J\}$ and the loss function is

$$L(j, a) = \begin{cases} 0 & \text{if } a = j \\ 1 & \text{if } a \neq j. \end{cases}$$

(i) Obtain the risk of a decision rule (which may be randomized).
(ii) Let Π be a prior probability measure on $\{1, ..., J\}$ with $\Pi(\{j\}) = \pi_j$, $j = 1, ..., J$. Obtain the Bayes risk of a decision rule.
(iii) Obtain a Bayes rule under the prior Π in (ii).
(iv) Assume that $J = 2$, $\pi_1 = \pi_2 = 0.5$, and $f_j(x) = \phi(x - \mu_j)$, where $\phi(x)$ is the Lebesgue density of the standard normal distribution and μ_j, $j = 1, 2$, are known constants. Obtain the Bayes rule in (iii).
(v) Obtain a minimax rule when $J = 2$.
Solution. (i) Let δ be a randomized decision rule. For any X, let $\delta(X, j)$ be the probability of taking action j under the rule δ. Let E_j be the expectation taking under f_j. Then

$$R_\delta(j) = E_j\left[\sum_{k=1}^J L(j,k)\delta(X,k)\right] = \sum_{k\neq j} E_j[\delta(X,k)] = 1 - E_j[\delta(X,j)],$$

since $\sum_{k=1}^J \delta(X, k) = 1$.
(ii) The Bayes risk of a decision rule δ is

$$r_\delta = \sum_{j=1}^J \pi_j R_\delta(j) = 1 - \sum_{j=1}^J \pi_j E_j[\delta(X,j)].$$

(iii) Let δ^* be a rule satisfying $\delta^*(X, j) = 1$ if and only if $\pi_j f_j(X) = g(X)$, where $g(X) = \max_{1\leq k\leq J} \pi_k f_k(X)$. Then δ_* is a Bayes rule, since, for any rule δ,

$$\begin{aligned} r_\delta &= 1 - \sum_{j=1}^J \int \pi_j \delta(x,j) f_j(x) d\nu \\ &\geq 1 - \sum_{j=1}^J \int \delta(x,j) g(x) d\nu \\ &= 1 - \int g(x) d\nu \\ &= 1 - \sum_{j=1}^J \int_{g(x)=\pi_j f_j(x)} \pi_j f_j(x) d\nu \\ &= r_{\delta^*}. \end{aligned}$$

(iv) From the result in (iii), the Bayes rule $\delta^*(X, j) = 1$ if and only if $\phi(x - \mu_j) > \phi(x - \mu_k)$, $k \neq j$. Since $\phi(x - \mu_j) = e^{-(x-\mu_j)^2/2}/\sqrt{2\pi}$, we can obtain a nonrandomized Bayes rule that takes action 1 if and only if $|X - \mu_1| < |X - \mu_2|$.
(v) Let c be a positive constant and consider a rule δ_c such that $\delta_c(X, 1) = 1$ if $f_1(X) > cf_2(X)$, $\delta_c(X, 2) = 1$ if $f_1(X) < cf_2(X)$, and $\delta_c(X, 1) = \gamma$ if $f_1(X) = cf_2(X)$. Since $\delta_c(X, j) = 1$ if and only if $\pi_j f_j(X) = \max_k \pi_k f_k(X)$, where $\pi_1 = 1/(c + 1)$ and $\pi_2 = c/(c + 1)$, it follows from part (iii) of the solution that δ_c is a Bayes rule. Let P_j be the probability corresponding to f_j. The risk of δ_c is $P_1(f_1(X) \leq cf_2(X)) - \gamma P_1(f_1(x) = cf_2(X))$ when $j = 1$ and $1 - P_2(f_1(X) \leq cf_2(X)) + \gamma P_2(f_1(x) = cf_2(X))$ when $j = 2$. Let $\psi(c) = P_1(f_1(X) \leq cf_2(X)) + P_2(f_1(X) \leq cf_2(X)) - 1$. Then ψ is nondecreasing in c, $\psi(0) = -1$, $\lim_{c\to\infty} \psi(c) = 1$, and $\psi(c) - \psi(c-) = P_1(f_1(X) = cf_2(X)) + P_2(f_1(X) = cf_2(X))$. Let $c_* = \inf\{c : \psi(c) \geq 0\}$. If $\psi(c_*) = \psi(c_*-)$, we set $\gamma = 0$; otherwise, we set $\gamma = \psi(c_*)/[\psi(c_*)-\psi(c_*-)]$. Then, the risk of δ_{c_*} is a constant. For any rule δ, $\sup_j R_\delta(j) \geq r_\delta \geq r_{\delta_{c_*}} = R_{\delta_{c_*}}(j) = \sup_j R_{\delta_{c_*}}(j)$. Hence, δ_{c_*} is a minimax rule. ∎

Exercise 46 (#2.94). Let $\hat{\theta}$ be an unbiased estimator of an unknown $\theta \in \mathcal{R}$.
(i) Under the squared error loss, show that the estimator $\hat{\theta}+c$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T, where $c \neq 0$ is a known constant.
(ii) Under the squared error loss, show that the estimator $c\hat{\theta}$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T, where $c \in (0, 1)$ is a known constant.
(iii) Consider the loss function $L(\theta, a) = (a - \theta)^2/\theta^2$ (assuming $\theta \neq 0$). Show that $\hat{\theta}$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any T.
Solution. (i) Under the squared error loss, the risk of $\hat{\theta} + c$ is

$$R_{\hat{\theta}+c}(P) = E(\hat{\theta} + c - \theta)^2 = c^2 + \mathrm{Var}(\hat{\theta}) = c^2 + R_{\hat{\theta}}(P).$$

Then

$$\sup_P R_{\hat{\theta}+c}(P) = c^2 + \sup_P R_{\hat{\theta}}(P)$$

and either $\sup_P R_{\hat{\theta}+c}(P) = \infty$ or $\sup_P R_{\hat{\theta}+c}(P) > \sup_P R_{\hat{\theta}}(P)$. Hence, the only case where $\hat{\theta} + c$ is minimax is when $\sup_P R_T(P) = \infty$ for any estimator T.
(ii) Under the squared error loss, the risk of $c\hat{\theta}$ is

$$R_{c\hat{\theta}}(P) = E(c\hat{\theta} - \theta)^2 = (1 - c)^2\theta^2 + c^2\mathrm{Var}(\hat{\theta}) = (1 - c)^2\theta^2 + c^2 R_{\hat{\theta}}(P).$$

Then, $\sup_P R_{\hat{\theta}+c}(P) = \infty$ and the only case where $\hat{\theta}+c$ is minimax is when $\sup_P R_T(P) = \infty$ for any estimator T.

(iii) Under the given loss function, the risk of $c\hat{\theta}$ is

$$R_{c\hat{\theta}}(P) = (1-c)^2 + c^2 R_{\hat{\theta}}(P).$$

If $\sup_P R_{\hat{\theta}}(P) = \infty$, then the result follows. Assume $\xi = \sup_P R_{\hat{\theta}}(P) < \infty$. Let $c = \xi/(\xi+1)$. Then

$$\sup_P R_{c\hat{\theta}}(P) = (1-c)^2 + c^2\xi = \frac{\xi^2}{(\xi+1)^2} + \frac{\xi}{(\xi+1)^2} = \frac{\xi}{\xi+1} < \xi.$$

Hence $\hat{\theta}$ is not minimax. ∎

Exercise 47 (#2.96). Let X be an observation from the binomial distribution with size n and probability $\theta \in (0,1)$, where n is a known integer ≥ 2. Consider testing hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 \in (0,1)$ is a fixed value. Let $\Im = \{T_j : j = 0, 1, ..., n-1\}$ be a class of nonrandomized decision rules, where $T_j(X) = 1$ (rejecting H_0) if and only if $X \geq j+1$. Consider the 0-1 loss function.
(i) When the uniform distribution on $(0,1)$ is used as the prior, show that the Bayes rule within the class $\Im$ is $T_{j^*}(X)$, where j^* is the largest integer in $\{0, 1, ..., n-1\}$ such that $B_{j+1,n-j+1}(\theta_0) \geq \frac{1}{2}$ and $B_{a,b}(\cdot)$ denotes the cumulative distribution function of the beta distribution with parameter (a, b).
(ii) Derive a minimax rule over the class $\Im$.
Solution. (i) Let P_θ be the probability law of X. Under the 0-1 loss, the risk of T_j is

$$\begin{aligned} R_{T_j}(\theta) &= P_\theta(X > j) I_{(0,\theta_0]}(\theta) + P_\theta(X \leq j) I_{(\theta_0,1)}(\theta) \\ &= \sum_{k=j+1}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} I_{(0,\theta_0]}(\theta) + \sum_{k=0}^{j} \binom{n}{k} \theta^k (1-\theta)^{n-k} I_{(\theta_0,1)}(\theta). \end{aligned}$$

Hence, the Bayes risk of T_j is

$$\begin{aligned} r_{T_j} &= \sum_{k=j+1}^{n} \binom{n}{k} \int_0^{\theta_0} \theta^k (1-\theta)^{n-k} d\theta + \sum_{k=0}^{j} \binom{n}{k} \int_{\theta_0}^{1} \theta^k (1-\theta)^{n-k} d\theta \\ &= \sum_{k=j+1}^{n} B_{k+1,n-k+1}(\theta_0) + \sum_{k=0}^{j} [1 - B_{k+1,n-k+1}(\theta_0)]. \end{aligned}$$

Then, for $j = 1, ..., n-1$,

$$r_{T_{j-1}} - r_{T_j} = 2B_{j+1,n-j+1}(\theta_0) - 1.$$

The family $\{B_{\beta+1,n-\beta+1}(y) : \beta > 0\}$ is an exponential family having monotone likelihood ratio in $\log y - \log(1-y)$. By Lemma 6.3 in Shao (2003), if

Y has distribution $B_{\beta+1,n-\beta+1}$, then $P(Y \leq t) = P(\log Y - \log(1-Y) \leq \log t - \log(1-t))$ is decreasing in β for any fixed $t \in (0,1)$. This shows that $B_{j+1,n-j+1}(\theta_0)$ is decreasing in j. Hence, if j^* is the largest integer j such that $B_{j+1,n-j+1}(\theta_0) \geq \frac{1}{2}$, then

$$r_{T_{j-1}} - r_{T_j} \geq 0 \quad j = 1, ..., j^*$$

and

$$r_{T_{j-1}} - r_{T_j} \leq 0 \quad j = j^*+1, ..., n-1.$$

Consequently,

$$r_{T_{j^*}} = \min_{j=0,1,...,n-1} r_{T_j}.$$

This shows that T_{j^*} is the Bayes rule over the class $\Im$.
(ii) Again, by Lemma 6.3 in Shao (2003), $P_\theta(X \leq j)$ is decreasing in θ and $P_\theta(X > j)$ is increasing in θ. Hence,

$$\sup_{\theta\in(0,1)} R_{T_j}(\theta) = P_{\theta_0}(X > j) = \sum_{k=j+1}^{n} \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k}.$$

Then, the minimax rule over the class $\Im$ is T_{n-1}. ∎

Exercise 48 (#2.99). Let $(X_1, ..., X_n)$ be a random sample from the Cauchy distribution with location parameter $\mu \in \mathcal{R}$ and a known scale parameter $\sigma > 0$. Consider the hypotheses $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, where μ_0 is a fixed constant. Calculate the size of the nonrandomized test $T_c(X) = I_{(c,\infty)}(\bar{X})$, where c is a fixed constant; find a c_α such that T_{c_α} has size $\alpha \in (0,1)$; and find the p-value for T_{c_α}.
Solution: Note that $\bar{X}$ has the same distribution as X_1. Hence, the size of $T_c(X)$ is

$$\begin{aligned} \sup_{\mu\leq\mu_0} E(T_c(X)) &= \sup_{\mu\leq\mu_0} P(\bar{X} > c) \\ &= \sup_{\mu\leq\mu_0} P\left(\frac{\bar{X}-\mu}{\sigma} > \frac{c-\mu}{\sigma}\right) \\ &= \sup_{\mu\leq\mu_0} \left[\frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c-\mu}{\sigma}\right)\right] \\ &= \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c-\mu_0}{\sigma}\right). \end{aligned}$$

Therefore, if $c_\alpha = \mu_0 + \sigma\tan\left(\pi(\frac{1}{2}-\alpha)\right)$, then the size of $T_{c_\alpha}(X)$ is exactly α. Note that

$$\alpha = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c_\alpha-\mu_0}{\sigma}\right) > \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\bar{X}-\mu_0}{\sigma}\right)$$

if and only if $\bar{X} > c_\alpha$ (i.e., $T_{c_\alpha}(X) = 1$). Hence, the p-value of $T_{c_\alpha}(X)$ is

$$\inf\{\alpha | T_{c_\alpha}(X) = 1\} = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\bar{X} - \mu_0}{\sigma}\right). \blacksquare$$

Exercise 49 (#2.101). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown parameters. Let $\alpha \in (0, 1)$ be given.
(i) Using $T_1(X) = \sum_{i=1}^n (X_i - X_{(1)})$, where $X_{(1)}$ is the smallest order statistic, construct a confidence interval for θ with confidence coefficient $1 - \alpha$ and find the expected interval length.
(ii) Using $T_1(X)$ and $T_2(X) = X_{(1)}$, construct a confidence interval for a with confidence coefficient $1 - \alpha$ and find the expected interval length.
(iii) Construct a confidence set for the two-dimensional parameter (a, θ) with confidence coefficient $1 - \alpha$.
Solution. (i) Let $W = T_1(X)/\theta$. Then W has the gamma distribution with shape parameter $n - 1$ and scale parameter 1. Let $c_1 < c_2$ such that $P(c_1 < W < c_2) = 1 - \alpha$. Then c_1 and c_2 can be chosen so that they do not depend on unknown parameters. A confidence interval for θ with confidence coefficient $1 - \alpha$ is

$$\left(\frac{T_1(X)}{c_2}, \frac{T_1(X)}{c_1}\right).$$

Its expected length is

$$\left(\frac{1}{c_1} - \frac{1}{c_2}\right) E(T_1) = \left(\frac{1}{c_1} - \frac{1}{c_2}\right)(n-1)\theta.$$

(ii) Using the result in Exercise 7(iii), $[T_2(X) - a]/T_1(X)$ has the Lebesgue density $n\left(1 + \frac{nt}{n-1}\right)^{-n} I_{(0,\infty)}(t)$, which does not depend on any unknown parameter. Choose two constants $0 < c_1 < c_2$ such that

$$\int_{c_1}^{c_2} n\left(1 + \frac{nt}{n-1}\right)^{-n} dt = 1 - \alpha.$$

Then a confidence interval for a with confidence coefficient $1 - \alpha$ is

$$(T_2 - c_2 T_1, T_2 - c_1 T_1).$$

Its expected length is

$$E[(c_2 - c_1)T_1] = (c_2 - c_1)(n-1)\theta.$$

(iii) Let $0 < a_1 < a_2$ be constants such that

$$P(a_1 < W < a_2) = \sqrt{1 - \alpha}$$

and let $0 < b_1 < b_2$ be constants such that

$$P\left(b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) = e^{-nb_1} - e^{-nb_2} = \sqrt{1-\alpha}.$$

Consider the region

$$C(X) = \left\{(\theta, a)\colon \frac{T_1(X)}{a_2} < \theta < \frac{T_1(X)}{a_1}, T_2(X) - b_2\theta < a < T_2(X) - b_1\theta\right\}.$$

By the result in Exercise 7(iii), $T_1(X)$ and $T_2(X)$ are independent. Hence

$$\begin{aligned}
P\big((a,\theta) \in C(X)\big) &= P\left(a_1 < \frac{T_1(X)}{\theta} < a_2,\ b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) \\
&= P\left(a_1 < \frac{T_1(X)}{\theta} < a_2\right) P\left(b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) \\
&= \sqrt{1-\alpha}\sqrt{1-\alpha} \\
&= 1 - \alpha.
\end{aligned}$$

Hence, $C(X)$ is a confidence region for (a, θ) with confidence coefficient $1 - \alpha$. ∎

Exercise 50 (#2.104). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathcal{R}$ is unknown. Let $X_{(j)}$ be the jth order statistic. Show that $(X_{(1)} + X_{(n)})/2$ is strongly consistent for θ and also consistent in mean squared error.

Solution. (i) For any $\epsilon > 0$,

$$\begin{aligned}
P\left(|X_{(1)} - (\theta - \tfrac{1}{2})| > \epsilon\right) &= P\left(X_{(1)} > \epsilon + (\theta - \tfrac{1}{2})\right) \\
&= \left[P\left(X_1 > \epsilon + \theta - \tfrac{1}{2}\right)\right]^n \\
&= (1-\epsilon)^n
\end{aligned}$$

and

$$\begin{aligned}
P\left(|X_{(n)} - (\theta + \tfrac{1}{2})| > \epsilon\right) &= P\left(X_{(n)} < (\theta + \tfrac{1}{2}) - \epsilon\right) \\
&= \left[P\left(X_1 < \theta + \tfrac{1}{2} - \epsilon\right)\right]^n \\
&= (1-\epsilon)^n.
\end{aligned}$$

Since $\sum_{i=1}^n (1-\epsilon)^n < \infty$, we conclude that $\lim_n X_{(1)} = \theta - \frac{1}{2}$ a.s. and $\lim_n X_{(n)} = \theta + \frac{1}{2}$ a.s. Hence $\lim_n (X_{(1)} + X_{(n)})/2 = \theta$ a.s.

(ii) A direct calculation shows that

$$E[X_{(n)} - (\theta + \tfrac{1}{2})] = n\int_0^1 x^n dx - 1 = -\frac{1}{n+1}$$

and

$$E[X_{(1)} - (\theta - \tfrac{1}{2})] = n\int_0^1 x(1-x)^{n-1}dx = \frac{1}{n+1}.$$

Hence $(X_{(1)} + X_{(n)})/2$ is unbiased for θ. Note that

$$\begin{aligned}
\mathrm{Var}(X_{(n)}) &= \mathrm{Var}\left(X_{(n)} - (\theta - \tfrac{1}{2})\right) \\
&= E[X_{(n)} - (\theta - \tfrac{1}{2})]^2 - [EX_{(n)} - (\theta - \tfrac{1}{2})]^2 \\
&= n\int_0^1 x^{n+1}dx - \left[\theta + \tfrac{1}{2} - \frac{1}{n+1} - (\theta - \tfrac{1}{2})\right]^2 \\
&= \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Similarly, $\lim_n \mathrm{Var}(X_{(1)}) = 0$. By the Cauchy-Schwarz inequality, $[\mathrm{Cov}(X_{(1)}, X_{(n)})]^2 \le \mathrm{Var}(X_{(1)})\mathrm{Var}(X_{(n)})$. Thus, $\lim_n \mathrm{Cov}(X_{(1)}, X_{(n)}) = 0$ and, consequently, $\lim_n E[(X_{(1)} + X_{(n)})/2 - \theta]^2 = \lim_n 4^{-1}[\mathrm{Var}(X_{(1)} + \mathrm{Var}(X_{(n)}) + 2\mathrm{Cov}(X_{(1)}, X_{(n)})] = 0$. ∎

Exercise 51 (#2.105). Let $(X_1, ..., X_n)$ be a random sample from a population with the Lebesgue density $f_\theta(x) = 2^{-1}(1 + \theta x)I_{(-1,1)}(x)$, where $\theta \in (-1, 1)$ is an unknown parameter. Find an estimator of θ that is strongly consistent and consistent in mean squared error.
Solution. By the strong law of large numbers, the sample mean $\bar{X}$ is strongly consistent for

$$EX_1 = \frac{1}{2}\int_{-1}^1 x(1+\theta x)dx = \frac{\theta}{2}\int_{-1}^1 x^2 dx = \frac{\theta}{3}.$$

Hence $3\bar{X}$ is a strongly consistent estimator of θ. Since $3\bar{X}$ is unbiased for θ and $\mathrm{Var}(3\bar{X}) = 9\mathrm{Var}(X_1)/n$, where

$$\mathrm{Var}(X_1) = EX_1^2 - (EX_1)^2 = \frac{1}{2}\int_{-1}^1 x^2(1+\theta x)dx - \frac{\theta^2}{9} = \frac{1}{2} - \frac{\theta^2}{9},$$

we conclude that $3\bar{X}$ is consistent in mean squared error. ∎

Exercise 52 (#2.106). Let $X_1, ..., X_n$ be a random sample. Suppose that T_n is an unbiased estimator of ϑ based on $X_1, ..., X_n$ such that for any n, $\mathrm{Var}(T_n) < \infty$ and $\mathrm{Var}(T_n) \le \mathrm{Var}(U_n)$ for any other unbiased estimator U_n of ϑ based on $X_1, ..., X_n$. Show that T_n is consistent in mean squared error.
Solution. Let $U_n = n^{-1}\sum_{i=1}^n T_1(X_i)$. Then U_n is unbiased for ϑ since $T_1(X_1)$ is unbiased for ϑ. By the assumption, $\mathrm{Var}(T_n) \le \mathrm{Var}(U_n)$. Hence $\lim_n \mathrm{Var}(T_n) = 0$ since $\lim_n \mathrm{Var}(U_n) = \lim_n \mathrm{Var}\big(T_1(X_1)\big)/n = 0$. ∎

Exercise 53 (#2.111). Let $X_1, ..., X_n$ be a random sample from P with unknown mean $\mu \in \mathcal{R}$ and variance $\sigma^2 > 0$, and let $g(\mu) = 0$ if $\mu \neq 0$ and $g(0) = 1$. Find a consistent estimator of $g(\mu)$.
Solution. Consider the estimator $T(X) = I_{(0,n^{-1/4})}(|\bar{X}|)$, where $\bar{X}$ is the sample mean. Note that $T = 0$ or 1. Hence, we only need to show that $\lim_n P(T = 1) = 1$ when $g(\mu) = 1$ (i.e., $\mu = 0$) and $\lim_n P(T = 1) = 0$ when $g(\mu) = 0$ (i.e., $\mu \neq 0$). If $\mu = 0$, by the central limit theorem, $\sqrt{n}\bar{X} \to_d N(0, \sigma^2)$ and, thus

$$\lim_n P(T(X) = 1) = \lim_n P(\sqrt{n}|\bar{X}| < n^{1/4}) = \lim_n \Phi(n^{1/4}) = 1,$$

where Φ is the cumulative distribution function of $N(0, 1)$. If $\mu \neq 0$, then by the law of large numbers, $|\bar{X}| \to_p |\mu| > 0$ and, hence, $n^{-1/4}/|\bar{X}| \to_p 0$. Then

$$\lim_n P(T(X) = 1) = \lim_n P(1 < n^{-1/4}/|\bar{X}|) = 0. \blacksquare$$

Exercise 54 (#2.115). Let $(X_1, ..., X_n)$ be a random sample of random variables from a population P with $EX_1^2 < \infty$ and $\bar{X}$ be the sample mean. Consider the estimation of $\mu = EX_1$.
(i) Let $T_n = \bar{X} + \xi_n/\sqrt{n}$, where ξ_n is a random variable satisfying $\xi_n = 0$ with probability $1 - n^{-1}$ and $\xi_n = n^{3/2}$ with probability n^{-1}. Show that the bias of T_n is not the same as the asymptotic bias of T_n for any P.
(ii) Let $T_n = \bar{X} + \eta_n/\sqrt{n}$, where η_n is a random variable that is independent of $X_1, ..., X_n$ and equals 0 with probability $1 - 2n^{-1}$ and $\pm\sqrt{n}$ with probability n^{-1}. Show that the asymptotic mean squared error of T_n, the asymptotic mean squared error of $\bar{X}$, and the mean squared error of $\bar{X}$ are the same, but the mean squared error of T_n is larger than the mean squared error of $\bar{X}$ for any P.
Note. The asymptotic bias and mean squared error are defined according to Definitions 2.11 and 2.12 in Shao (2003).
Solution. (i) Since $E(\xi_n) = n^{3/2}n^{-1} = n^{1/2}$, $E(T_n) = E(\bar{X}) + n^{-1/2}E(\xi_n)$ $= \mu + 1$. This means that the bias of T_n is 1. Since $\xi_n \to_p 0$ and $\bar{X} \to_p \mu$, $T_n \to_p \mu$. Thus, the asymptotic bias of T_n is 0.
(ii) Since $\eta_n \to_p 0$ and $\sqrt{n}(\bar{X} - \mu) \to_d N(0, \sigma^2)$, where $\sigma^2 = \mathrm{Var}(X_1)$, by Slutsky's theorem, $\sqrt{n}(T_n - \mu) = \sqrt{n}(\bar{X} - \mu) + \eta_n \to_d N(0, \sigma^2)$. Hence, the asymptotic mean squared error of T_n is the same as that of $\bar{X}$ and is equal to σ^2/n, which is the mean squared error of $\bar{X}$. Since $E(\eta_n) = 0$, $E(T_n) = E(\bar{X}) = \mu$ and the mean squared error of T_n is

$$\mathrm{Var}(T_n) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\eta_n/\sqrt{n}) = \frac{\sigma^2}{n} + \frac{2}{n} > \frac{\sigma^2}{n},$$

which is the mean squared error of $\bar{X}$. $\blacksquare$

Exercise 55 (#2.116(b)). Let $(X_1, ..., X_n)$ be a random sample of random variables with finite $\theta = EX_1$ and $\text{Var}(X_1) = \theta$, where $\theta > 0$ is unknown. Consider the estimation of $\sqrt{\theta}$. Let $T_{1n} = \sqrt{\bar{X}}$ and $T_{2n} = \bar{X}/S$, where $\bar{X}$ and S^2 are the sample mean and sample variance. Obtain the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.
Solution. Since $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta)$, by the δ-method with $g(t) = \sqrt{t}$ and $g'(t) = (2\sqrt{t})^{-1}$, $\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\theta}) \to_d N(0, \frac{1}{4})$. From Example 2.8 in Shao (2003),

$$\sqrt{n}(\bar{X} - \theta, S^2 - \theta) \to_d N_2(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \theta & \mu_3 \\ \mu_3 & \mu_4 - \theta^2 \end{pmatrix}$$

and $\mu_k = E(X_1 - \theta)^k$, $k = 3, 4$. By the δ-method with $g(x, y) = x/\sqrt{y}$, $\partial g/\partial x = 1/\sqrt{y}$ and $\partial g/\partial y = -x/(2y^{3/2})$, we obtain that

$$\sqrt{n}(T_{2n} - \sqrt{\theta}) \to_d N\left(0, \theta^{-1}[\theta^2 - \mu_3 + (\mu_4 - \theta^2)/4]\right).$$

Hence, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is $4\theta - 4\theta^{-1}\mu_3 + \theta^{-1}(\mu_4 - \theta^2)$. ∎

Exercise 56 (#2.118). Let $(X_1, ..., X_n)$ be a random sample from the $N(0, \sigma^2)$ distribution with an unknown $\sigma > 0$. Consider the estimation of σ. Find the asymptotic relative efficiency of $T_{1n} = \sqrt{\pi/2}\sum_{i=1}^n |X_i|/n$ with respect to $T_{2n} = (\sum_{i=1}^n X_i^2/n)^{1/2}$.
Solution. Since $E(\sqrt{\pi/2}|X_1|) = \sigma$ and $\text{Var}(\sqrt{\pi/2}|X_1|) = (\frac{\pi}{2} - 1)\sigma^2$, by the central limit theorem, we obtain that

$$\sqrt{n}(T_{1n} - \sigma) \to_d N\left(0, \left(\tfrac{\pi}{2} - 1\right)\sigma^2\right).$$

Since $EX_1^2 = \sigma^2$ and $\text{Var}(X_1) = 2\sigma^4$, $\sqrt{n}(n^{-1}\sum_{i=1}^n X_i^2 - \sigma^2) \to_d N(0, 2\sigma^4)$. By the δ-method with $g(t) = \sqrt{t}$ and $g'(t) = (2\sqrt{t})^{-1}$, we obtain that

$$\sqrt{n}(T_{2n} - \sigma) \to_d N\left(0, \tfrac{1}{2}\sigma^2\right).$$

Hence, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is equal to $\frac{1}{2}/(\frac{\pi}{2} - 1) = (\pi - 2)^{-1}$. ∎

Exercise 57 (#2.121). Let $X_1, ..., X_n$ be a random sample of random variables with $EX_i = \mu$, $\text{Var}(X_i) = 1$, and $EX_i^4 < \infty$. Let $T_{1n} = n^{-1}\sum_{i=1}^n X_i^2 - 1$ and $T_{2n} = \bar{X}^2 - n^{-1}$ be estimators of μ^2, where $\bar{X}$ is the sample mean.
(i) Find the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.
(ii) Show that the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is no larger than 1 if the distribution of $X_i - \mu$ is symmetric about 0 and

$\mu \neq 0$.
(iii) Find a distribution P for which the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is larger than 1.
Solution. (i) Since $EX_1^2 = \mathrm{Var}(X_1) + \mu^2 = 1 + \mu^2$, by applying the central limit theorem to $\{X_i^2\}$ we obtain that

$$\sqrt{n}(T_{1n} - \mu^2) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n X_i^2 - (1+\mu^2)\right] \to_d N(0,\gamma),$$

where $\gamma = \mathrm{Var}(X_1^2)$. Also, by the central limit theorem, $\sqrt{n}(\bar{X} - \mu) \to_d N(0,1)$. When $\mu \neq 0$, by the δ-method and Slutsky's theorem,

$$\sqrt{n}(T_{2n} - \mu^2) = \sqrt{n}(\bar{X}^2 - \mu^2) - \frac{1}{\sqrt{n}} \to_d N(0, 4\mu^2).$$

When $\mu = 0$, $\sqrt{n}\bar{X} \to_d N(0,1)$ and, thus,

$$n(T_{2n} - \mu^2) = n\bar{X}^2 - 1 = (\sqrt{n}\bar{X})^2 - 1 \to_d W - 1,$$

where W has the chi-square distribution χ_1^2. Note that $E(W-1) = 0$ and $\mathrm{Var}(W-1) = 2$. Therefore, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is equal to

$$e = \begin{cases} \frac{4\mu^2}{\mathrm{Var}(X_1^2)} & \mu \neq 0 \\ \frac{2}{n\mathrm{Var}(X_1^2)} & \mu = 0. \end{cases}$$

(ii) If the distribution of $X_1 - \mu$ is symmetric about 0, then $E(X_1-\mu)^3 = 0$ and, thus,

$$\begin{aligned} \mathrm{Var}(X_1^2) &= EX_1^4 - (EX_1^2)^2 \\ &= E[(X_1-\mu)+\mu]^4 - (1+\mu^2)^2 \\ &= E(X_1-\mu)^4 + 4\mu E(X_1-\mu)^3 + 6\mu^2 E(X_1-\mu)^2 \\ &\quad + 4\mu^3 E(X_1-\mu) + \mu^4 - (1+2\mu^2+\mu^4) \\ &= E(X_1-\mu)^4 + 4\mu^2 - 1 \\ &\geq 4\mu^2, \end{aligned}$$

where the inequality follows from the Jensen's inequality $E(X_1-\mu)^4 \geq [E(X_1-\mu)^2]^2 = 1$. Therefore, when $\mu \neq 0$, the asymptotic relative efficiency $e \leq 1$.
(iii) Let the common distribution of X_i be the distribution of $Y/\sqrt{p(1-p)}$, where Y is a binary random variable with $P(Y=1) = p$ and $P(Y=0) = 1-p$. Then $EX_i = \sqrt{p/(1-p)} = \mu$, $\mathrm{Var}(X_1) = 1$, and $EX_1^4 < \infty$. Note that

$$\mathrm{Var}(X_1^2) = \mathrm{Var}(Y^2)/[p^2(1-p)^2] = \mathrm{Var}(Y)/[p^2(1-p)^2] = [p(1-p)]^{-1}.$$

Then the asymptotic relative efficiency is $e = 4\mu^2/\mathrm{Var}(X_1^2) = 4p^2$, which is larger than 1 if $p \in (1/2, 1)$. ∎

Exercise 58 (#2.119). Let $(X_1, ..., X_n)$ be a random sample of random variables with unknown mean $\mu \in \mathcal{R}$, unknown variance $\sigma^2 > 0$, and $EX_1^4 < \infty$. Consider the estimation of μ^2 and the following three estimators: $T_{1n} = \bar{X}^2$, $T_{2n} = \bar{X}^2 - S^2/n$, $T_{3n} = \max\{0, T_{2n}\}$, where $\bar{X}$ and S^2 are the sample mean and variance.
(i) Show that the asymptotic mean squared errors of T_{jn}, $j = 1, 2, 3$, are the same when $\mu \neq 0$.
(ii) When $\mu = 0$, obtain the asymptotic relative efficiency of T_{2n} with respect to T_{1n} and the asymptotic relative efficiency of T_{3n} with respect to T_{2n}. Find out which estimator is asymptotically more efficient.
Solution. (i) By the central limit theorem and the δ-method,

$$\sqrt{n}(\bar{X}^2 - \mu^2) \to_d N(0, 4\mu^2\sigma^2).$$

By the law of large numbers, $S^2 \to_p \sigma^2$ and, hence, $S^2/\sqrt{n} \to_p 0$. By Slutsky's theorem,

$$\sqrt{n}(T_{2n} - \mu^2) = \sqrt{n}\bar{X}^2 - S^2/\sqrt{n} \to_d N(0, 4\mu^2\sigma^2).$$

This shows that, when $\mu \neq 0$, the asymptotic mean squared error of T_{2n} is the same as that of $T_{1n} = \bar{X}^2$. When $\mu \neq 0$, $\bar{X}^2 \to_p \mu^2 > 0$. Hence

$$\lim_n P(T_{2n} \neq T_{3n}) = \lim_n P(T_{2n} < 0) = \lim_n P(\bar{X}^2 < S^2/n) = 0,$$

since $S^2/n \to_p 0$. Therefore, the limiting distribution of $\sqrt{n}(T_{3n} - \mu^2)$ is the same as that of $\sqrt{n}(T_{2n} - \mu^2)$.
(ii) Assume $\mu = 0$. From $\sqrt{n}\bar{X} \to_d N(0, \sigma^2)$, we conclude that $n\bar{X}^2 \to_d \sigma^2 W$, where W has the chi-square distribution χ_1^2. Since $\mu = 0$, this shows that $n(T_{1n} - \mu^2) \to_d \sigma^2 W$ and, hence, the asymptotic mean squared error of T_{1n} is $\sigma^4 EW^2/n^2 = 3\sigma^4/n^2$. On the other hand, by Slutsky's theorem, $n(T_{2n} - \mu^2) = n\bar{X} - S^2 \to_p \sigma^2 W - \sigma^2$, since $S^2 \to_p \sigma^2$. Hence, the asymptotic mean squared error of T_{2n} is $\sigma^4 E(W-1)^2/n^2 = \sigma^4 \mathrm{Var}(W)/n^2 = 2\sigma^4/n^2$. The asymptotic relative efficiency of T_{2n} with respect to T_{1n} is $3/2$. Hence T_{2n} is asymptotically more efficient than T_{1n}. Note that

$$n(T_{3n} - \mu^2) = n\max\{0, T_{2n}\} = \max\{0, nT_{2n}\} \to_d \max\{0, \sigma^2(W-1)\},$$

since $\max\{0, t\}$ is a continuous function of t. Then the asymptotic mean squared error of T_{3n} is $\sigma^4 E(\max\{0, W-1\})^2/n^2$ and The asymptotic relative efficiency of T_{3n} with respect to T_{2n} is $E(W-1)^2/E(\max\{0, W-1\})^2$. Since

$$E(\max\{0, W-1\})^2 = E[(W-1)^2 I_{\{W>1\}}] < E(W-1)^2,$$

we conclude that T_{3n} is asymptotically more efficient than T_{jn}, $j = 1, 2$. ∎

Exercise 59. Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution $\theta^{-1}e^{-x/\theta}I_{(0,\infty)}(x)$, where $\theta \in (0, \infty)$. Consider the hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is a fixed constant. Let $T_c = I_{(c,\infty)}(\bar{X})$, where $\bar{X}$ is the sample mean.
(i) For any given level of significance $\alpha \in (0, 1)$, find a $c_{n,\alpha}$ such that the test $T_{c_{n,\alpha}}$ has size α and show that $T_{c_{n,\alpha}}$ is a consistent test, i.e., the power of $T_{c_{n,\alpha}}$ converges to 1 as $n \to \infty$ for any $\theta > \theta_0$.
(ii) Find a sequence $\{b_n\}$ such that the test T_{b_n} is consistent and the size of T_{b_n} converges to 0 as $n \to \infty$.
Solution. (i) Note that $\bar{X}/\theta$ has the gamma distribution with shape parameter n and scale parameter θ/n. Let $G_{n,\theta}$ denote the cumulative distribution function of this distribution and $c_{n,\alpha}$ be the constant satisfying $G_{n,\theta_0}(c_{n,\alpha}) = 1 - \alpha$. Then,

$$\sup_{\theta \leq \theta_0} P(T_{c_{n,\alpha}} = 1) = \sup_{\theta \leq \theta_0} [1 - G_{n,\theta}(c_{n,\alpha})] = 1 - G_{n,\theta_0}(c_{n,\alpha}) = \alpha,$$

i.e., the size of $T_{c_{n,\alpha}}$ is α.

Since the power of $T_{c_{n,\alpha}}$ is $P(T_{c_{n,\alpha}} = 1) = P(\bar{X} > c_{n,\alpha})$ for $\theta > \theta_0$ and, by the law of large numbers, $\bar{X} \to_p \theta$, the consistency of the test $T_{c_{n,\alpha}}$ follows if we can show that $\lim_n c_{n,\alpha} = \theta_0$. By the central limit theorem, $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta^2)$. Hence, $\sqrt{n}(\frac{\bar{X}}{\theta} - 1) \to_d N(0, 1)$. By Pólya's theorem (e.g., Proposition 1.16 in Shao, 2003),

$$\lim_n \sup_t \left| P\left(\sqrt{n} \left(\tfrac{\bar{X}}{\theta} - 1 \right) \leq t \right) - \Phi(t) \right| = 0,$$

where Φ is the cumulative distribution function of the standard normal distribution. When $\theta = \theta_0$,

$$\alpha = P(\bar{X} \geq c_{n,\alpha}) = P\left(\sqrt{n} \left(\tfrac{\bar{X}}{\theta_0} - 1 \right) \geq \sqrt{n} \left(\tfrac{c_{n,\alpha}}{\theta_0} - 1 \right) \right).$$

Hence

$$\lim_n \Phi\left(\sqrt{n} \left(\tfrac{c_{n,\alpha}}{\theta_0} - 1 \right) \right) = 1 - \alpha,$$

which implies $\lim_n \sqrt{n}(\frac{c_{n,\alpha}}{\theta_0} - 1) = \Phi^{-1}(1 - \alpha)$ and, thus, $\lim_n c_{n,\alpha} = \theta_0$.
(ii) Let $\{a_n\}$ be a sequence of positive numbers such that $\lim_n a_n = 0$ and $\lim_n \sqrt{n}a_n = \infty$. Let $\alpha_n = 1 - \Phi(\sqrt{n}a_n)$ and $b_n = c_{n,\alpha_n}$, where $c_{n,\alpha}$ is defined in the proof of part (i). From the proof of part (i), the size of T_{b_n} is α_n, which converges to 0 as $n \to \infty$ since $\lim_n \sqrt{n}a_n = \infty$.

Using the same argument as that in the proof of part (i), we can show that

$$\lim_n \left| 1 - \alpha_n - \Phi\left(\sqrt{n} \left(\tfrac{c_{n,\alpha_n}}{\theta_0} - 1 \right) \right) \right| = 0,$$

which implies that

$$\lim_n \frac{\sqrt{n}}{\Phi^{-1}(1-\alpha_n)}\left(\tfrac{c_{n,\alpha_n}}{\theta_0}-1\right)=1.$$

Since $1-\alpha_n=\Phi(\sqrt{n}a_n)$, this implies that $\lim_n c_{n,\alpha_n}=\theta_0$. Since $b_n = c_{n,\alpha_n}$, the test T_{b_n} is consistent. ∎

Exercise 60 (#2.130). Let (Y_i,Z_i), $i=1,...,n$, be a random sample from a bivariate normal distribution and let ρ be the correlation coefficient between Y_1 and Z_1. Construct a confidence interval for ρ that has asymptotic significance level $1-\alpha$, based on the sample correlation coefficient

$$\hat{\rho}=\frac{1}{(n-1)\sqrt{S_Y^2S_Z^2}}\sum_{i=1}^n(Y_i-\bar{Y})(Z_i-\bar{Z}),$$

where $\bar{Y}=n^{-1}\sum_{i=1}^n Y_i$, $\bar{Z}=n^{-1}\sum_{i=1}^n Z_i$, $S_Y^2=(n-1)^{-1}\sum_{i=1}^n(Y_i-\bar{Y})^2$, and $S_Z^2=(n-1)^{-1}\sum_{i=1}^n(Z_i-\bar{Z})^2$.

Solution. Assume first that $EY_1=EZ_1=0$ and $\mathrm{Var}(Y_1)=\mathrm{Var}(Z_1)=1$. From Exercise 9, $\sqrt{n}(\hat{\rho}-\rho)\to_d N(0,c^2)$ with

$$\begin{aligned}c^2 &= \rho^2[E(Y_1^4)+E(Z_1^4)+2E(Y_1^2Z_1^2)]/4\\ &\quad -\rho[E(Y_1^3Z_1)+E(Y_1Z_1^3)]+E(Y_1^2Z_1^2).\end{aligned}$$

We now derive the value of c^2. Under the normality assumption, $E(Y_1^4)=E(Z_1^4)=3$. Let $U=Y_1+Z_1$ and $V=Y_1-Z_1$. Then U is distributed as $N(0,2(1+\rho))$, V is distributed as $N(0,2(1-\rho))$ and U and V are independent, since $\mathrm{Cov}(U,V)=E(UV)=E(Y_1^2-Z_1^2)=0$. Note that $Y_1=(U+V)/2$ and $Z_1=(U-V)/2$. Then,

$$\begin{aligned}E(Y_1^2Z_1^2) &= \frac{E\left[(U+V)^2(U-V)^2\right]}{16}\\ &= \frac{E\left(U^4+V^4-2U^2V^2\right)}{16}\\ &= \frac{EU^4+EV^4-2EU^2EV^2}{16}\\ &= \frac{3[2(1+\rho)]^2+3[2(1-\rho)]^2-2[2(1+\rho)][2(1-\rho)]}{16}\\ &= \frac{3[(1+\rho)^2+(1-\rho)^2]-2(1-\rho^2)}{4}\\ &= \frac{3(2+2\rho^2)-2+2\rho^2}{4}\\ &= 1+2\rho^2\end{aligned}$$

and

$$\begin{aligned}
E(Y_1^3 Z_1) &= \frac{E[(U+V)^3(U-V)]}{16} \\
&= \frac{E[(U+V)^3 U] - E[(U+V)^3 V]}{16} \\
&= \frac{EU^4 + 3E(U^2V^2) - EV^4 - 3E(U^2V^2)}{16} \\
&= \frac{3[2(1+\rho)]^2 - 3[2(1-\rho)]^2}{16} \\
&= \frac{3(1+\rho)^2 - 3(1-\rho)^2}{4} \\
&= 3\rho.
\end{aligned}$$

By symmetry, $E(Y_1 Z_1^3) = 3\rho$. Using these results, we obtain that

$$\begin{aligned}
c^2 &= \rho^2[3+3+2(1+2\rho^2)]/4 - 2\rho(3\rho) + 1 + 2\rho^2 \\
&= \rho^2(2+\rho^2) - 6\rho^2 + 1 + 2\rho^2 \\
&= \rho^4 - 2\rho^2 + 1 \\
&= (1-\rho^2)^2.
\end{aligned}$$

In general, the distribution of $\hat{\rho}$ does not depend on the parameter vector $(EY_1, EZ_1, \mathrm{Var}(Y_1), \mathrm{Var}(Z_1))$, which can be shown by considering the transformation $(Y_i - EY_i)/\sqrt{\mathrm{Var}(Y_i)}$ and $(Z_i - EZ_i)/\sqrt{\mathrm{Var}(Z_i)}$. Hence,

$$\sqrt{n}(\hat{\rho} - \rho) \to_d N(0, (1-\rho)^2)$$

always holds, which implies that $\hat{\rho} \to_p \rho$. By Slutsky's theorem,

$$\frac{\sqrt{n}(\hat{\rho} - \rho)}{1 - \hat{\rho}^2} \to_d N(0,1).$$

Hence

$$\lim_n P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\rho} - \rho)}{1 - \hat{\rho}^2} \leq z_{\alpha/2}\right) = 1 - \alpha,$$

where z_a is the $(1-a)$th quantile of the standard normal distribution. Thus, a confidence interval for ρ that has asymptotic significance level $1-\alpha$ is

$$\left[\ \hat{\rho} - (1-\hat{\rho}^2) z_{\alpha/2}/\sqrt{n}, \hat{\rho} + (1-\hat{\rho}^2) z_{\alpha/2}/\sqrt{n}\ \right]. \ \blacksquare$$

http://www.springer.com/978-0-387-24970-4

Mathematical Statistics: Exercises and Solutions
Shao, J.
2005, XXVIII, 360 p., Softcover
ISBN: 978-0-387-24970-4