ANTONIO LIJOI

MATHEMATICAL STATISTICS: EXERCISES WITH SOLUTIONS

ADVANCED STATISTICAL METHODS
BOCCONI UNIVERSITY
A.Y. 2022–2023

The exercises are mostly taken from the following textbook:

Samaniego, F.J. *Stochastic Modeling and Mathematical Statistics*. Chapman & Hall, Boca Raton, FL.

The numbering is the same as in the book, with indication of the page where it can be found.

Exercise 6.7.1 (p. 291)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Pareto}(3, \theta)$.

(a) Find the value of c that makes $\hat{\theta}_1 = c \sum_{i=1}^n X_i$ and unbiased estimator of θ.

(b) Determine the probability distribution of $X_{(1)}$ and the constant d such that $\hat{\theta}_2 = d\, X_{(1)}$ is an unbiased estimator of θ.

(c) Which estimator would you prefer?

Solution.
Note that

$$f(x|\theta) = \frac{3\,\theta^3}{x^4}\,\mathbb{1}_{(\theta,+\infty)}(x), \qquad F_\theta(x) = \left\{1 - \left(\frac{\theta}{x}\right)^3\right\}\mathbb{1}_{[\theta,+\infty)}(x).$$

Before addresing (a), note that

$$\mathbb{E}X_1 = \int_\theta^{+\infty} x\,\frac{3\theta^3}{x^4}\,\mathrm{d}x = 3\,\theta^3 \int_\theta^{+\infty} \frac{1}{x^3}\,\mathrm{d}x = \frac{3}{2}\,\theta$$

$$\mathrm{Var}(X_1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \int_\theta^{+\infty} x^2\,\frac{3\theta^3}{x^4}\,\mathrm{d}x - \left(\frac{3}{2}\,\theta\right)^2 = 3\,\theta^3 \int_\theta^{+\infty} \frac{1}{x^2}\,\mathrm{d}x - \frac{9\theta^2}{4}$$
$$= 3\theta^2 - \frac{9\theta^2}{4} = \frac{3}{4}\,\theta^2$$

As for (a), note that

$$\mathbb{E}\,c\sum_{i=1}^n X_i = c\sum_{i=1}^n \mathbb{E}X_i = c\,\frac{3\theta}{2}\,n$$

and one deduces that $\mathbb{E}\hat{\theta}_1 = \theta$ if and only if $c = 2/(3n)$. An unbiased estimator of θ is, then,

$$\hat{\theta}_1 = \frac{2}{3n}\sum_{i=1}^n X_i$$

Finally,

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\Big(\frac{2}{3n}\sum_{i=1}^n X_i\Big) = \frac{4}{9n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{4}{9n^2}\,\frac{3\theta^2}{4}\,n = \frac{\theta^2}{3n}$$

As for (b), since

$$f_{X_{(1)}} = nf(x|\theta)\,\{1 - F_\theta(x)\}^{n-1} = \frac{3n\,\theta^{3n}}{x^{3n+1}}\,\mathbb{1}_{(\theta,+\infty)}(x)$$

and $X_{(1)} \sim \text{Pareto}(3n, \theta)$. Moreover

$$\mathbb{E}X_{(1)} = \int_\theta^{+\infty} x\,\frac{3n\,\theta^{3n}}{x^{3n+1}}\,\mathrm{d}x = 3n\,\theta^{3n}\int_\theta^{+\infty}\frac{1}{x^{3n}}\,\mathrm{d}x = \frac{3n}{3n-1}\,\theta.$$

Hence,

$$\mathbb{E}\hat{\theta}_2 = \mathbb{E}dX_{(1)} = d\,\mathbb{E}X_{(1)} = d\,\frac{3n}{3n-1}\,\theta$$

and $\mathbb{E}\hat{\theta}_2 = \theta$ if and only if $d = (3n-1)/(3n)$. Another unbiased estimator of θ is

$$\hat{\theta}_2 = \frac{3n-1}{3n}\,X_{(1)}$$

Note, further, that

$$\begin{aligned}
\text{Var}(\hat{\theta}_2) &= \text{Var}\Big(\frac{3n-1}{3n}\,X_{(1)}\Big) = \frac{(3n-1)^2}{9n^2}\,\text{Var}(X_{(1)}) \\
&= \frac{(3n-1)^2}{9n^2}\left\{\int_\theta^{+\infty} x^2\,\frac{3n\,\theta^{3n}}{x^{3n+1}}\,\mathrm{d}x - \frac{9n^2\,\theta^2}{(3n-1)^2}\right\} \\
&= \frac{(3n-1)^2}{9n^2}\left\{\frac{3n}{3n-2}\,\theta^2 - \frac{9n^2\,\theta^2}{(3n-1)^2}\right\} \\
&= \frac{(3n-1)^2}{9n^2}\,\frac{3n\theta^2}{(3n-1)^2(3n-2)} = \frac{\theta^2}{3n(3n-2)}
\end{aligned}$$

In view of these calculations one can now address (c), by noting that

$$\text{RE}(\hat{\theta}_1; \hat{\theta}_2) = \frac{\theta^2/[3n(3n-2)]}{\theta^2/(3n)} = \frac{1}{3n-2} < 1$$

and $\hat{\theta}_2$ is, thus, preferred to $\hat{\theta}_1$.

Exercise 6.7.9 (p. 292)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\,\cdot\,|\theta)$, with

$$f(x|\theta) = \frac{1+\theta x}{2}\,\mathbb{1}_{(-1,1)}(x)$$

Determine the constant c that makes $T = c\sum_{i=1}^n X_i$ an unbiased estimator of θ.

Solution.
Since $\mathbb{E}T = c\sum_{i=1}^n \mathbb{E}X_i = c\,n\,\mathbb{E}X_1$, we need to compute

$$\mathbb{E}X_1 = \int_{-1}^{1} x\,\frac{1+\theta x}{2}\,\mathrm{d}x = \frac{1}{2}\int_{-1}^{1}(x+\theta x^2)\,\mathrm{d}x = \frac{\theta}{3}.$$

Hence

$$\mathbb{E}T = cn\frac{\theta}{3}$$

and $\mathbb{E}T = \theta$ if and only if $c = 3/n$. We concluIde that an unbiased estimator of θ is

$$T = \frac{3}{n}\sum_{i=1}^{n} X_i.$$

Exercise 6.7.12 (p. 293)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$. If S is the sample standard deviation, identify the constant c that makes $\hat{\sigma} = cS$ unbiased for σ.

Solution.
Recall that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad\Longrightarrow\quad Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

In order to determine $\mathbb{E}\hat{\sigma}$, we need to identify the probability distribution of S. Since

$$f_Y(y) = \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\, y^{\frac{n-1}{2}-1}\, \mathrm{e}^{-\frac{y}{2}}\, \mathbb{1}_{(0,+\infty)}(y)$$

from $S^2 = \sigma^2 Y/(n-1)$, an application of the change of variable technique yields the following density function of S^2

$$f_{S^2}(w) = \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} w^{\frac{n-1}{2}-1}\, \mathrm{e}^{-\frac{n-1}{2\sigma^2}w}\, \mathbb{1}_{(0,+\infty)}(w).$$

Since $S = \sqrt{S^2}$, one can evaluate

$$\begin{aligned}
\mathbb{E}S &= \int_0^\infty \sqrt{w}\, f_{S^2}(w)\, \mathrm{d}w \\
&= \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} \int_0^\infty w^{\frac{n}{2}-1}\, \mathrm{e}^{-\frac{n-1}{2\sigma^2}w}\, \mathrm{d}w \\
&= \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n}{2}}} \\
&= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\sqrt{\frac{2}{n-1}}\,\sigma.
\end{aligned}$$

Hence $\mathbb{E}cS = \sigma$ if and only if $c = \Gamma(\frac{n-1}{2})\sqrt{n-1}/[\Gamma(\frac{n}{2})\sqrt{2}]$ and an unbiased estimator of σ is

$$\hat{\sigma} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\frac{n-1}{2}}\, S.$$

Exercise 6.7.20 (p. 294)
Let $X_i \overset{\text{ind}}{\sim} \mathrm{U}(0, i\theta)$, for $i = 1, \ldots, n$. Hence, the X_i's are independent, but not identically distributed.

(a) Show that

$$\hat{\theta}_1 = \frac{2}{n} \sum_{i=1}^{n} \frac{X_i}{i}$$

is an unbiased estimator of θ.

(b) Show that also

$$\hat{\theta}_2 = \left(1 + \frac{1}{n}\right) \max_{i=1,\ldots,n} \frac{X_i}{i}$$

is unbiased for θ.

(c) Which of these two estimators is preferred for estimating θ?

Solution.
Since $X_i \sim \mathrm{U}(0, i\theta)$, one has $\mathbb{E}X_i = (i\theta)/2$. Hence,

$$\mathbb{E}\hat{\theta}_1 = \frac{2}{n} \sum_{i=1}^{n} \frac{1}{i}\, \mathbb{E}X_i = \frac{2}{n} \sum_{i=1}^{n} \frac{1}{i}\, \frac{i\theta}{2} = \frac{1}{n}\, n\,\theta = \theta$$

and this shows (a).
From $X_i \sim \mathrm{U}(0, i\theta)$ one deduces

$$X_i^* = \frac{X_i}{i} \sim \mathrm{U}(0,\theta) \quad \& \quad \max_{1 \le i \le n} \frac{X_i}{i} = X_{(n)}^*$$

where $X_{(n)}^*$ is the n-th order statistic of a sample $X_1^*, \ldots, X_n^* \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$. This implies that

$$f_{X_{(n)}^*}(x) = \frac{n\, x^{n-1}}{\theta^n}\, \mathbb{1}_{(0,\theta)}(x)$$

and

$$\mathbb{E}X_{(n)}^* = \int_0^\theta x\, \frac{n\, x^{n-1}}{\theta^n}\, \mathrm{d}x = \frac{n}{n+1}\, \theta$$

To sum up

$$\mathbb{E}\hat{\theta}_2 = \mathbb{E}\left(1+\frac{1}{n}\right)\max_{1\le i\le n}\frac{X_i}{i} = \left(1+\frac{1}{n}\right)\mathbb{E}\max_{1\le i\le n}\frac{X_i}{i}$$
$$= \left(1+\frac{1}{n}\right)\mathbb{E}X^*_{(n)} = \frac{n+1}{n}\,\frac{n}{n+1}\,\theta = \theta$$

which proves (b).
Since $X_i \sim \mathrm{U}(0, i\theta)$, one has $\mathrm{Var}(X_i) = (i\theta)^2/12$ and

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\Big(\frac{2}{n}\sum_{i=1}^{n}\frac{X_i}{i}\Big) = \frac{4}{n^2}\sum_{i=1}^{n}\frac{1}{i^2}\,\frac{(i\theta)^2}{12} = \frac{\theta^2}{3n}.$$

Moreover, since

$$\mathrm{Var}(X^*_{(n)}) = \mathbb{E}(X^*_{(n)})^2 - (\mathbb{E}X^*_{(n)})^2 = \int_0^\theta x^2\,\frac{n}{\theta^n}\,x^{n-1}\,\mathrm{d}x - \frac{n^2}{(n+1)^2}\theta^2$$
$$= \frac{n}{(n+2)}\,\theta^2 - \frac{n^2}{(n+1)^2}\,\theta^2$$
$$= \frac{n}{(n+2)(n+1)^2}\,\theta^2$$

one has

$$\mathrm{Var}(\hat{\theta}_2) = \left(1+\frac{1}{n}\right)^2\mathrm{Var}\Big(\max_{1\le i\le n}\frac{X_i}{i}\Big) = \frac{(n+1)^2}{n^2}\,\mathrm{Var}(X^*_{(n)})$$
$$= \frac{(n+1)^2}{n^2}\,\frac{n}{(n+2)(n+1)^2}\,\theta^2 = \frac{\theta^2}{n(n+2)}.$$

The two estimators can be now compared in terms of

$$\mathrm{RE}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)} = \frac{\theta^2/[n(n+2)]}{\theta^2/(3n)} = \frac{3}{n+2} < 1$$

for any $n \ge 2$ and $\hat{\theta}_2$ is preferred over $\hat{\theta}_1$.

Exercise 6.1.3 (p. 260)
Let $X_1, X_2, X_3 \overset{\text{iid}}{\sim} \mathrm{U}(\theta - 1/2,\ \theta + 1/2)$. Show that the estimators

$$\hat{\theta}_1 = \bar{X} \qquad \hat{\theta}_2 = \frac{X_{(1)} + X_{(3)}}{2}$$

are both unbiased for estimating θ. Is one of the two preferable over the other?

Slightly more complicated exercise

Solution.
As for $\hat{\theta}_1$, since $\mathbb{E}X_1 = \theta$, one has

$$\mathbb{E}\,\hat{\theta}_1 = \theta \qquad \forall\theta \in \mathbb{R}$$

and $\hat{\theta}_1$ is unbiased for θ.
As for $\hat{\theta}_2$, it involves the order statistics of a sample of size $n = 3$. Notice that if $Y \sim \mathrm{U}(0,1)$, then

$$X = Y + \theta - \frac{1}{2} \sim \mathrm{U}(\theta - 1/2,\, \theta + 1/2).$$

Moreover,

$$X_{(1)} = Y_{(1)} + \theta - \frac{1}{2}, \quad X_{(2)} = Y_{(2)} + \theta - \frac{1}{2}, \quad X_{(3)} = Y_{(3)} + \theta - \frac{1}{2}$$

and we can then work with the order statistics from a sample of size $n = 3$ from a $\mathrm{U}(0,1)$ distribution, which is easier! Since

$$f(y) = \mathbb{1}_{(0,1)}(y) \qquad F(y) = y\,\mathbb{1}_{[0,1)}(y) + \mathbb{1}_{[1,+\infty)}(y),$$

from general formulae of order statistics one has

$$f_{Y_{(1)}}(y) = 3f(y)\,\{1 - F(y)\}^2 = 3(1-y)^2\,\mathbb{1}_{(0,1)}(y)$$
$$f_{Y_{(3)}}(y) = 3f(y)F^2(y) = 3y^2\,\mathbb{1}_{(0,1)}(y).$$

Hence,

$$Y_{(1)} \sim \mathrm{Beta}(1,3) \quad \Longrightarrow \quad \mathbb{E}Y_{(1)} = \frac{1}{4}$$
$$Y_{(3)} \sim \mathrm{Beta}(3,1) \quad \Longrightarrow \quad \mathbb{E}Y_{(1)} = \frac{3}{4}$$

and from this one deduces that

$$\mathbb{E}X_{(1)} = \mathbb{E}Y_{(1)} + \theta - \frac{1}{2} = \theta - \frac{1}{4}, \qquad \mathbb{E}X_{(3)} = \mathbb{E}Y_{(3)} + \theta - \frac{1}{2} = \theta + \frac{1}{4}.$$

Hence,

$$\mathbb{E}\,\hat{\theta}_2 = \frac{1}{2}\Big(\mathbb{E}X_{(1)} + \mathbb{E}X_{(3)}\Big) = \theta$$

and $\hat{\theta}_2$ is unbiased for θ.
As for the comparison between the two estimators, this is carried out in terms of precision. To this end, recall that

$$X \sim \mathrm{U}(a,b) \quad \Longrightarrow \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

$$X \sim \mathrm{Beta}(\alpha,\beta) \quad \Longrightarrow \quad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

In view of this, one has

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X_1)}{3} = \frac{1/12}{3} = \frac{1}{36}.$$

In order to determine the variance $\hat{\theta}_2$, from

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\Big(\frac{X_{(1)} + X_{(2)}}{2}\Big) = \mathrm{Var}\Big(\frac{Y_{(1)} + Y_{(3)}}{2} + \theta - \frac{1}{2}\Big) = \mathrm{Var}\Big(\frac{Y_{(1)} + Y_{(3)}}{2}\Big)$$
$$= \frac{1}{4}\,\Big(\mathrm{Var}(Y_{(1)}) + \mathrm{Var}(Y_{(3)}) + 2\,\mathrm{Cov}(Y_{(1)}, Y_{(3)})\Big)$$

it is apparent we need to identify the joint distribution of $(Y_{(1)}, Y_{(3)})$ in order to evaluate $\mathrm{Cov}(Y_{(1)}, Y_{(3)})$. A general formula for order statistics yields

$$f_{Y_{(1)},Y_{(3)}}(x,y) = \frac{3!}{1!} f(x)f(y)\{F(y) - F(x)\} = 6(y-x)\,\mathbb{1}_{(0,1)}(x)\,\mathbb{1}_{(x,1)}(y)$$

and from this we deduce

$$\begin{aligned}
\mathrm{Cov}(Y_{(1)}, Y_{(3)}) &= \mathbb{E}Y_{(1)}Y_{(3)} - (\mathbb{E}Y_{(1)})(\mathbb{E}Y_{(3)}) \\
&= \int_0^1 \int_x^1 xy\,6(y-x)\,\mathrm{d}y\,\mathrm{d}x - \frac{1}{4}\frac{3}{4} \\
&= 6\int_0^1 x\Big(\int_x^1 (y^2 - xy)\mathrm{d}y\Big)\,\mathrm{d}x - \frac{3}{16} \\
&= 6\int_0^1 x\left[\frac{y^3}{3} - x\frac{y^2}{2}\right]_{y=x}^{y=1} \mathrm{d}x - \frac{3}{16} \\
&= 6\int_0^1 \Big(\frac{x}{3} - \frac{x^2}{2} - \frac{x^4}{3} + \frac{x^4}{2}\Big)\,\mathrm{d}x - \frac{3}{16} \\
&= 6\left[\frac{x^2}{6} - \frac{x^3}{6} + \frac{x^5}{30}\right]_{x=0}^{x=1} - \frac{3}{16} = \frac{1}{5} - \frac{3}{16} = \frac{1}{80}.
\end{aligned}$$

Moreover, since $Y_{(1)} \sim \mathrm{Beta}(1,3)$ and $Y_{(3)} \sim \mathrm{Beta}(3,1)$, one has

$$\mathrm{Var}(Y_{(1)}) = \mathrm{Var}(Y_{(3)}) = \frac{3}{5\cdot 4^2} = \frac{3}{80}.$$

We are now in a position to evaluate $\mathrm{Var}(\hat{\theta}_2)$ and compare it to $\mathrm{Var}(\hat{\theta}_1)$ and since

$$\mathrm{Var}(\hat{\theta}_2) = \frac{1}{4}\left\{\frac{3}{80} + \frac{3}{80} + 2\,\frac{1}{80}\right\} = \frac{1}{40} < \frac{1}{36} = \mathrm{Var}(\hat{\theta}_1)$$

for any $\theta \in \mathbb{R}$, one concludes that $\hat{\theta}_2$ is preferred over $\hat{\theta}_1$.

Exercise 6.7.23 (p. 294)
Let $X \sim \mathrm{NB}(r,p)$, with $r \geq 1$ fixed and known and $p \in (0,1)$ unknown. Identify the Fisher information $\mathcal{I}_{X_1}(p)$ and the Cramér–Rao lower bound (CRLB) for the variance of unbiased estimators of p.

Solution.
Since

$$f(x|p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}\,\mathbb{1}_{\{r,r+1,\dots\}}(x)$$

satisfies the regularity conditions, and

$$\log f(x|p) = \log\binom{x-1}{r-1} + r\log p + (x-r)\log(1-p),$$

one has

$$\frac{\partial}{\partial p}\log f(x|p) = \frac{r}{p} - \frac{x-r}{1-p}$$
$$\frac{\partial^2}{\partial p^2}\log f(x|p) = -\frac{r}{p^2} - \frac{x-r}{(1-p)^2}.$$

Since $\mathbb{E}X_1 = r/p$, one has

$$\mathcal{I}_{X_1}(p) = -\mathbb{E}\frac{\partial^2}{\partial p^2}\log f(X_1|p) = \mathbb{E}\Big(\frac{r}{p^2} + \frac{X_1 - r}{(1-p)^2}\Big)$$
$$= \frac{r}{p^2} + \frac{\mathbb{E}X_1 - r}{(1-p)^2} = \frac{r}{p^2} + \frac{r}{p(1-p)} = \frac{r}{p^2(1-p)}.$$

Hence, the Cramér–Rao lower bound for the variance of unbiased estimators of p is

$$\frac{1}{\mathcal{I}_{X_1,\ldots,X_n}(p)} = \frac{1}{(nr)/[p^2(1-p)]} = \frac{p^2(1-p)}{rn}$$

Exercise (from the May 2018 exam)
Let $X_1, \ldots, X_n$ be an iid sample from a distribution whose density is

$$f(x|\beta) = \frac{1}{\beta^2}\, x^{-3}\, \mathrm{e}^{-\frac{1}{\beta x}}\, \mathbb{1}_{(0,+\infty)}(x)$$

(a) Identify a sufficient and complete statistic for β and motivate the answer.

(b) Determine the Cramér–Rao lower bound for unbiased estimators of β

(c) Propose a MVUE for β

Solution.
Since the density of the data is

$$f(x_1,\ldots,x_n|\beta) = \prod_{i=1}^n f(x_i|\beta) = \prod_{i=1}^n \frac{1}{\beta^2}\, x_i^{-3}\, \mathrm{e}^{-\frac{1}{\beta x_i}}\, \mathbb{1}_{(0,+\infty)}(x_i)$$
$$= \frac{\prod_{i=1}^n x_i^{-3}}{\beta^{2n}}\, \mathrm{e}^{-\frac{1}{\beta}\sum_{i=1}^n \frac{1}{x_i}} \prod_{i=1}^n \mathbb{1}_{(0,+\infty)}(x_i)$$
$$= \exp\Big\{-\frac{1}{\beta}\sum_{i=1}^n \frac{1}{x_i} - 3\sum_{i=1}^n \log(x_i) - 2n\log(\beta)\Big\} \prod_{i=1}^n \mathbb{1}_{(0,+\infty)}(x_i)$$

it is is easily seen that this is a one–parameter exponential family with $\Omega = \{x \in \mathbb{R} : \ f(x|\beta) > 0\} = (0,1)$, $A(\beta) = -1/\beta$, $C(x) = -3\log x$, $D(\beta) = -2\log\beta$ and $B(x) = 1/x$. Hence,

$$\sum_{i=1}^{n} B(X_i) = \sum_{i=1}^{n} \frac{1}{X_i}$$

is the sufficient and complete statistic. And this asswers (a).
As for (b), we need to determine the Fisher information and recall that

$$\begin{aligned}
\mathcal{I}_{X_1}(\beta) &= -\mathbb{E}\,\frac{\partial^2}{\partial\beta^2}\,\log f(X_1|\beta) \\
&= -\mathbb{E}\,\frac{\partial^2}{\partial\beta^2}\Big(-2\log\beta - 3\log X_1 - \frac{1}{\beta X_1}\Big) \\
&= -\,\mathbb{E}\left\{\frac{2}{\beta^2} - \frac{2}{\beta^3 X_1}\right\} \\
&= \frac{2}{\beta^2}\,\mathbb{E}\Big(\frac{1}{\beta X_1} - 1\Big) = \frac{2}{\beta^2}\Big\{\frac{1}{\beta}\,\mathbb{E}\frac{1}{X_1} - 1\Big\}
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\frac{1}{X_1} &= \frac{1}{\beta^2}\int_0^\infty \frac{1}{x^4}\,\mathrm{e}^{-\frac{1}{\beta x}}\,\mathrm{d}x = \frac{1}{\beta^2}\int_0^\infty y^2\,\mathrm{e}^{-\frac{y}{\beta}}\,\mathrm{d}y \\
&= 2\beta.
\end{aligned}$$

Hence

$$\mathcal{I}_{X_1}(\beta) = \frac{2}{\beta^2}\Big(\frac{1}{\beta}2\beta - 1\Big) = \frac{2}{\beta^2}$$

and

$$\mathrm{CRLB} = \frac{1}{n\,\mathcal{I}_{X_1}(\beta)} = \frac{\beta^2}{2n}$$

In order to identify a MVUE of β, from

$$\mathbb{E}\sum_{i=1}^{n}\frac{1}{X_i} = \sum_{i=1}^{n}\mathbb{E}\frac{1}{X_i} = 2n\beta.$$

we deduce that the estimator

$$T_n = \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{X_i}$$

is unbiased for estimating β, i.e. $\mathbb{E}T_n = \beta$. Furthermore, since it is a function the complete and sufficiente statistic, it is a MVUE of β.

Exercise (from the June 2018 exam)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ namely

$$f(x|\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2\sigma^2}}, \qquad \sigma > 0.$$

(a) Find an unbiased estimator for σ^2 based only on the first observation X_1, and determine its variance.

(b) Compute the Crámer–Rao lower bound for the class of unbiased estimators of σ^2.

(c) Identify a sufficient and complete statistic for σ^2 and establish whether the solution of point (a) is MVUE for σ^2 (motivate the answer). In case it is not, which estimator is MVUE for σ^2?

Solution.
Since $X_1 \sim \mathrm{N}(0, \sigma^2)$, then

$$\frac{X_1^2}{\sigma^2} \sim \chi^2_{(1)}.$$

Further recall that is $Y \sim \chi^2_{(q)}$, then $\mathbb{E}Y = q$ and $\mathrm{Var}(Y) = 2q$. This implies that $\mathbb{E}(X_1^2/\sigma^2) = 1$ and $\mathrm{Var}(X_1^2/\sigma^2) = 2$. Hence, if we set $T = T(X_1) = X_1^2$, we have $\mathbb{E}T = \mathbb{E}X_1^2 = \sigma^2$ and the estimator is unbiased. Moreover

$$\mathrm{Var}(T) = \mathrm{Var}(X_1^2) = 2\sigma^4$$

and this completes the answer to (a).
In order to address (b), we deterimne the Fisher information $\mathcal{I}_{X_1}(\sigma^2)$, after noting that

$$\log f(x; \sigma^2) = -\frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi) - \frac{x^2}{2\sigma^2}$$

which yields

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log f(x;\sigma^2) = \frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}.$$

Since we have already argued that $\mathbb{E}X_1^2 = \sigma^2$, we have

$$\begin{aligned}\mathcal{I}_{X_1}(\sigma^2) &= -\mathbb{E}\,\frac{\partial^2}{\partial(\sigma^2)^2}\log f(X_1;\sigma^2)\\ &= -\frac{1}{2\,\sigma^4} + \frac{\mathbb{E}X_1^2}{\sigma^6} = \frac{1}{2\sigma^4}\end{aligned}$$

Finally, the Cramér–Rao lower bound for the variance of unbiased estimators of σ^2 is

$$\mathrm{CRLB} = \frac{1}{n\mathcal{I}_{X_1}(\sigma^2)} = \frac{2\sigma^4}{n},$$

which answers (b).
It is immediate to note that

$$f(x|\sigma) = \mathrm{e}^{-\frac{1}{2\sigma^2}x^2 - \log(\sigma\sqrt{2\pi})}\,\mathbb{1}_{\mathbb{R}}(x)$$

and, hence, $\{f(\cdot|\sigma) : \ \sigma > 0\}$ is a one–parameter exponential family with $A(\sigma) = -1/(2\sigma^2)$, $B(x) = x^2$, $C(x) = 0$ and $D(\sigma) = -\log(\sigma\sqrt{2\pi})$. Hence

$$\sum_{i=1}^{n} B(X_i) = \sum_{i=1}^{n} X_i^2$$

is a sufficient and complete statistic for the model. The estimator $T = X_1$ discussed in (a) is not a MVUE of σ^2 unless $n = 1$. Indeed, for $n > 1$, T is not a function of the sufficient and complete statistics. On the other hand, the estimator

$$T' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

is unbiased, i.e. $\mathbb{E}T' = \mathbb{E}(\sum_{i=1}^{n} X_i^2/n) = \sigma^2$ and it is a function of the sufficient and complete statistic. Hence, T' is a MVUE of σ^4. This conclusion could have been achieved, in this case, by noting that

$$\mathrm{Var}(T') = \frac{\mathrm{Var}(X_1^2)}{n} = \frac{2\sigma^4}{n} = \mathrm{CRLB}$$

so that he variance of T' attains the Cramér–Rao lower bound for the variance of unbiased estimators of σ^2 that has been identified in (b). This completes the solution of point (c).

Exercise

Let $X_1, \ldots, X_n$ be an iid sample from from a distribution whose probability density function is

$$f(x|\alpha) = \frac{\alpha\, 3^{\alpha}}{x^{\alpha+1}}\, \mathbb{1}_{(3,+\infty)}(x) \qquad \alpha > 0$$

(a) What is the sufficient statistic for this family of distributions?

(b) Determine the Cramér–Rao lower bound for the variance of unbiased estimators of the parameter α.

Solution.

From

$$\begin{aligned} f(x_1,\ldots,x_n|\alpha) &= \prod_{i=1}^{n} f(x_i|\alpha) = \prod_{i=1}^{n} \frac{\alpha\, 3^{\alpha}}{x^{\alpha+1}}\, \mathbb{1}_{(3,+\infty)}(x_i) \\ &= \frac{\alpha^n\, 3^{n\alpha}}{(\prod_{i=1}^{n} x_i)^{\alpha+1}} \prod_{i=1}^{n} \mathbb{1}_{(3,+\infty)}(x_i) \\ &= \nu(t;\alpha)\, w(x_1,\ldots,x_n) \end{aligned}$$

where $t = \prod_{i=1}^{n} x_i$,

$$v(t;\alpha) = \frac{\alpha^n\, 3^{n\alpha}}{t^{\alpha+1}}, \qquad w(x_1,\ldots,x_n) = \prod_{i=1}^{n} \mathbb{1}_{(3,+\infty)}(x_i).$$

From the factorization theorem we can conclude that $T = \prod_{i=1}^{n} X_i$ is sufficient.
As for the Cramér–Rao lower bound in (b),

$$\log f(x|\alpha) = \log\alpha + \alpha\log 3 - (\alpha+1)\log x.$$

From this one has

$$\frac{\partial}{\partial\alpha}\log f(x|\alpha) = \frac{1}{\alpha} + \log 3 - \log x$$
$$\frac{\partial^2}{\partial\alpha^2}\log f(x|\alpha) = -\frac{1}{\alpha^2}$$

and, then, the Fisher information is

$$\mathcal{I}_{X_1}(\alpha) = -\mathbb{E}\,\frac{\partial^2}{\partial\alpha^2}\log f(X_1|\alpha) = \frac{1}{\alpha^2}.$$

To complete the answer to (b), it is now enough to note that

$$\text{CRLB} = \frac{1}{n\mathcal{I}_{X_1}(\alpha)} = \frac{\alpha^2}{n}.$$

Exercise 6.7.29 (p.295)
Let $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \text{Gamma}(\alpha_0, 1/\beta)$ where α_0 is a known positive number, while $\beta > 0$ is unknown. The density function is

$$f(x|\beta) = \frac{(1/\beta)^{\alpha_0}}{\Gamma(\alpha_0)}\, x^{\alpha_0-1}\, \mathrm{e}^{-x/\beta}\, \mathbb{1}_{(0,+\infty)}(x).$$

(a) Show that
$$\hat\beta_0 = \frac{1}{\alpha_0 n}\sum_{j=1}^{n} X_j$$
is an unbiased estimator of β.

(b) Determine the CRLB for the variance of unbiased estimators of β. Is $\hat\beta$ MVUE of β?

Solution.

First note that

$$\begin{aligned}
\mathbb{E}X_1 &= \int_{-\infty}^{+\infty} x\, f(x|\beta)\, \mathrm{d}x = \int_0^\infty x\, \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)}\, x^{\alpha_0 - 1}\, \mathrm{e}^{-x/\beta}\, \mathrm{d}x \\
&= \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)} \int_0^\infty x^{\alpha_0+1-1}\, \mathrm{e}^{-x/\beta}\, \mathrm{d}x = \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)}\, \frac{\Gamma(\alpha_0+1)}{(1/\beta)^{\alpha_0+1}} \\
&= \alpha_0 \beta
\end{aligned}$$

since $\Gamma(\alpha_0+1) = \alpha_0\, \Gamma(\alpha_0)$. As for (a), one can use this result to obtain

$$\mathbb{E}\,\hat{\beta} = \frac{1}{\alpha_0 n} \sum_{i=1}^n \mathbb{E}X_i = \frac{1}{\alpha_0 n} n \alpha_0\, \beta = \beta,$$

which shows that $\hat{\beta}$ is unbiased for *beta*.
As for (b), we first determine the Fisher information

$$\mathcal{I}_{X_1}(\beta) = -\mathbb{E} \frac{\partial^2}{\partial \beta^2} \log f(X_1|\beta)$$

Since

$$\log f(x|\beta) = -\log \Gamma(\alpha_0) - \alpha_0 \log \beta + (\alpha_0 - 1) \log x - \frac{x}{\beta}$$

one has

$$\begin{aligned}
\frac{\partial}{\partial \beta} \log f(x|\beta) &= -\frac{\alpha_0}{\beta} + \frac{x}{\beta^2} \\
\frac{\partial^2}{\partial \beta^2} \log f(x|\beta) &= \frac{\alpha_0}{\beta^2} - \frac{2x}{\beta^3}.
\end{aligned}$$

Hence

$$\mathcal{I}_{X_1}(\beta) - \mathbb{E}\Big(\frac{\alpha_0}{\beta^2} - \frac{2X_1}{\beta^3}\Big) = -\frac{\alpha_0}{\beta^2} + \frac{2\,\mathbb{E}X_1}{\beta^3} = \frac{\alpha_0}{\beta^2}$$

and the Cramér–Rao lower bound is

$$\frac{1}{\mathcal{I}_{X_1,\dots,X_n}(\beta)} = \frac{1}{n\, \mathcal{I}_{X_1}(\beta)} = \frac{\beta^2}{n\, \alpha_0}.$$

In order to establish whether $\mathrm{Var}(\hat{\beta})$ attains the CRLB above, we need to determine

$$\begin{aligned}
\mathrm{Var}(X_1) &= \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)} \int_0^\infty x^2\, x^{\alpha_0-1}\, \mathrm{e}^{-x/\beta}\, \mathrm{d}x - (\alpha_0\beta)^2 \\
&= \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)} \int_0^\infty x^{\alpha_0+2-1}\, \mathrm{e}^{-x/\beta}\, \mathrm{d}x - \alpha_0^2 \beta^2 \\
&= \frac{1}{\beta^{\alpha_0}\, \Gamma(\alpha_0)}\, \beta^{\alpha_0+2}\, \Gamma(\alpha_0+2) - \alpha_0^2\, \beta^2 = \alpha_0(\alpha_0+1)\beta^2 - \alpha_0^2\, \beta^2 \\
&= \alpha_0 \beta^2
\end{aligned}$$

where we've used the fact that $\Gamma(\alpha_0+2)=\alpha_0(\alpha_0+1)\Gamma(\alpha_0)$. We can now determine

$$\operatorname{Var}(\hat{\beta})=\frac{1}{\alpha_0^2\,n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i)=\frac{1}{\alpha_0^2\,n^2}\,n\alpha_0\beta^2=\frac{\beta^2}{n\,\alpha_0}=\text{CRLB}.$$

Hence, $\hat{\beta}$ is a MVUE of β.

Exercise 6.7.44 (p. 296)
Suppose $X_1\sim \mathrm{N}(0,\sigma^2)$. Show that $|X_1|$ is a sufficient statistic for σ^2.

Solution.
With $T=|X_1|$, from $\mathbb{P}[X_1<0]=\mathbb{P}[X_1>0]=1/2$, one has

$$\mathbb{P}[X=x|T=t]=\frac{1}{2}\,\mathbb{1}_{\{t\}}(x)+\frac{1}{2}\,\mathbb{1}_{\{-t\}}(x)$$

and such a probability distribution does not depend on σ^2. Hence, $T=|X_1|$ is sufficient.

Exercise 6.7.45 (p. 296)
Suppose $X_i\overset{\text{ind}}{\sim}\mathrm{E}(1/(i\theta))$, for $i=1,\dots,n$. Show that $\sum_{i=1}^n X_i/i$ is a sufficient statistic for θ.

Solution.
Letting $f_i(\,\cdot\,|\theta)$ denote the density function of X_i, i.e.

$$f_i(x|\theta)=\frac{1}{i\theta}\,\mathrm{e}^{-\frac{x}{i\theta}}\,\mathbb{1}_{(0,+\infty)}(x).$$

We can resort to the factorization theorem, since

$$\begin{aligned}f(x_1,\dots,x_n|\theta)&=\prod_{i=1}^{n}f_i(x_i|\theta)=\prod_{i=1}^{n}\frac{1}{i\theta}\,\mathrm{e}^{-\frac{x}{i\theta}}\,\mathbb{1}_{(0,+\infty)}(x_i)\\&=\frac{1}{\theta^n\,n!}\,\mathrm{e}^{-\frac{1}{\theta}\sum_{i=1}^n\frac{x_i}{i}}\,\mathbb{1}_{(0,+\infty)}(x_{(1)})\\&=\nu(t,\theta)\,w(x_1,\dots,x_n)\end{aligned}$$

where $x_{(1)}=\min_{1\le i\le n}x_i$,

$$t=\sum_{i=1}^{n}\frac{x_i}{i},\quad \nu(t,\theta)=\frac{1}{\theta^n}\,\mathrm{e}^{-\frac{t}{\theta}},\quad w(x_1,\dots,x_n)=\frac{1}{n!}\,\mathbb{1}_{(0,+\infty)}(x_{(1)})$$

and $T=\sum_{i=1}^n X_i/i$ is sufficient.

Exercise 6.7.47 (p. 296)
Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Beta}(\alpha, \beta)$. Identify a two–dimensional sufficient statistic for (α, β).

Solution.
Since

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}\, \mathbb{1}_{(0,1)}(x),$$

the joint density of the data is

$$\begin{aligned} f(x_1,\dots,x_n|\alpha,\beta) &= \prod_{j=1}^n f(x_j|\alpha,\beta) = \prod_{j=1}^n \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x_j^{\alpha-1}(1-x_j)^{\beta-1}\, \mathbb{1}_{(0,1)}(x_j) \\ &= \frac{\Gamma^n(\alpha+\beta)}{\Gamma^n(\alpha)\,\Gamma^n(\beta)}\, \mathrm{e}^{\alpha\sum_{j=1}^n \log x_j + \beta \sum_{j=1}^n \log(1-x_j)} \prod_{j=1}^n \frac{1}{x_j(1-x_j)}\, \mathbb{1}_{(0,1)}(x_j) \\ &= v(t_1,t_2;\alpha,\beta)\, w(x_1,\dots,x_n) \end{aligned}$$

where $t_1 = \sum_{j=1}^n \log x_j$, $t_2 = \sum_{j=1}^n \log(1-x_j)$ and

$$\begin{aligned} v(t_1,t_2;\alpha,\beta) &= \frac{\Gamma^n(\alpha+\beta)}{\Gamma^n(\alpha)\,\Gamma^n(\beta)}\, \mathrm{e}^{\alpha t_1 + \beta t_2} \\ w(x_1,\dots,x_n) &= \prod_{j=1}^n \frac{1}{x_j(1-x_j)}\, \mathbb{1}_{(0,1)}(x_j). \end{aligned}$$

Hence, the factorization theorem implies that

$$(T_1, T_2) = \Big(\sum_{j=1}^n \log X_j,\ \sum_{j=1}^n \log(1-X_j) \Big)$$

is sufficient for (α, β).

Exercise 6.7.49 (p. 296)
Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \theta^2)$, namely

$$f(x|\theta) = \frac{1}{|\theta|\,\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2\theta^2}(x-\theta)^2} \qquad x \in \mathbb{R}, \quad \theta \in \mathbb{R}.$$

Identify a sufficient statistic for θ.

Solution.
We will use the factorization theorem. To this end, we determine the

joint density of $(X_1, \ldots, X_n)$

$$\begin{aligned} f(x_1, \ldots, x_n|\theta) &= \prod_{j=1}^{n} f(x_j|\theta) = \prod_{j=1}^{n} \frac{1}{|\theta|\,\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2\theta^2}(x_j-\theta)^2} \\ &= \frac{1}{|\theta|^n\,(2\pi)^{n/2}}\,\mathrm{e}^{-\frac{1}{2\theta^2}\,\sum_{j=1}^n (x_j-\theta)^2} \\ &= \frac{1}{|\theta|^n}\,\mathrm{e}^{-\frac{\sum_{j=1}^n x_j^2}{2\theta^2}+\frac{\sum_{j=1}^n x_j}{\theta}}\,\frac{1}{(2\pi)^{n/2}}\,\mathrm{e}^{-\frac{n}{2}} \\ &= \nu(t_1, t_2; \theta)\,w(x_1, \ldots, x_n), \end{aligned}$$

where $t_1 = \sum_{j=1}^n x_j^2$, $t_2 = \sum_{j=1}^n x_j$ and

$$\begin{aligned} \nu(t_1, t_2; \theta) &= \frac{1}{|\theta|^n}\,\mathrm{e}^{-\frac{t_1}{2\theta^2}+\frac{t_2}{\theta}} \\ w(x_1, \ldots, x_n) &= \frac{1}{(2\pi)^{n/2}}\,\mathrm{e}^{-\frac{n}{2}} \end{aligned}$$

We can, then, conclude that

$$(T_1, T_2) = \Big(\sum_{j=1}^{n} X_j^2, \sum_{j=1}^{n} X_j\Big)$$

is sufficient for θ.

Exercise 6.7.55 (p. 296)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu_0, \sigma^2)$, where $\mu_0 \in \mathbb{R}$ is known and $\sigma > 0$ is unknown. Identify a sufficient statistic for $\theta = \sigma$.

Solution.
In this case

$$f(x|\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu_0)^2}$$

and, then,

$$\begin{aligned} f(x_1, \ldots, x_n|\sigma) &= \prod_{j=1}^{n} f(x_j|\sigma) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x_j-\mu_0)^2} \\ &= \frac{1}{\sigma^n\,(2\pi)^{n/2}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}\,\sum_{j=1}^n (x_j-\mu_0)^2} \\ &= \nu(t, \sigma)\,w(x_1, \ldots, x_n) \end{aligned}$$

where $t = \sum_{j=1}^n (x_j - \mu_0)^2$,

$$\nu(t, \sigma) = \frac{1}{\sigma^n}\,\mathrm{e}^{-\frac{t}{2\sigma^2}}, \quad w(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}}.$$

By the factorization theorem, $T = \sum_{j=1}^{n}(X_j - \mu_0)^2$ is sufficient for σ.

Exercise 6.7.59 (p. 298)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Po}(\lambda)$. Recall that $T = \sum_{i=1}^{n} X_i$ is sufficient and complete.

(a) Show that there doesn't exist any unbiased estimator of $\theta = 1/\lambda$ which is a function of T.

(b) Show that $\hat{\theta} = n/(\sum_{i=1}^{n} X_i + 1)$ is asymptotically unbiased, namely
$$\lim_{n\to\infty} \mathbb{E}\hat{\theta} = \theta.$$

Solution.
Recall that $T \sim \text{Po}(n\lambda)$ and suppose there exists a function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}g(T) = \theta = 1/\lambda$. If so, then
$$\frac{1}{\lambda} = \mathbb{E}g(T) = \sum_{t=0}^{\infty} g(t)\, \frac{(n\lambda)^t\, \mathrm{e}^{-n\lambda}}{t!}$$
and this implies that for any $\lambda > 0$ one must have
$$1 = \lambda \sum_{t=0}^{\infty} \frac{g(t)\, n^t\, \mathrm{e}^{-n\lambda}}{t!}\, \lambda^t = \sum_{t=0}^{\infty} \frac{g(t)\, n^t\, \mathrm{e}^{-n\lambda}}{t!}\, \lambda^{t+1}.$$
the right–hand side above is a polynomial in λ and it cannot equal 1 for any choice of λ and of the function g. Hence, there cannot exist and unbiased estimator of $\theta = 1/\lambda$. One also says that θ is *not extimable*. This shows (a).
As for (b),
$$\mathbb{E}\,\hat{\theta} = \mathbb{E}\frac{n}{1 + \sum_{i=1}^{n} X_i} = \sum_{t=0}^{\infty} \frac{n}{(t+1)}\, \frac{(n\lambda)^t\, \mathrm{e}^{-n\lambda}}{t!}$$
[use the fact that (t+1) t!=(t+1)!]
$$= \sum_{t=0}^{\infty} \frac{n^{t+1}\, \lambda^t\, \mathrm{e}^{-n\lambda}}{(t+1)!}$$
[by the change of variable $t + 1 = x$]
$$= \mathrm{e}^{-n\lambda} \sum_{x=1}^{\infty} \frac{n^x\, \lambda^{x-1}}{x!} = \frac{\mathrm{e}^{-n\lambda}}{\lambda} \sum_{x=1}^{\infty} \frac{n^x\, \lambda^x}{x!}$$
$$= \frac{\mathrm{e}^{-n\lambda}}{\lambda} \left\{ \sum_{x=0}^{\infty} \frac{n^x\, \lambda^x}{x!} - 1 \right\} = \frac{\mathrm{e}^{-n\lambda}}{\lambda} \left\{ \mathrm{e}^{n\lambda} - 1 \right\}$$
$$= \frac{1}{\lambda} - \frac{\mathrm{e}^{-n\lambda}}{\lambda} \neq \lambda$$

and, as expected from (a), $\hat{\theta}$ is biased for estimating θ. Nonetheless

$$\lim_{n\to\infty} \mathbb{E}\,\hat{\theta} = \lim_{n\to\infty} \left\{ \frac{1}{\lambda} - \frac{\mathrm{e}^{-n\lambda}}{\lambda} \right\} = \frac{1}{\lambda} = \theta$$

and, hence, $\hat{\theta}$ is asymptotically unbiased.

> **Exercise 6.7.61** (p. 298)
> Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(a, b)$. Identify a sufficient statistic (T_1, T_2) for (a, b). Obtain unbiased estimators of $\mathbb{E}X_1 = (a + b)/2$ and of the range $(b - a)$, based on (T_1, T_2).

Solution.
With $\theta = (a, b)$, for $a < b$, one has

$$f(x|\theta) = \frac{1}{b-a}\,\mathbb{1}_{(a,b)}(x).$$

Hence, the joint density of the data is

$$f(x_1, \ldots, x_n|\theta) = \prod_{j=1}^{n} f(x_j|\theta) = \prod_{j=1}^{n} \frac{1}{b-a}\,\mathbb{1}_{(a,b)}(x_j) = \frac{1}{(b-a)^n} \prod_{j=1}^{n} \mathbb{1}_{(a,b)}(x_j).$$

Notice, now, that all the observations $x_1, \ldots, x_n$ are in (a, b) if and only if the minimum is above a, i.e. $x_{(1)} > a$, and the maximum is below b, i.e. $x_{(n)} < b$, where we adopt the usual notation

$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad x_{(n)} = \max\{x_1, \ldots, x_n\}.$$

This implies

$$\prod_{j=1}^{n} \mathbb{1}_{(a,b)}(x_j) = \mathbb{1}_{(a,+\infty)}(x_{(1)})\,\mathbb{1}_{(-\infty,b)}(x_{(n)})$$

and we can ,then, rewrite

$$\begin{aligned} f(x_1, \ldots, x_n|\theta) &= \frac{1}{(b-a)^n}\,\mathbb{1}_{(a,+\infty)}(x_{(1)})\,\mathbb{1}_{(-\infty,b)}(x_{(n)}) \\ &= v(t_1, t_2; a, b)\,w(x_1, \ldots, x_n) \end{aligned}$$

where $t_1 = x_{(1)}$, $t_2 = x_{(n)}$ and

$$\begin{aligned} v(t_1, t_2; a, b) &= \frac{1}{(b-a)^n}\,\mathbb{1}_{(a,+\infty)}(x_{(1)})\,\mathbb{1}_{(-\infty,b)}(x_{(n)}) \\ w(x_1, \ldots, x_n) &= 1. \end{aligned}$$

By the factorization theorem one has that $(T_1, T_2) = (X_{(1)}, X_{(n)})$ is sufficient for (a, b).
Recall that

$$F_\theta(x) = \frac{x-a}{b-a}\,\mathbb{1}_{[a,b)}(x) + \mathbb{1}_{[b,+\infty)}(x)$$

and a simple application of general formulae for order statistics yields

$$
\begin{aligned}
f_{X_{(1)}}(x|\theta) &= n\, f(x|\theta)\, \{1 - F_\theta(x)\}^{n-1} = \frac{n}{b-a}\Big(1 - \frac{x-a}{b-a}\Big)^{n-1} \mathbb{1}_{(a,b)}(x) \\
&= \frac{n}{(b-a)^n}\, (b-x)^{n-1}\, \mathbb{1}_{(a,b)}(x) \\
f_{X_{(n)}}(x|\theta) &= n\, f(x|\theta)\, F_\theta^{n-1}(x) = \frac{n}{b-a}\, \Big(\frac{x-a}{b-a}\Big)^{n-1} \mathbb{1}_{(a,b)}(x) \\
&= \frac{n}{(b-a)^n}\, (x-a)^{n-1}\, \mathbb{1}_{(a,b)}(x).
\end{aligned}
$$

In order to identify unbiased estimators of the expectation $\mathbb{E}X_1 = (a+b)/2$ and of the range $b-a$, we first evaluate

$$
\mathbb{E}X_{(1)} = \int_{-\infty}^{+\infty} x\, f_{X_{(1)}}(x|\theta)\, \mathrm{d}x = \int_a^b x\, \frac{n}{(b-a)^n} (b-x)^{n-1}\, \mathrm{d}x
$$

[apply the change of variable $w = (b-x)/(b-a)$]

$$
\begin{aligned}
&= \frac{n}{(b-a)^n} \int_0^1 (b - w(b-a))\, w^{n-1} (b-a)^{n-1}\, (b-a)\, \mathrm{d}w \\
&= n\, b \int_0^1 w^{n-1}\, \mathrm{d}w - n\,(b-a) \int_0^1 w^n\, \mathrm{d}w \\
&= b - (b-a)\, \frac{n}{n+1} = \frac{b+na}{n+1}
\end{aligned}
$$

$$
\mathbb{E}X_{(n)} = \int_{-\infty}^{+\infty} x\, f_{X_{(1)}}(x|\theta)\, \mathrm{d}x = \int_a^b x\, \frac{n}{(b-a)^n} (x-a)^{n-1}\, \mathrm{d}x
$$

[apply the change of variable $w = (x-a)/(b-a)$]

$$
\begin{aligned}
&= \frac{n}{(b-a)^n} \int_0^1 (a + w(b-a))\, w^{n-1}\, (b-a)^{n-1}\, (b-a)\, \mathrm{d}w \\
&= n\, a \int_0^1 w^{n-1}\, \mathrm{d}w + n\,(b-a) \int_0^1 w^n\, \mathrm{d}w \\
&= a + \frac{n}{n+1}\, (b-a) = \frac{a+n\,b}{n+1}.
\end{aligned}
$$

A natural guess of a possible unbiased estimator of $(a+b)/2$, as a function of (T, T_2), is $(X_{(1)} + X_{(n)})/2$. Note that

$$
\begin{aligned}
\mathbb{E}\, \frac{X_{(1)} + X_{(n)}}{2} &= \frac{1}{2}\, (\mathbb{E}X_{(1)} + \mathbb{E}X_{(n)}) = \frac{1}{2}\Big(\frac{b+n\,a}{n+1} + \frac{a+n\,b}{n+1}\Big) \\
&= \frac{a+b}{2}
\end{aligned}
$$

and, hence, $(X_{(1)} + X_{(n)})/2$ is unbiased for estimating $\mathbb{E}X_1 = (a+b)/2$.

As for the range, $b-a$, it is natural to start with the sample range $X_{(n)} - X_{(1)}$ and note that

$$\mathbb{E}(X_{(n)} - X_{(1)}) = \mathbb{E}X_{(n)} - \mathbb{E}X_{(1)} = \frac{b+n\,a}{n+1} - \frac{a+n\,b}{n+1} = \frac{n-1}{n+1}\,(b-a).$$

Though $X_{(n)} - X_{(1)}$ is biased for estimating $b-a$, one may take a slight modification, $(n+1)(X_{(n)} - X_{(1)})/(n-1)$ and

$$\begin{aligned}\mathbb{E}\,\frac{n+1}{n-1}(X_{(n)} - X_{(1)}) &= \frac{n+1}{n-1}\,\mathbb{E}\,(X_{(n)} - X_{(1)}) = \frac{n+1}{n-1}\,\frac{n-1}{n+1}\,(b-a) \\ &= b-a\end{aligned}$$

and we have, then, obtained an unbiased estimator of $b-a$.

Exercise 6.4.3 (p. 283)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Beta}(\theta, 1)$, i.e.

$$f(x|\theta) = \theta\, x^{\theta-1}\, \mathbb{1}_{(0,1)}(x)$$

Fine the complete and sufficient statistic for θ and identify a MVUE of $1/\theta$.

Solution.
Since

$$f(x|\theta) = \mathrm{e}^{(\theta-1)\log x + \log\theta}\, \mathbb{1}_{\{0,1\}}(x)$$

the statistical model is a one–parameter exponential family with $B(x) = \log x$. Hence $\sum_{j=1}^n \log X_j$ is a sufficient and complete statistics. As we wish to determine an unbiased estimator based on $\sum_{j=1}^n \log X_j$, note that

$$\mathbb{E}\log X_1 = \int_0^1 (\log x)\,\theta x^{\theta-1}\,\mathrm{d}x$$

(by the change of variable $w = -\log x$)

$$\begin{aligned}&= -\int_0^\infty w\,\theta\,\mathrm{e}^{-(\theta-1)w}\,\mathrm{e}^{-w}\,\mathrm{d}w = -\int_0^\infty w\,\theta\mathrm{e}^{-\theta w}\,\mathrm{d}w \\ &= -\frac{1}{\theta}.\end{aligned}$$

this implies that $\mathbb{E}(-\log X_1) = 1/\theta$. Hence the statistic

$$\hat{\theta} = -\frac{1}{n}\sum_{j=1}^n \log X_j$$

is such that

$$\mathbb{E}\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(-\log X_j) = \frac{1}{n}\,\frac{n}{\theta} = \frac{1}{\theta}.$$

Hence, $\hat{\theta}$ is unbiased for estimating θ and since it is a function of the complete and sufficient statistic $\sum_{j=1}^{n}\log X_j$ it is also a MVUE of $(1/\theta)$ because of the Lehmann-Scheffé theorem.

Exercise 6.4.5 (p.283).
Let $X_1, \dots, X_n$ be iid from the following density function

$$f(x|\theta) = \frac{2x}{\theta}\,\mathrm{e}^{-\frac{x^2}{\theta}}\,\mathbb{1}_{(0,+\infty)}(x)$$

(a) Show that $T = \sum_{j=1}^{n} X_j^2$ is sufficient for θ.

(b) Show that

$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} X_j^2$$

is an unbiased estimator for θ. Is it MVUE? Motivate the answer.

Solution.
Note that one may rewrite

$$f(x|\theta) = \mathrm{e}^{-\frac{x^2}{\theta}+\log(2x)-\log\theta}\,\mathbb{1}_{(0,+\infty)}(x)$$

and the model is, thus, a one–parameter exponential family with $B(x) = x^2$. Hence, $T = \sum_{j=1}^{n} B(X_j) = \sum_{j=1}^{n} X_j^2$ is a sufficient statistic for the model.
As for (b), from

$$\mathbb{E}X_1^2 = \int_0^\infty x^2\,\frac{2x}{\theta}\,\mathrm{e}^{-\frac{x^2}{\theta}}\,\mathrm{d}x = \frac{2}{\theta}\int_0^\infty x^3\,\mathrm{e}^{-\frac{x^2}{\theta}}\,\mathrm{d}x$$

(by the change of variable $w = x^2$)

$$\begin{aligned} &= \frac{2}{\theta}\int_0^\infty w^{3/2}\,\mathrm{e}^{-\frac{w}{\theta}}\,\frac{1}{2\sqrt{w}}\,\mathrm{d}w \\ &= \frac{1}{\theta}\int_0^\infty w\,\mathrm{e}^{-\frac{w}{\theta}}\,\mathrm{d}w = \frac{1}{\theta}\,\frac{\Gamma(2)}{(1/\theta)^2} \\ &= \theta \end{aligned}$$

one has $\mathbb{E}T = \sum_{j=1}^{n}\mathbb{E}X_j^2 = n\theta$, and this implies that

$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} X_j^2$$

is unbiased. Since it is further a function of the sufficient and complete statistics $\sum_{j=1}^{n} X_j^2$, by virtue of the Lehmann-Scheffé theorem $\hat{\theta}$ is a MVUE of θ

Exercise 63 (p. 298)
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Po}(\lambda)$. Find a MVUE of

$$\theta = \lambda \, \mathrm{e}^{-\lambda}$$

Solution.
Recall that

$$f(x|\lambda) = \frac{\lambda^x \, \mathrm{e}^{-\lambda}}{x!} \, \mathbb{1}_{\{0,1,2,\ldots\}}(x) = \mathrm{e}^{x \, \log \lambda - \lambda - \log(x!)} \, \mathbb{1}_{\{0,1,2,\ldots\}}(x)$$

and the statistical model is a one–parameter exponential family with $B(x) = x$ and, then, $\sum_{j=1}^{n} B(X_j) = \sum_{j=1}^{n} X_j = T$ is the sufficient and complete statistic. By the Lehmann-Scheffé theorem, a MVUE of θ is a function of $T = \sum_{j=1}^{n} X_j$. In order to obtain such an estimator, note that

$$\mathbb{P}[X_1 = 1] = \lambda \, \mathrm{e}^{-\lambda} = \theta$$

and $\mathbb{1}_{\{1\}}(X_1)$ is unbiased for θ, since $\mathbb{E}\mathbb{1}_{\{1\}}(X_1) = \mathbb{P}[X_1 = 1] = \lambda \, \mathrm{e}^{-\lambda}$. We can, now, identify the MVUE by considering

$$\hat{\theta} = \mathbb{E}[\mathbb{1}_{\{1\}}(X_1) \,|\, T]$$

which is an unbiased estimator of θ

$$\mathbb{E}\,\mathbb{E}[\mathbb{1}_{\{1\}}(X_1) \,|\, T] = \mathbb{E}\mathbb{1}_{\{1\}}(X_1) = \theta = \lambda \, \mathrm{e}^{-\lambda}$$

where the first equality follows from a well–known property of conditional expectation. Moreover, $\hat{\theta}$ is a function of T and, by Lehmann–Scheffé theorem is a MVUE of θ. We now determine an explicit representation for $\hat{\theta}$. It is worth reminding that $X_1, \ldots, X_q \overset{\text{iid}}{\sim} \text{Po}(\lambda)$ implies that $\sum_{j=1}^{q} X_j \sim \text{Po}(q\lambda)$. By the definition of conditional expectation and probability one has

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{\{1\}}(X_1) \mid T = t] = \mathbb{P}[X_1 = 1 \mid T = t] &= \frac{\mathbb{P}[X_1 = 1, \sum_{j=1}^n X_j = t]}{\mathbb{P}[\sum_{j=1}^n X_j = t]]} \\
&= \frac{\mathbb{P}[X_1 = 1, \sum_{j=2}^n X_j = t-1]}{\mathbb{P}[\sum_{j=1}^n X_j = t]]} \\
&= \frac{\mathbb{P}[X_1 = 1]\, \mathbb{P}[\sum_{j=2}^n X_j = t-1]}{\mathbb{P}[\sum_{j=1}^n X_j = t]} \\
&= \frac{\lambda\, \mathrm{e}^{-\lambda}\, \frac{((n-1)\lambda)^{t-1}\, \mathrm{e}^{-(n-1)\lambda}}{(t-1)!}}{\frac{(n\lambda)^t\, \mathrm{e}^{-n\lambda}}{t!}} \\
&= \frac{(n-1)^{t-1}}{n^t}\, t.
\end{aligned}$$

This allows us to conlude that

$$\hat{\theta} = \frac{(n-1)^{\sum_{j=1}^n X_j - 1}}{n^{\sum_{j=1}^n X_j}} \sum_{j=1}^n X_j$$

is a MVUE of $\theta = \lambda\, \mathrm{e}^{-\lambda}$.

Exercise

Let $X_1, \ldots, X_n$ be iid from a uniform in $(\theta, 1)$, namely

$$f(x|\theta) = \frac{1}{1-\theta}\, \mathbb{1}_{(\theta,1)}(x) \qquad \theta \in (-\infty, 1).$$

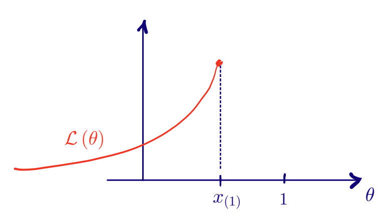
(a) Show that $\hat{\theta} = X_{(1)}$ is the maximum likelihood estimator of θ.

(b) Show that $\hat{\theta}$ is consistent for estimating θ and determine the asymptotic distribution of $n(\hat{\theta} - \theta)$ as $n \to \infty$.

Solution.

The likelihood function is

$$\begin{aligned}
\mathcal{L}(\theta) = \prod_{j=1}^n f(x_j|\theta) &= \prod_{j=1}^n \frac{1}{1-\theta}\, \mathbb{1}_{(\theta,1)}(x_j) \\
&= \frac{1}{(1-\theta)^n}\, \mathbb{1}_{(\theta,+\infty)}(x_{(1)})\, \mathbb{1}_{(-\infty,1)}(x_{(n)}) \\
&= \frac{1}{(1-\theta)^n}\, \mathbb{1}_{(-\infty,x_{(1)})}(\theta)\, \mathbb{1}_{(-\infty,1)}(x_{(n)})
\end{aligned}$$



See the plot on the right, from which it is apparent that the maximum is achieved at $x_{(1)} = \min\{x_1, \ldots, x_n\}$, namely

$$\mathcal{L}(x_{(1)}) = \max_{\theta<1} \mathcal{L}(\theta)$$

This shows that $\hat{\theta} = X_{(1)}$ is the MLE of θ. As for (b), one first needs to determine c.d.f. $F_{X_{(1)}}$ of $X_{(1)}$. Since

$$F(x|\theta) = \int_{-\infty}^{x} f(s|\theta)\,\mathrm{d}s = \frac{x-\theta}{1-\theta}\,\mathbb{1}_{(\theta,1)}(x) + \mathbb{1}_{[1,+\infty)}(x),$$

one has

$$F_{X_{(1)}}(x) = 1 - \Big\{1 - F(x|\theta)\Big\}^n = 1 - \Big(\frac{1-x}{1-\theta}\Big)^n \mathbb{1}_{(\theta,1)}(x) + \mathbb{1}_{[1,+\infty)}(x).$$

If $Y_n = n(\hat{\theta} - \theta)$, it can be seen that with n large enough

$$F_{Y_n}(y) = F_{X_{(1)}}\Big(\frac{y}{n} + \theta\Big) = 1 - \Big(1 - \frac{y}{(1-\theta)n}\Big)^n \to 1 - \mathrm{e}^{-y/(1-\theta)}.$$

Henve, the limiting distribution is negative exponential with parameter $1/(1-\theta)$.

Exercise

Let $X_1, \ldots, X_n$ be i.i.d. random variables with common pdf

$$f(x|\theta) = \theta^2 x\,\mathrm{e}^{-\theta x}\mathbb{1}_{(0,+\infty)}(x), \qquad \theta > 0.$$

(a) Determine the MLE $\hat{\theta}_n$ of θ.

(b) Show that the MLE $\hat{\theta}_n$ satisfies

$$\frac{\sqrt{2n}(\hat{\theta}_n - \theta)}{\theta} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1)$$

where $\mathrm{N}(0,1)$ denotes a standard normal random variable.

(c) Consider the prior distribution $g(\theta) := \mathrm{e}^{-\theta}\mathbb{1}_{(0,+\infty)}(\theta)$ for the parameter θ, determine the posterior distribution of θ.

Solution.

As for (a), note that the likelihood function is

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \theta^2 x_i\,\mathrm{e}^{-\theta x_i} = \theta^{2n}\Big(\prod_{i=1}^{n} x_i\Big)\,\mathrm{e}^{-\theta\sum_{i=1}^{n} x_i}$$

for any $x_1, \ldots, x_n > 0$. The log-likelihood, then, equals

$$\log \mathcal{L}(\theta) = 2n\log\theta + \sum_{i=1}^{n}\log x_i - \theta\sum_{i=1}^{n} x_i.$$

From the equation

$$\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta) = \frac{2n}{\theta} - \sum_{j=1}^{n} x_j = 0$$

one obtains $\hat{\theta}_n = 2/\bar{x}_n$. Since one further has

$$\frac{\partial^2}{\partial\theta^2}\log\mathcal{L}(\theta)\Big|_{\theta=\hat{\theta}_n} = -\frac{2n}{\hat{\theta}_n^2} < 0$$

one has that $\hat{\theta}_n$ actually is the MLE of θ. Moreover, note that the Fisher information equals

$$\mathcal{I}_{X_1}(\theta) = -\mathbb{E}_\theta\frac{\partial^2}{\partial\theta^2}\Big(2\log\theta + \log X_1 - \theta X_1\Big) = \frac{2}{\theta^2},$$

by properties of MLEs one has $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{D}} \mathrm{N}(0, 2/\theta^2)$ and the limiting result in (b) follows.

Finally, with a negative exponential prior, with parameter 1, on θ, the posterior is

$$g(\theta|x_1,\ldots,x_n) \propto \theta^{2n}\,\mathrm{e}^{-\theta\sum_{i=1}^{n}x_i}\,\mathrm{e}^{-\theta}\,\mathbb{1}_{(0,+\infty)}(\theta)$$

which is seen to be a gamma density function with parameters $(2n+1, \sum_{j=1}^{n}x_j + 1)$, namely

$$g(\theta|x_1,\ldots,x_n) = \frac{(\sum_{j=1}^{n}x_j+1)^{2n+1}}{\Gamma(2n+1)}\,\theta^{2n}\,\mathrm{e}^{-\theta\sum_{i=1}^{n}x_i}\,\mathrm{e}^{-\theta}\,\mathbb{1}_{(0,+\infty)}(\theta)$$

and this completes (c).

Exercise

Let $X_1,\ldots,X_n$ be i.i.d. random variables with common pdf

$$f(x|\alpha) = \frac{\alpha}{x^{\alpha+1}}\mathbb{1}_{(1,+\infty)}(x), \qquad \alpha > 0.$$

(a) Determine the MLE $\hat{\tau}_n$ of $\tau = 2/\alpha$.

(b) Consider the prior distribution $g(\alpha) := e^{-\alpha}\mathbb{1}_{(0,+\infty)}(\alpha)$ for the parameter α, determine the Bayes estimator of α under a squared loss function.

(c) Consider the same prior of point (b), identify the predictive density $f(x_2|x_1)$ of X_2 given $X_1 = x_1$.

Solution.

The MLE of $\tau = 2/\alpha$ can be determined through the invariance property, once we have ontained the MLE of α. To this end, note that, for any $x_1, \dots, x_n > 1$,

$$\mathcal{L}(\alpha) = \frac{\alpha^n}{\prod_{i=1}^n x_i^{\alpha+1}} \quad \Longrightarrow \quad \log \mathcal{L}(\alpha) = n \log \alpha - (\alpha + 1) \sum_{i=1}^n \log x_i$$

From the score equation

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} \log \mathcal{L}(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^n \log x_i = 0$$

one has $\alpha_n = n / \sum_{i=1}^n \log x_i$. Moreover, since

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \log \mathcal{L}(\alpha) \Big|_{\alpha=\hat{\alpha}_n} = -\frac{n}{\hat{\alpha}_n^2} < 0$$

then $\hat{\alpha}_n$ actually is the MLE of α. The answer to (a) is, then, completed upon using the invariance property of MLEs hich in this case yields

$$\hat{\tau}_n = \frac{2}{\hat{\alpha}_n} = \frac{2 \sum_{i=1}^n \log x_i}{n}$$

As for (b), note that the posterior density function of α is

$$g(\alpha | x_1, \dots, x_n) \propto \frac{\alpha^n}{\prod_{i=1}^n x_i^{\alpha}} \, \mathrm{e}^{-\alpha} \, \mathbb{1}_{(0,+\infty)}(\alpha) = \alpha^n \, \mathrm{e}^{-\alpha(1+\sum_{i=1}^n \log x_i)} \, \mathbb{1}_{(0,+\infty)}(\alpha)$$

which can be identified as the density function of a Gamma distribution with parameters $(n+1,\, 1 + \sum_{i=1}^n \log x_i)$, namely

$$g(\alpha | x_1, \dots, x_n) = \frac{\left(1 + \sum_{i=1}^n \log x_i\right)^{n+1}}{\Gamma(n+1)} \, \alpha^n \, \mathrm{e}^{-\alpha(1+\sum_{i=1}^n \log x_i)} \, \mathbb{1}_{(0,+\infty)}(\alpha).$$

The Bayesian estimator of α under square loss is known to be $\hat{\alpha}_{n,\mathrm{B}} = \mathbb{E}[\alpha | x_1, \dots, x_n]$ and, in view of the posterior distribution we have determined, one has

$$\begin{aligned}
\hat{\alpha}_{n,\mathrm{B}} &= \int_0^\infty \alpha \, g(\alpha | x_1, \dots, x_n) \, \mathrm{d}\alpha \\
&= \frac{\left(1 + \sum_{i=1}^n \log x_i\right)^{n+1}}{\Gamma(n+1)} \int_0^\infty \alpha^{n+2} \, \mathrm{e}^{-\alpha(1+\sum_{i=1}^n \log x_i)} \, \mathrm{d}\alpha \\
&= \frac{\left(1 + \sum_{i=1}^n \log x_i\right)^{n+1}}{\Gamma(n+1)} \, \frac{\Gamma(n+2)}{\left(1 + \sum_{i=1}^n \log x_i\right)^{n+2}} \\
&= \frac{n+1}{1 + \sum_{i=1}^n \log x_i}
\end{aligned}$$

As for (c), one may evaluate the predictive density $f(x_2 | x_1)$ through different calculations. For example, one may first determine the marginal

$f(x_1) = \int_0^\infty f(x_1|\alpha)\, g(\alpha)\, \mathrm{d}\alpha$ and the joint $f(x_1, x_2) = \int_0^\infty f(x_1|\alpha) f(x_2|\alpha) g(\alpha)\, \mathrm{d}\alpha$ densities and, then, obtain $f(x_2|x_1) = f(x_1, x_2)/f(x_1)$. Alternatively, by simple integration with respect to the posterior, for any $x_1 > 1$ one has

$$\begin{aligned}
f(x_2|x_1) &= \int_0^\infty f(x_2|\alpha)\, g(\alpha|x_1)\, \mathrm{d}\alpha \\
&= \int_0^\infty \frac{\alpha}{x_2^{\alpha+1}} \frac{(1+\log x_1)^2}{\Gamma(2)} \alpha\, \mathrm{e}^{-\alpha(1+\log x_1)}\, \mathrm{d}\alpha\, \mathbb{1}_{(1,+\infty)}(x_2) \\
&= \frac{(1+\log x_1)^2}{x_2} \int_0^\infty \alpha^2\, \mathrm{e}^{-\alpha(1+\log x_1+\log x_2)}\, \mathrm{d}\alpha\, \mathbb{1}_{(1,+\infty)}(x_2) \\
&= \frac{2(1+\log x_1)^2}{x_2(1+\log x_1+\log x_2)^3}\, \mathbb{1}_{(1,+\infty)}(x_2)
\end{aligned}$$

Exercise

Suppose $X_1, \dots, X_n$ are the measurements of radii of ball bearings produced by a factory. It is assumed that $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, 0.64)$.

(a) Determine the MLE of the radius mean μ and identify its probability distribution.

(b) Determine the MLE $\hat{h}$ of the area $h(\mu) = \pi\mu^2$ and, if biased, propose an unbiased estimator that is based on it.

(c) Let
$$\hat{h}^* = \frac{\pi}{n} \sum_{i=1}^n X_i^2 - 0.64\,\pi$$
Show that $\hat{h}^*$ is consistent for estimating $h(\mu)$.

Solution.

Since the likelihood function is

$$\mathcal{L}(\mu) = \prod_{i=1}^n \frac{1}{0.8\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{1.28}(x_i-\mu)^2} = \frac{1}{0.8^n\,(2\pi)^{n/2}}\, \mathrm{e}^{-\frac{1}{1.28}\sum_{i=1}^n (x_i-\mu)^2}$$

one has

$$\log \mathcal{L}(\mu) = -n\log 0.8 - \frac{n}{2}\log(2\pi) - \frac{1}{1.28}\sum_{i=1}^n (x_i-\mu)^2$$

and it, then, follows that the solution of the scoring equation

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\log \mathcal{L}(\mu) = \frac{1}{1.28}\sum_{i=1}^n (x_i-\mu) = 0$$

is the sample mean of the n measurements $x_1, \dots, x_n$, namely $\hat{\mu} = \bar{x}$. Moreover,

$$\frac{d^2}{d\mu^2} \log \mathcal{L}(\mu) \Big|_{\mu=\hat{\mu}} = -\frac{n}{1.28} < 0$$

and $\hat{\mu}$ actually is a MLE. The answer to (a) is completed upon noting that, due to well–known properties of Gaussian distributions, one has $\hat{\mu} \sim \mathrm{N}(\mu, 0.64/n)$.

As for (b), one may resort to the invariance property of MLEs and note that

$$\hat{h} = h(\hat{\mu}) = \pi \bar{X}^2$$

is the MLE of $h(\mu)$. Since

$$\mathbb{E}\bar{X}^2 = Var(\bar{X}) + \left(\mathbb{E}\bar{X}\right)^2 = \frac{0.64}{n} + \mu^2,$$

one has

$$\mathbb{E}\hat{h} = \pi \frac{0.64}{n} + \pi \mu^2$$

and the MLE $\hat{h}$ is biased. An unbiased estimator is obtained if one sets

$$\hat{h}_{\mathrm{u}} = \hat{h} - \pi \frac{0.64}{n}.$$

As for (c), note that $X_1^2, \dots, X_n^2$ sono indipendenti e identicamente distribuite con $\mathbb{E}X_1^2 = 0.64 + \mu^2$ and $\mathrm{Var}(X_1) < \infty$. Hence, by the Storng Law of Large Numbers

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{\text{as}} 0.64 + \mu^2$$

and, *a fortiori*,

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{\text{p}} 0.64 + \mu^2.$$

This implies

$$\hat{h}^* = \frac{\pi}{n} \sum_{i=1}^{n} X_i^2 - 0.64\pi \xrightarrow{\text{p}} \pi\{0.64 + \mu^2\} - 0.64\pi = \pi\mu^2 = h(\mu)$$

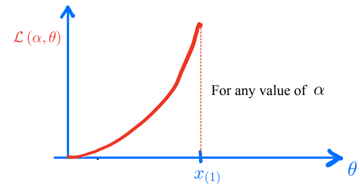
which shows that the estimator $\hat{h}^*$ is consistent for estimating $h(\mu)$ and, thus, answers (c).

Exercise

Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Pareto}(\alpha, \theta)$, namely

$$f(x|\alpha, \theta) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \mathbb{1}_{[\theta, +\infty)}(x).$$

Determine the MLE of (α, θ).



Solution.
Note that for any value of $\alpha > 0$, the likelihood function is

$$\mathcal{L}(\alpha,\theta) = \frac{\alpha^n\,\theta^{n\alpha}}{\prod_{i=1}^n x_i^{\alpha+1}}\,\mathbb{1}_{(0,x_{(1)}]}(\theta)$$

where $x_{(1)} = \min\{x_1,\ldots,x_n\}$ is the first order statistic. From the plot it can be seen that for any value of α the MLE is $\hat{\theta} = x_{(1)}$.

Since $\hat{\theta}$ does not depend on α, one may plug it in the likelihood function and determine the MLE of α. In other terms, we aim at determining $\hat{\alpha}$ such that

$$\mathcal{L}(\hat{\alpha},\hat{\theta}) = \max_{\alpha>0}\mathcal{L}(\alpha, x_{(1)})$$

and to this end, we shall consider

$$\log\mathcal{L}(\alpha,x_{(1)}) = n\log\alpha + \alpha n\log x_{(1)} - (\alpha+1)\sum_{i=1}^n \log x_i.$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\mathcal{L}(\alpha,x_{(1)}) = \frac{n}{\alpha} + n\log x_{(1)} - \sum_{i=1}^n\log x_i = 0,$$

it turns out that the solution is $\hat{\alpha} = n/\{\sum_{i=1}^n \log(x_i/x_{(1)})\}$. Moreover,

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\log\mathcal{L}(\alpha,x_{(1)})\Big|_{\alpha=\hat{\alpha}} = -\frac{n}{\hat{\alpha}^2} < 0$$

and we can conclude that

$$(\hat{\alpha},\hat{\theta}) = \left(\frac{n}{\sum_{i=1}^n\log(x_i/x_{(1)})},\, x_{(1)}\right)$$

is the MLE of (α,θ).

Exercise
Suppose $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \text{Beta}(\theta,1)$, namely

$$f(x|\theta) = \theta\,x^{\theta-1}\,\mathbb{1}_{(0,1)}(x).$$

Determine the MLE of θ. Determine its asymptotic distribution, as $n\to\infty$.

Solution.
Since the likelihood function is

$$\mathcal{L}(\theta) = \prod_{i=1}^n \theta x_i^{\theta-1}\,\mathbb{1}_{(0,1)}(x_i)$$

for any $x_1, \ldots, x_n$ in $(0,1)$ one has

$$\log \mathcal{L}(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i$$

and the solution of the $(\mathrm{d}/\mathrm{d}\theta) \log \mathcal{L}(\theta) = 0$ is $\hat{\theta} = -n/\{\sum_{i=1}^{n} \log x_i\}$. Since

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log \mathcal{L}(\theta) \Big|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$$

and $\hat{\theta}$ is the MLE of θ. As for its asymnptotic distribution, note that

$$\log f(x|\theta) = \log \theta + (\theta - 1) \log x$$

which implies that

$$\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) = -\frac{1}{\theta^2}$$

Hence, $mathcalI_{X_1}(\theta) = 1/\theta^2$ and the requested asymptotic distribution is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathrm{D}} \mathrm{N}(0, \theta^2).$$

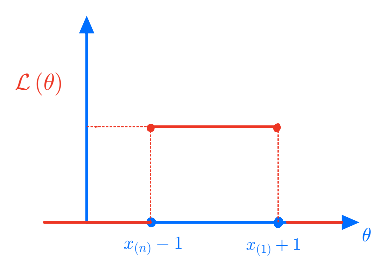
Exercise
Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}(\theta - 1, \theta + 1)$, namely

$$f(x|\theta) = \frac{1}{2}\, \mathbb{1}_{(\theta-1,\theta+1)}(x).$$

Determine the MLE of θ.

Solution.
Since the likelihood function is

$$\mathcal{L}(\theta) = \frac{1}{2^n} \prod_{i=1}^{n} \mathbb{1}_{(\theta-1,\theta+1)}(x_i) = \frac{1}{2^n}\, \mathbb{1}_{(\theta-1,+\infty)}(x_{(1)})\, \mathbb{1}_{(-\infty,\theta+1)}(x_{(n)})$$
$$= \frac{1}{2^n}\, \mathbb{1}_{(x_{(n)}-1, x_{(1)}+1)}(\theta)$$



from its plot it is apparent that the MLE is not unique and any point in the interval $(x_{(n)} - 1, x_{(1)} + 1)$ is a MLE for θ. Also $|(\theta - 1, \theta + 1)| = 2$ implies that $x_{(n)} - x_{(1)} < 2$ so that one has $x_{(n)} - 1 < x_{(1)} + 1$ as represented in the plot on the right.

Exercise

Let $X_1, \dots, X_n$ be i.i.d. random variables with common pdf

$$f(x|\sigma) = x\sigma \mathrm{e}^{-\sigma x^2/2} \mathbb{1}_{(0,+\infty)}(x), \qquad \sigma > 0.$$

(a) Determine the MLE $\hat{\sigma}_n$ of σ and show it is consistent.

(b) Consider the prior distribution $g(\sigma) = \mathrm{e}^{-\sigma} \mathbb{1}_{(0,+\infty)}(\sigma)$ for the parameter σ, determine the posterior distribution of σ, given the sample $X_1 = x_1, \dots, X_n = x_n$.

(c) Consider the same prior of point (b), determine the Bayes estimator $\hat{\sigma}_g$ of σ under a squared loss function.

Solution.

For any $x_1, \dots, x_n > 0$ the likelihood function is

$$\mathcal{L}(\sigma) = \sigma^n \left(\prod_{i=1}^n x_i \right) \mathrm{e}^{-\frac{\sigma}{2} \sum_{i=1}^n x_i^2}$$

which yields

$$\log \mathcal{L}(\sigma) = n \log \sigma + \sum_{i=1}^n \log x_i - \frac{\sigma}{2} \sum_{i=1}^n x_i^2.$$

The solution to

$$\frac{\mathrm{d}}{\mathrm{d}\sigma} \mathcal{L}(\sigma) = \frac{n}{\sigma} - \frac{1}{2} \sum_{i=1}^n x_i^2 = 0$$

is $\hat{\sigma}_n = 2n / \{\sum_{i=1}^n x_i^2\}$. This is the MLE of σ since

$$\frac{\mathrm{d}^2}{\mathrm{d}\sigma^2} \log \mathcal{L}(\sigma) \Big|_{\sigma = \hat{\sigma}_n} = -\frac{n}{\hat{\sigma}_n^2} < 0.$$

Incidentally, note that

$$\mathbb{E} X_1^2 = \sigma \int_0^{+\infty} x^3 \, \mathrm{e}^{-\frac{\sigma}{2} x} \, \mathrm{d}x = \frac{\sigma}{2} \int_0^{+\infty} w \, \mathrm{e}^{-\frac{\sigma}{2} w} \, \mathrm{d}w$$

where the second equality follows upon considering the change of variable $x^2 = w$. By solving the gamma integral above, one obtains $\mathbb{E} X_1^2 = 2/\sigma$. Hence, by the SLLN

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\text{as}} \frac{2}{\sigma} \quad \Longrightarrow \quad \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\text{p}} \frac{2}{\sigma}.$$

In view of this, one has

$$\hat{\sigma}_n = \frac{2}{\frac{1}{n} \sum_{i=1}^n X_i^2} \xrightarrow{\text{p}} \frac{2}{2/\sigma} = \sigma$$

which proves consistency of $\hat{\sigma}_n$ and completes (a).

If one further specifies a prior distribution for σ and, in particular it is assumed that $\sigma \sim$ neg-exp(1), one has

$$g(\sigma|x_1,\ldots,x_n) \propto \sigma^n \,\mathrm{e}^{-\frac{\sigma}{2}\sum_{i=1}^n x_i^2}\,\mathrm{e}^{-\sigma}\,\mathbb{1}_{(0,+\infty)}(\sigma)$$

which, up to a proportionality constant, is the density function of a Gamma$(n+1, \sum_{i=1}^n x_i^2/2+1)$ distribution, namely

$$g(\sigma|x_1,\ldots,x_n) = \frac{\left(\frac{1}{2}\sum_{i=1}^n x_i^2+1\right)^{n+1}}{\Gamma(n+1)}\,\sigma^n\,\mathrm{e}^{-(\frac{1}{2}\sum_{i=1}^n x_i^2+1)\,\sigma}\,\mathbb{1}_{(0,+\infty)}(\sigma)$$

The Bayesian estimator of σ under square loss is

$$\hat{\sigma}_g = \mathbb{E}[\sigma|x_1,\ldots,x_n] = \frac{n+1}{\frac{1}{2}\sum_{i=1}^n x_i^2+1}$$

Exercise

Let $X_1,\ldots,X_n$ be i.i.d. random variables with common pdf

$$f(x|\theta) = \mathrm{e}^{-(x-\theta)}\mathbb{1}_{(\theta,+\infty)}(x), \qquad \theta > 0.$$

Determine the MLE $\hat{\theta}_n$ of θ and establish whether it is unbiased and consistent.

Solution.

The determination of the MLE is immediate, once one observes that the likelihood function has the following form

$$\mathcal{L}(\theta) = \prod_{i=1}^n \mathrm{e}^{-(x_i-\theta)}\,\mathbb{1}_{(\theta,+\infty)}(x_i) = \mathrm{e}^{n\theta-\sum_{i=1}^n x_i}\,\mathbb{1}_{(0,x_{(1)})}(\theta)$$

and a simple representation of the likelihood function shows that the MLE is $\hat{\theta}_n = x_{(1)}$. Unbiasedness and consistency can be checked once we have determined the density function of $X_{(1)}$. To this end, note that

$$F_\theta(x) = \mathbb{1}_{(\theta,+\infty)}(x)\int_\theta^s \mathrm{e}^{-(s-\theta)}\,\mathrm{d}s = \{1-\mathrm{e}^{-(x-\theta)}\}\,\mathbb{1}_{(\theta,+\infty)}(x)$$

which implies that the probability density function of $X_{(1)}$ is

$$\begin{aligned} f_{X_{(1)}}(x|\theta) &= n\,\mathrm{e}^{-(x-\theta)}\,\{1-1+\mathrm{e}^{-(x-\theta)}\}^{n-1}\,\mathbb{1}_{(\theta,+\infty)}(x) \\ &= n\,\mathrm{e}^{-n(x-\theta)}\,\mathbb{1}_{(\theta,+\infty)}(x). \end{aligned}$$

With this, one can show that

$$\mathbb{E}_\theta \hat{\theta}_n = \mathbb{E}_\theta X_{(1)} = n \int_\theta^{+\infty} x\, \mathrm{e}^{-n(x-\theta)}\, \mathrm{d}x = n \int_0^{+\infty} (y+\theta)\, \mathrm{e}^{-ny}\, \mathrm{d}y$$
$$= \frac{1}{n} + \theta \neq \theta$$

which implies that $X_{(1)}$ is biased. Nonetheless, it is asymptotically unibased since $\lim_{n\to\infty} \mathbb{E}_\theta X_{(1)} = \theta$ for any $\theta > 0$. As for consistency, one has

$$\mathbb{P}_\theta[|X_{(1)} - \theta| > \varepsilon] = \mathbb{P}_\theta[X_{(1)} > \theta + \varepsilon] = \int_{\theta+\varepsilon}^{+\infty} n\, \mathrm{e}^{-n(x-\theta)}\, \mathrm{d}x = \mathrm{e}^{-n\varepsilon}$$

and, since $\varepsilon > 0$, consistency follows

$$\lim_{n\to\infty} \mathbb{P}_\theta[|X_{(1)} - \theta| > \varepsilon] = \lim_{n\to\infty} \mathrm{e}^{-n\varepsilon} = 0$$

so that $X_{(1)} \xrightarrow{\mathrm{P}} \theta$.

Exercise

Let $X_1, \ldots, X_n$ be i.i.d. random variables with common pdf

$$f(x|\theta) = \frac{2x}{\theta^2}\, \mathbb{1}_{[0,\theta]}(x), \qquad \theta > 0.$$

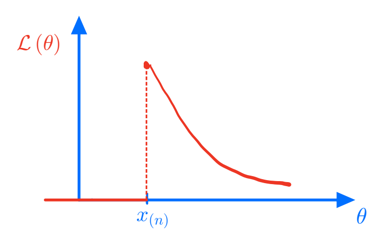
(a) Identify a sufficient statistic for θ

(b) Determine the MLE $\hat{\theta}_n$ of θ and establish whether it is unbiased and consistent.

Solution.

Since

$$\prod_{j=1}^n f(x_j|\theta) = \prod_{j=1}^n \frac{2x_j}{\theta}\, \mathbb{1}_{([0,\theta]}(x_j) = \frac{2^n}{\theta^n} \left\{ \prod_{j=1}^n x_j \right\} \mathbb{1}_{(-\infty,\theta]}(x_{(n)})\, \mathbb{1}_{[0,+\infty)}(x_{(1)})$$



from the factorization theorem one concludes that the n–th order $T = X_{(n)} = \max\{X_1, \ldots, x_n\}$ is the sufficient statistic for θ. As for (b), from previous calculations the likelihood function has the following representation

$$\mathcal{L}(\theta) = \frac{2^n \prod_{j=1}^n x_j}{\theta^n}\, \mathbb{1}_{[0,+\infty)}(x_{(1)})\, \mathbb{1}_{[x_{(n)},+\infty)}(\theta)$$

The function is not differentiable with respect to θ, but the maximum can be easily identified by relying on the plot of $\mathcal{L}$, on the side of this page. It can be seen that the maximum is achieved at $\hat{\theta}_= X_{(n)}$.

Moreover, since

$$F(x|\theta) = \mathbb{1}_{[0,\theta)}(x) \int_0^x \frac{2s}{\theta^2}\, \mathrm{d}s + \mathbb{1}_{[\theta,+\infty)}(x) = \frac{x^2}{\theta^2}\, \mathbb{1}_{[0,\theta)}(x) + \mathbb{1}_{[\theta,+\infty)}(x),$$

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

$$f_{X(n)}(x|\theta) = n\, \frac{2x}{\theta^2} \Big\{ \frac{x^2}{\theta^2} \Big\}^{n-1} \mathbb{1}_{[0,\theta]}(x) = \frac{2n\, x^{2n-1}}{\theta^{2n}}\, \mathbb{1}_{[0,\theta]}(x)$$

and from this one obtains

$$\begin{aligned} P_\theta[|X_{(n)} - \theta| < \varepsilon] = P_\theta[\theta - \varepsilon < X_{(n)} < \theta + \varepsilon] &= \int_{\theta-\varepsilon}^{\theta} \frac{2n\, x^{2n-1}}{\theta^{2n}}\, \mathrm{d}x \\ &= 1 - \Big(\frac{\theta - \varepsilon}{\theta} \Big)^{2n} \end{aligned}$$

which implies that $\lim_{n\to\infty} P_\theta[|X_{(n)} - \theta| < \varepsilon] = 1$ for any $\varepsilon > 0$ and $\theta > 0$ and $\hat{\theta}_n = X_{(n)}$ is, thus, consistent for estimating θ.

Exercise
Let $X_1, \dots, X_n$ be i.i.d. random variables from a $\mathrm{N}(\theta, \theta)$ distribution, with $\theta > 0$. Determine the MLE of θ and show it is consistent for estimating θ.

Solution.
Since the likelihood is

$$\mathcal{L}(\theta) = \frac{1}{\theta^{n/2}(2\pi)^{n/2}}\, \mathrm{e}^{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \theta)^2}$$

the log-likelihood boils down to

$$\log \mathcal{L}(\theta) = -\frac{n}{2} \log \theta - \frac{n}{2} \log(2\pi) - \frac{1}{2\theta} \sum_{i=1}^n x_i^2 + \frac{1}{2} \sum_{i=1}^n x_i - \frac{n\theta}{2}$$

By simple evaluation of the first derivative, one has

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log \mathcal{L}(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 - \frac{n}{2} = 0$$

This equaiton has two solutions, though only one of them is positive and, hence, in $\Theta = (0, +\infty)$. Such a solution is

$$\hat{\theta}_n = \frac{1}{2} \Big\{ \sqrt{1 + \frac{4}{n} \sum_{i=1}^n x_i^2} - 1 \Big\}$$

The fact that this is the point where the maximum is attained can be verified by observing that

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log \mathcal{L}(\theta) \Big|_{\theta = \hat{\theta}_n} = \frac{n}{2\hat{\theta}_n^2} - \frac{1}{\hat{\theta}_n^3} \sum_{i=1}^n x_i^2 = \frac{n}{2\hat{\theta}_n^3} \Big(\hat{\theta}_n - \frac{2}{n} \sum_{i=1}^n x_i^2 \Big) < 0$$

since

$$\hat{\theta}_n - \frac{2}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{2}\Big(\sqrt{1+\frac{4}{n}\sum_{i=1}^{n} x_i^2} - 1 - \frac{4}{n}\sum_{i=1}^{n} x_i^2\Big) < 0$$

In order to check consistency, it is worth noting that if a sequence $(X_n)_{n\geq 1}$ of random variables converges in probability to a random variable X, namely $X_n \xrightarrow{\text{p}} X$, for any continuous function $g : \mathbb{R} \to \mathbb{R}$ one has $g(X_n) \xrightarrow{\text{p}} g(X)$. By the strong law of large numbers

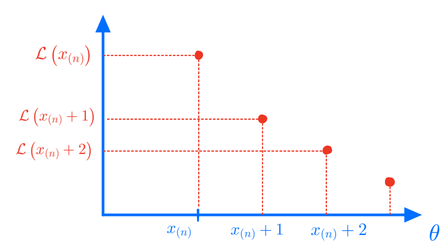
$$\frac{1}{n}\sum_{X_i^2} \xrightarrow{\text{as}} \mathbb{E}_\theta X^2 = \theta + \theta^2$$

and, a fortiori, $(1/n)\sum_{i=1}^{n} X_i^2 \xrightarrow{\text{p}} \theta + \theta^2$. By virtue of the property that we have just recalled

$$\hat{\theta}_n \xrightarrow{\text{p}} \frac{1}{2}\left(\sqrt{1+4\theta+4\theta^2} - 1\right) = \frac{1}{2}(1+2\theta-1) = \theta$$

which shows consistency.

> **Exercise**
> Let $X_1, \ldots, X_n$ be i.i.d. random variables from adiscrete uniform distribution on the set $\{1, \ldots, \theta\}$, where the parameter space $\theta \in \mathbb{N} = \Theta$ is the set of positive integers. Determine the MLE of θ and study its asymptotic beahviour.



Solution.
The likelihood function is given by

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{1}_{\{1,\ldots,\theta\}}(x_i) = \frac{1}{\theta^n}\mathbb{1}_{\{1,2,\ldots\}}(x_{(1)})\,\mathbb{1}_{\{\ldots,-1,0,1,\ldots,\theta\}}(x_{(n)})$$
$$= \frac{1}{\theta^n}\mathbb{1}_{\{1,2,\ldots\}}(x_{(1)})\,\mathbb{1}_{\{x_{(n)},x_{(n)}+1,\ldots\}}(\theta)$$

and is depicted in the plot on the side of the page. From the plot it is apparent that the MLE $\hat{\theta}_n = X_{(n)}$ is the n–th order statistic.

In order to study consistency, note that

$$F(x|\theta) = \sum_{1\leq s\leq \min\{x,\theta\}} \frac{1}{\theta} = \frac{[x]}{\theta}\mathbb{1}_{\{1,\ldots,\theta\}}(x) + \mathbb{1}_{\{\theta+1,\theta+2,\ldots\}}(x)$$

where $[x]$ denotes the integer part of x, namely the largest integer less than or equal to x. This implies that

$$F_{X_{(n)}}(x|\theta) = \frac{[x]^n}{\theta^n}\mathbb{1}_{\{1,\ldots,\theta\}}(x) + \mathbb{1}_{\{\theta+1,\theta+2,\ldots\}}(x)$$

and

$$
\begin{aligned}
P_\theta[|X_{(n)} - \theta| < \varepsilon] = P_\theta[X_{(n)} > \theta - \varepsilon] &= 1 - F_{X_{(n)}}(\theta - \varepsilon|\theta) \\
&= 1 - \Big(\frac{[\theta + \varepsilon - 1]}{\theta}\Big)^n = 1 - \Big(\frac{\theta - 1}{\theta}\Big)^n \\
&\to 1 \qquad\qquad (\text{as } n \to \infty)
\end{aligned}
$$

so that $X^{(n)}$ is consistent for estimating θ.

Exercise

Let $X_1, \ldots, X_{20} \,|\, \theta \overset{\text{iid}}{\sim} \text{Bern}(\theta)$ and suppose $\theta \sim \text{Beta}(2,2)$, namely

$$\mathbb{P}[X_i = x \,|\, \theta] = \theta^x (1-\theta)^{1-x} \mathbb{1}_{\{0,1\}}(x), \qquad g(\theta) = 6\,\theta(1-\theta)\mathbb{1}_{[0,1]}(\theta).$$

(a) Identify the posterior distribution of θ, given $X_1 = x_1, \ldots, X_{20} = x_{20}$

(b) Determine the predictive distribution of X_{21}, given the observed sample, namely $\mathbb{P}[X_{21} = 1 \,|\, x_1, \ldots, x_{20}]$

(c) If the observed sample is such that $s_{20} = \sum_{i=1}^{20} x_i = 10$, determine

$$\mathbb{P}[\theta < 0.5 \,|\, x_1, \ldots, x_{20}]$$

Solution.

By Bayes' theorem, the posterior distribution has probability density function

$$
\begin{aligned}
g(\theta|x_1, \ldots, x_n) \propto \mathcal{L}(\theta)\, g(\theta) &= \Big\{ \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} \Big\}\, \theta(1-\theta)\, \mathbb{1}_{(0,1)}(\theta) \\
&= \theta^{\sum_{i=1}^n x_i + 1} (1-\theta)^{n - \sum_{i=1}^n x_i + 1}\, \mathbb{1}_{(0,1)}(\theta)
\end{aligned}
$$

and this implies that $\theta|(x_1, \ldots, x_n) \sim \text{Beta}(s_n + 2, n - s_n + 2)$ where $s_n = \sum_{i=1}^n x_i$. Hence, with $n = 20$, the posterior distribution of θ is $\text{Beta}(s_{20} + 2, 20 - s_{20} + 2)$ As for (b), the predictive distribution is determined as follows

$$\mathbb{P}[X_{21} = 1 \,|\, x_1, \ldots, x_{20}] = \int_0^1 \mathbb{P}[X_{21} = 1 \,|\, x_1, \ldots, x_{20}, \theta]\, g(\theta|x_1, \ldots, x_{20})\, \mathrm{d}\theta$$

and since the X_i's are conditionally independent, given θ, one has

$$\mathbb{P}[X_{21} = 1 \,|\, x_1, \ldots, x_{20}, \theta] = \mathbb{P}[X_{21} = 1 \,|\, \theta].$$

Hence, the calculations boil down to

$$
\begin{aligned}
\mathbb{P}[X_{21} = 1 \,|\, x_1, \ldots, x_{20}] &= \int_0^1 \mathbb{P}[X_{21} = 1|\theta]\, g(\theta|x_1, \ldots, x_{20})\, \mathrm{d}\theta \\
&= \int_0^1 \theta \, \frac{\Gamma(24)}{\Gamma(s_{20}+2)\Gamma(20-s_{20}+2)}\, \theta^{s_{20}+1}(1-\theta)^{20-s_{20}+1}\, \mathrm{d}\theta \\
&= \frac{\Gamma(24)}{\Gamma(s_{20}+2)\Gamma(20-s_{20}+2)} \int_0^1 \theta^{s_{20}+2}(1-\theta)^{20-s_{20}+1}\, \mathrm{d}\theta \\
&= \frac{\Gamma(24)}{\Gamma(s_{20}+2)\Gamma(20-s_{20}+2)} \, \frac{\Gamma(s_{20}+3)\Gamma(20-s_{20}+2)}{\Gamma(25)} \\
&= \frac{s_{20}+2}{24}
\end{aligned}
$$

Finally, note that with $s_{20} = 10$, the posterior distribution of θ is $\mathrm{Beta}(12, 12)$ which is symmetric around $\theta = 0.5$ on the interval $[0, 1]$. Hence, one has

$$
\mathbb{P}[\theta < 0.5|x_1, \ldots, x_{20}] = 0.5,
$$

and this answers (c).

Exercise

Let $X_1 \,|\, \theta \overset{\text{iid}}{\sim} \mathrm{Binom}(10, \theta)$ and suppose $\theta \sim \mathrm{Unif}(0, 1)$, namely

$$
\mathbb{P}[X_1 = x \,|\, \theta] = \binom{10}{x} \theta^x (1-\theta)^{10-x}\, \mathbb{1}_{\{0,1,\ldots,10\}}(x), \qquad g(\theta) = \mathbb{1}_{[0,1]}(\theta).
$$

(a) Identify the posterior distribution of θ, given $X_1 = x_1$

(b) Determine the predictive distribution of X_2, given the $X_1 = x_1$, namely $\mathbb{P}[X_2 = 1 \,|\, X_1 = x_1]$

(c) If $x_1 = 6$, compute the Bayesian estimator of θ corresponding to a quadratic loss function.

Solution.

For any $x_1 \in \{0, 1, \ldots, 10\}$, the posterior distribution is obtained as

$$
g(\theta|x_1) \propto \binom{10}{x_1} \theta^{x_1} (1-\theta)^{10-x_1}\, \mathbb{1}_{(0,1)}(\theta)
$$

and one deduces that $\theta|x_1 \sim \mathrm{Beta}(x_1 + 1, 10 - x_1 + 1)$. The predictive distribution in (b) follows from the following calculation

$$
\begin{aligned}
\mathbb{P}[X_2 = 1 \mid X_1 = x_1] &= \int_0^1 \mathbb{P}[X_2 = 1 \mid X_1 = x_1,\, \theta]\, g(\theta|x_1)\, \mathrm{d}\theta \\
&= \int_0^1 \mathbb{P}[X_2 = 1 \mid \theta]\, g(\theta|x_1)\, \mathrm{d}\theta \\
&= \int_0^1 \binom{10}{1} \theta^1 (1-\theta)^9 \, \frac{\Gamma(12)}{\Gamma(x_1+1)\Gamma(10-x_1+1)}\, \theta^{x_1}(1-\theta)^{10-x_1}\, \mathrm{d}\theta \\
&= \frac{10\,\Gamma(12)}{\Gamma(x_1+1)\Gamma(10-x_1+1)} \int_0^1 \theta^{x_1+1}(1-\theta)^{19-x_1}\, \mathrm{d}\theta \\
&= \frac{10\,\Gamma(12)}{\Gamma(x_1+1)\Gamma(10-x_1+1)}\, \frac{\Gamma(x_1+2)\Gamma(20-x_1)}{\Gamma(22)} \\
&= \frac{10\,\Gamma(12)}{\Gamma(22)}\, \frac{\Gamma(20-x_1)}{\Gamma(11-x_1)}\, (x_1+1)
\end{aligned}
$$

Finally, conditioning on $x_1 = 6$, he posterior distribution of θ is $\text{Beta}(7, 4)$ and the Bayesian estimator of θ under squared loss is

$$
\hat{\theta} = \mathbb{E}[\theta|X_1 = 6] = \int_0^1 \theta\, \frac{\Gamma(11)}{\Gamma(7)\Gamma(4)} \theta^6 (1-\theta)^3\, \mathrm{d}\theta = \frac{7}{11}.
$$

> **Exercise**
> Let $X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim} f(\,\cdot\,|\theta)$ where
>
> $$f(x|\theta) = \theta\, \mathrm{e}^{-\theta x}\, \mathbb{1}_{(0,+\infty)}(x), \qquad \theta > 0.$$
>
> Suppose further that the prior distribution for the parameter θ is $g(\theta) = \mathrm{e}^{-\theta}\, \mathbb{1}_{(0,+\infty)}(\theta)$.
>
> (a) Identify the posterior distribution of θ. It is useful to remember that the gamma density with parameters (a, b) is
>
> $$g(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \mathrm{e}^{-b\theta}, \qquad \theta > 0.$$
>
> (b) Determine the Bayes estimator of θ under a squared loss function.
>
> (c) Assume that $n = 1$ (the sample is X_1), identify the predictive density $f(x_2|x_1)$.

Solution.
As for the posterior distribution, the usual calculations apply and one

has

$$g(\theta|x_1,\dots,x_n) \propto \left\{\prod_{i=1}^{n} \theta\, \mathrm{e}^{-\theta x_i}\right\} \mathrm{e}^{-\theta}\, \mathbb{1}_{(0,+\infty)}(\theta)$$
$$= \theta^n\, \mathrm{e}^{-(s_n+1)\,\theta}\, \mathbb{1}_{(0,+\infty)}(\theta)$$

where $s_n = \sum_{i=1}^{n} x_i$. Hence, $\theta|(x_1,\dots,x_n) \sim \text{Gamma}(n+1, s_n+1)$. As for the Bayesian estimator in (b), one has

$$\hat{\theta} = \mathbb{E}[\theta|x_1,\dots,x_n] = \int_0^{+\infty} \theta\, \frac{(s_n+1)^{n+1}}{\Gamma(n+1)}\, \theta^n\, \mathrm{e}^{-(s_n+1)\theta}\, \mathrm{d}\theta$$
$$= \frac{(s_n+1)^{n+1}}{n!} \int_0^{+\infty} \theta^{n+1}\, \mathrm{e}^{-(s_n+1)\theta}\, \mathrm{d}\theta$$
$$= \frac{(s_n+1)^{n+1}}{n!}\, \frac{\Gamma(n+2)}{(s_n+1)^{n+2}} = \frac{n+1}{s_n+1}$$

Finally, for any $x_1 > 0$, the predictive density is (recall that $s_1 = x_1$)

$$f(x_2|x_1) = \mathbb{1}_{(0,+\infty)}(x_2) \int_0^{+\infty} f(x_2|\theta)\, g(\theta|x_1)\, \mathrm{d}\theta$$
$$= \mathbb{1}_{(0,+\infty)}(x_2) \int_0^{+\infty} \theta\, \mathrm{e}^{-\theta x_2}\, \frac{(x_1+1)^2}{\Gamma(2)}\, \theta\, \mathrm{e}^{-(x_1+1)\theta}\, \mathrm{d}\theta$$
$$= (x_1+1)^2 \int_0^{+\infty} \theta^2\, \mathrm{e}^{-(s_2+1)\theta}\, \mathrm{d}\theta$$
$$= \mathbb{1}_{(0,+\infty)}(x_2)\, (x_1+1)^2\, \frac{\Gamma(3)}{(s_2+1)^3} = \frac{2(x_1+1)^2}{(s_2+1)^3}\, \mathbb{1}_{(0,+\infty)}(x_2)$$

where $s_2 = x_1 + x_2$.

Exercise 9.6.7 (p. 409)
Suppose $X_1,\dots,X_n|\theta \overset{\text{iid}}{\sim} \text{Unif}(0,\theta)$ and let $\theta \sim \text{Pareto}(\alpha, \theta_0)$, i.e. the prior probability density function is

$$g(\theta) = \frac{\alpha\, \theta_0^\alpha}{\theta^{\alpha+1}}\, \mathbb{1}_{(\theta_0,+\infty)}(\theta)$$

where it is assumed that $\theta_0 > 0$ and $\alpha > 0$ are known.

(a) Determine the posterior density function of θ, given a sample $X_1 = x_1,\dots,X_n = x_n$.

(b) When $n > 2$, determine the Bayesian estimator of θ, with squared error loss.

Solution.
Since the likelihood function is

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta}\, \mathbb{1}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\, \mathbb{1}_{(0,+\infty)}(x_{(1)})\, \mathbb{1}_{(-\infty,\theta)}(x_{(n)}) \\ &= \frac{1}{\theta^n}\, \mathbb{1}_{(x_{(n)},+\infty)}(\theta)\, \mathbb{1}_{(0,+\infty)}(x_{(1)})\end{aligned}$$

by Bayes' theorem the posterior probability density function of θ, given $x_1, \ldots, x_n$, is

$$\begin{aligned}g(\theta|x_1,\ldots,x_n) &\propto \frac{1}{\theta^n}\, \mathbb{1}_{(x_{(n)},+\infty)}(\theta)\, \frac{1}{\theta^{\alpha+1}}\, \mathbb{1}_{(\theta_0,+\infty)}(\theta) \\ &= \frac{1}{\theta^{\alpha+n+1}}\, \mathbb{1}_{(\max\{\theta_0, x_{(n)}\},+\infty)}(\theta)\end{aligned}$$

and this implies that $\theta|(x_1,\ldots,x_n) \sim \text{Pareto}(\alpha+n, \theta_n^*)$, where $\theta_n^* = \max\{\theta, x_{(n)}\}$. The Bayesian estimator under squared error loss is

$$\begin{aligned}\hat{\theta} &= \mathbb{E}[\theta|x_1,\ldots,x_n] = \int_{\theta_n^*}^{+\infty} \theta\, \frac{(\alpha+n)\,(\theta_n^*)^{\alpha+n}}{\theta^{\alpha+n-1}}\, \mathrm{d}\theta \\ &= (\alpha+n)\,(\theta_n^*)^{\alpha+n} \int_{\theta_n^*}^{+\infty} \frac{1}{\theta^{\alpha+n-2}}\, \mathrm{d}\theta \\ &= (\alpha+n)\,(\theta_n^*)^{\alpha+n}\, \frac{1}{\alpha+n-1}\, \frac{1}{(\theta_n^*)^{\alpha+n-1}} \\ &= \frac{\alpha+n}{\alpha+n-1}\, \theta_n^*\end{aligned}$$

Exercise.
Recall that a random variable Y has an inverse-gamma distribution, with parameters (α, β) if

$$f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \frac{1}{y^{\alpha+1}}\, \mathrm{e}^{-\frac{\beta}{y}}\, \mathbb{1}_{(0,+\infty)}(y)$$

and we use the notation $Y \sim \text{IG}(\alpha, \beta)$. This is equivalent to saying that $1/Y \sim \text{Gamma}(\alpha, \beta)$.
Let, now, $X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim} \text{Gamma}(2, 1/\theta)$ and suppose that $\theta \sim \text{IG}(1, \lambda)$ for some pre-specified $\lambda > 0$.

(a) Determine the posterior distribution of θ, given $X_1 = x_1, \ldots, X_n = x_n$.

(b) Evaluate the predictive probability density function of X_{n+1}, given observations $X_1 = x_1, \ldots, X_n = x_b$.

(c) Determine the Bayesian estimate of θ under squared error loss and identify its limiting value as $\lambda \to 0$.

Solution.
By Bayes' theorem

$$
\begin{aligned}
g(\theta|x_1,\dots,x_n) &\propto \frac{\lambda}{\theta^2}\,\mathrm{e}^{-\frac{\lambda}{\theta}}\,\prod_{i=1}^{n}\frac{1}{\theta^2}\,x_i\,\mathrm{e}^{-\frac{x_i}{\theta}} \\
&\propto \frac{1}{\theta^{2n+2}}\,\mathrm{e}^{\cdot\frac{1}{\theta}\,(\sum_{i=1}^{n}x_i+\lambda)}
\end{aligned}
$$

Hence $\theta|(x_1,\dots,x_n) \sim \mathrm{IG}(2n+1,\,\lambda+\sum_{i=1}^{n}x_i)$, namely

$$
g(\theta|x_1,\dots,x_n) = \frac{(\sum_{i=1}^{n}x_i+\lambda)^{2n+1}}{\Gamma(2n+1)}\,\frac{1}{\theta^{2n+2}}\,\mathrm{e}^{\cdot\frac{1}{\theta}\,(\sum_{i=1}^{n}x_i+\lambda)}\,\mathbb{1}_{(0,+\infty)}(\theta).
$$

Henceforth, we use the notation $s_n = \sum_{i=1}^{n}x_i$. The evaluation of the predictive density function is as follows

$$
\begin{aligned}
f(x_{n+1}|x_1,\dots,x_n) &= \int_0^{+\infty} f(x_{n+1}|\theta)\,g(\theta|x_1,\dots,x_n)\,\mathrm{d}\theta \\
&= \mathbb{1}_{(0,+\infty)}(x_{n+1})\,\int_0^{+\infty}\frac{1}{\theta^2}\,x_{n+1}\,\mathrm{e}^{-\frac{x_{n+1}}{\theta}} \\
&\quad\times\frac{(s_n+\lambda)^{2n+1}}{\Gamma(2n+1)}\,\frac{1}{\theta^{2n+2}}\,\mathrm{e}^{-\frac{1}{\theta}\,(s_n+\lambda)}\,\mathrm{d}\theta \\
&= \frac{(s_n+\lambda)^{2n+1}}{\Gamma(2n+1)}\,\frac{\Gamma(2n+3)}{(s_n+\lambda+x_{n+1})^{2n+3}}\,x_{n+1}\,\mathbb{1}_{(0,+\infty)}(x_{n+1}) \\
&= (2n+2)(2n+1)\frac{(s_n+\lambda)^{2n+1}\,x_{n+1}}{(s_n+\lambda+x_{n+1})^{2n+3}}\,\mathbb{1}_{(0,+\infty)}(x_{n+1})
\end{aligned}
$$

As for (c), in this case the Bayesian estimator is

$$
\begin{aligned}
\hat{\theta} &= \mathbb{E}[\theta|x_1,\dots,x_n] \\
&= \int_0^{+\infty}\theta\,\frac{(s_n+\lambda)^{2n+1}}{\Gamma(2n+1)}\,\frac{1}{\theta^{2n+2}}\,\mathrm{e}^{-\frac{1}{\theta}(s_n+\lambda)}\,\mathrm{d}\theta \\
&= \frac{(s_n+\lambda)^{2n+1}}{\Gamma(2n+1)}\,\frac{\Gamma(2n)}{(s_n+\lambda)^{2n}} \\
&= \frac{s_n+\lambda}{2n}.
\end{aligned}
$$

Finally,

$$
\lim_{\lambda\to 0}\hat{\theta} = \frac{s_n}{2n}
$$

and it can be shown that such a limit is the (frequentist) MLE of θ (check!).

Exercise

Suppose a survey question asks to choose among three possibile answers: "Yes", "Maybe" and "No". A sample of n respondents is considered and to each respondent a random vector $\boldsymbol{X}_i = (X_{1,i}, X_{2,i})$ taking values in $\mathbb{X} = \{(1,0),(0,1),(0,0))\}$, is associated. Here, $(1,0)$ means that the answer is "Yes", $(0,1)$ is "Maybe" and $(0,0)$ is "No". Is is assumed that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n | (p_1, p_2) \overset{\text{iid}}{\sim}$ $\text{Mult}_3(1; p_1, p_2)$, with $p_1, p_2 \in (0,1)$ are such that $p_1 + p_2 \leq 1$. This means that the probability mass function of each $\boldsymbol{X}_i$ is

$$f(x_1, x_2|\theta) = p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{1-x_1-x_2} \mathbb{1}_{\mathbb{X}}(x_1, x_2).$$

For the vector of probabilities (p_1, p_2), the following prior is specified

$$g(p_1, p_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \\ \times (1 - p_2 - p_2)^{\alpha_3 - 1} \mathbb{1}_{\Delta_2}(p_1, p_2)$$

where $\Delta_2 = \{(p_1, p_2) \in [0,1]^2 : \; p_1 + p_2 \leq 1\}$ is the 2-dimensional simplex. Such a prior is know as Dirichlet distribution with parameters $(\alpha_1, \alpha_2, \alpha_3)$, with $\alpha_1, \alpha_2 \alpha_3 > 0$.

(a) Determine the posterior distribution of (p_1, p_2), given $X_1 = x_1, \ldots, X_n = x_n$.

(b) Determine the probability that the $(n+1)$–the respondent will opt for a "Yes", given the responses of the previous n, namely

$$\mathbb{P}[\boldsymbol{X}_{n+1} = (1,0) \,|\, X_1 = x_1, \ldots, X_n = x_n]$$

Solution.

Let us denote as $s_{1,n} = \sum_{i=1}^n x_{1,i}$ the number of "Yes" answers, $s_{2,n} = \sum_{i=1}^n x_{2,i}$ the number of "Maybe" answers and $s_{3,n} = n - s_{1,n} - s_{2,n}$ the number of "No"s. The posterior probability density function is

$$\begin{aligned} g(p_1, p_2 | x_1, \ldots, x_n) &\propto p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} (1 - p_1 - p_2)^{\alpha_3 - 1} \mathbb{1}_{\Delta_2}(p_1, p_2) \\ &\quad \times \prod_{i=1}^n p_1^{x_{1,i}} \, p_2^{x_{2,i}} \, (1 - p_1 - p_2)^{1 - x_{1,i} - x_{2,i}} \\ &= p_1^{\alpha_1 + s_{1,n} - 1} p_2^{\alpha_2 + s_{2,n} - 1} (1 - p_1 - p_2)^{\alpha_3 + s_{3,n} - 1} \mathbb{1}_{\Delta_2}(p_1, p_2) \end{aligned}$$

and it turns out that the posterior distribution of the vector (p_1, p_2) is again Dirichlet with parameters $(\alpha_1 + s_{1,n}, \alpha_2 + s_{2,n}, \alpha_3 + s_{3,n})$, namely the posterior density is

$$g(p_1, p_2|x_1, \dots, x_n) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3 + n)}{\Gamma(\alpha_1 + s_{1,n})\Gamma(\alpha_2 + s_{2,n})\Gamma(\alpha_3 + s_{3,n})} \times p_1^{\alpha_1 + s_{1,n} - 1} p_2^{\alpha_2 + s_{2,n} - 1} (1 - p_1 - p_2)^{\alpha_3 + s_{3,n} - 1} \mathbb{1}_{\Delta_2}(p_1, p_2)$$

Note that the both the prior and the posterior distributions are Dirichlet. The posterior is obtained by updating the parameters of the prior through the frequency of observations in each of the 3 categories, i.e. the number of "Yes", "Maybe" and "No" answers. The determination of the predictive probability in (b) easily follows if one notes that the Dirichlet distribution shares the property below

$$(w_1, w_2) \sim \text{Dirichlet}(c_1, c_2, c_3) \quad \Longrightarrow \quad w_1 \sim \text{Beta}(c_1, c_2 + c_3)$$

$$(w_1, w_2) \sim \text{Dirichlet}(c_1, c_2, c_3) \quad \Longrightarrow \quad w_2 \sim \text{Beta}(c_2, c_1 + c_3)$$

Since $(p_1, p_2)|(x_1, \dots, x_n) \sim \text{Dirichlet}(\alpha_1 + s_{1,n}, \alpha_2 + s_{2,n}, \alpha_3 + s_{3,n})$, then $p_1|(x_1, \dots, x_n) \sim \text{Beta}(\alpha_1 + s_{1,n}, \alpha_2 + \alpha_3 + s_{2,n} + s_{3,n})$. When it comes to evaluating the predictive probability in (b), one has

$$\begin{aligned}
&\mathbb{P}[\boldsymbol{X}_{n+1} = (1, 0) \,|\, X_1 = x_1, \dots, X_n = x_n] \\
&= \int_0^1 p_1 \, g(p_1|x_1, \dots, x_n) \, \mathrm{d}p_1 \\
&= \int_0^1 p_1 \, \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3 + n)}{\Gamma(\alpha_1 + s_{1,n})\Gamma(\alpha_2 + +\alpha_3 + s_{2,n} + s_{3,n})} \\
&\qquad \times p_1^{\alpha_1 + s_{1,n} - 1} (1 - p_1)^{\alpha_2 + \alpha_3 + s_{2,n} + s_{3,n} - 1} \, \mathrm{d}p_1 \\
&= \frac{\alpha_1 + s_{1,n}}{\alpha_1 + \alpha_2 + \alpha_3 + n}
\end{aligned}$$

where we have used the fact that $s_{1,n} + s_{2,n} + s_{3,n} = n$. As a side note, it is interesting to point out that such a predictive probability has a nice intuitive structure. Indeed, setting $\bar{\alpha} = \alpha_1 + \alpha_2 + \alpha_3$ and noting that

$$\frac{\alpha_1}{\bar{\alpha}} = \mathbb{E}p_1 = \int_{\Delta_2} g(p_1, p_2) \, \mathrm{d}p_1, \mathrm{d}p_2, \qquad \frac{s_{1,n}}{n} \hat{p}_1 = \text{ MLE of } p_1$$

one has

$$\begin{aligned}
&\mathbb{P}[\boldsymbol{X}_{n+1} = (1, 0) \,|\, X_1 = x_1, \dots, X_n = x_n] \\
&= \frac{\bar{\alpha}}{\bar{\alpha} + n} \, \frac{\alpha_1}{\bar{\alpha}} + \frac{n}{\bar{\alpha} + n} \, \frac{s_{1,n}}{n} \\
&= \frac{\bar{\alpha}}{\bar{\alpha} + n} \, \mathbb{E}p_1 + \frac{n}{\bar{\alpha} + n} \, \hat{p}_1
\end{aligned}$$

and the predictive distribution is a linear combination of the prior guess at the value of p_1, i.e. $\mathbb{E}p_1$, and a frequentist estimator of p_1, namely the MLE $\hat{p}_1$.

Exercise
Let $X_1, X_2 \overset{\text{iid}}{\sim} \text{Unif}(\theta, 1)$, with $\theta \in (0,1)$. Hence, the density function of each X_i is

$$f_{X_i}(x) = \frac{1}{1-\theta}\,\mathbb{1}_{(\theta,1)}(x) \qquad i = 1, 2.$$

(a) For testing $H_0 : \theta = 0.2$ vs $H_1 : \theta = 0.8$, a procedure based only on X_1 has been proposed and is such that

$$\text{reject } H_0 \text{ whenever } X_1 > 0.96$$

Determine probabilities of type I and type II error associated to such a test.

(b) What is the cumulative distribution function of $X_{(1)} = \min\{X_1, X_2\}$ both under H_0 and under H_1?

(c) Suppose now that for testing $H_0 : \theta = 0.2$ vs $H_1 : \theta = 0.8$, one uses an alternative decision rule according to which

$$\text{reject } H_0 \text{ whenever } X_{(1)} > 1 - 0.8\sqrt{0.05}$$

Would one prefer it to the test in (a)? Motivate the answer.

Solution.
Since

$$F_{X_1}(x|\theta) = P_\theta[X_1 \le x] = \frac{x-\theta}{1-\theta}\,\mathbb{1}_{(\theta,1)}(x) + \mathbb{1}_{[1,+\infty)}(x),$$

it is easily seen that hte probabilities of type I and type II error are

$$\alpha = \mathbb{P}_{\theta=0.2}[X_1 > 0.96] = 1 - F_{X_1}(0.96\,|\,0.2) = 1 - \frac{0.96-0.2}{1-0.2}$$

$$= \frac{0.04}{0.8} = 0.05$$

$$\beta = \mathbb{P}_{\theta=0.8}[X_1 \le 0.96] = F_{X_1}(0.96\,|\,0.8) = \frac{0.96-0.8}{1-0.8} = \frac{0.16}{0.2} = 0.8$$

Since for any $n \ge 1$, the cumulative distribution function of $X_{(1)} = \min\{X_1, \ldots, X_n\}$ is

$$F_{X_{(1)}}(x|\theta) = 1 - \Big\{1 - F_{X_1}(x|\theta)\Big\}^n.$$

In our case, $n = 2$ and, hence,

$$F_{X_{(1)}}(x|0.2) = 1 - \left\{1 - \frac{x-0.2}{0.8}\,\mathbb{1}_{(0.2,1)}(x) + \mathbb{1}_{[1,+\infty)}(x)\right\}^2$$

$$F_{X_{(1)}}(x|0.8) = 1 - \left\{1 - \frac{x-0.8}{0.2}\,\mathbb{1}_{(0.8,1)}(x) + \mathbb{1}_{[1,+\infty)}(x)\right\}^2$$

and this answers (b). One can now rely on this findings to solve (c) and note that

$$\alpha = \mathbb{P}_{\theta=0.2}[X_{(1)} > 1 - 0.8\sqrt{0.05}] = \left\{1 - \frac{1-0.8\sqrt{0.05}-0.2}{0.8}\right\}^2$$
$$= 0.05$$
$$\beta = \mathbb{P}_{\theta=0.8}[X_{(1)} \le 1 - 0.8\sqrt{0.05}] = 1 - \left\{1 - \frac{1-0.8\sqrt{0.05}-0.8}{0.2}\right\}^2$$
$$= 1 - 16 \times 0.05 = 0.2$$

Hence this test is preferred over the previous one: it has the same probability of type I error ($\alpha = 0.05$) and has lower probability of type II error.

Exercise

Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{E}(\theta)$, namely

$$f(x|\theta) = \theta\, \mathrm{e}^{-\theta x}\, \mathbb{1}_{\mathbb{R}^+}(x).$$

(a) A sample of size $n = 1$ is used for testing $H_0 : \theta = 1$ vs $H_1 : \theta = 0.5$ and the following decision rule has been proposed:

$$\text{reject } H_0 \text{ when } X_1 > 3.5$$

Determine the probabilities of type 1 and type 2 errors.

(b) Still with a sample of size $n = 1$, identify the Neyman–Pearson test of size $\alpha = 0.05$ for the same problem as in (a), namely $H_0 : \theta = 1$ vs $H_1 : \theta = 0.5$

(c) Compare the power functions of the tests in (a) and in (b) and explain why the test in (b) is preferable to the test in (a).

Solution.

Since $F_{X_1}(x|\theta) = (1 - \mathrm{e}^{-\theta x})\, \mathbb{1}_{(0,+\infty)}(x)$, one has

$$\alpha = P_{\theta=1}\left[X_1 > 3.5\right] = \mathrm{e}^{-3.5} \approx 0.03$$
$$\beta = P_{\theta=0.5}\left[X_1 \le 3.5\right] = 1 - \mathrm{e}^{-3.5/2} \approx 0.8262$$

As for (b), the Neyman–Pearson test has critical region

$$\text{Reject } H_0 \text{ when } \frac{0.5\,\mathrm{e}^{-0.5x_1}}{\mathrm{e}^{-x_1}} > k.$$

Since the likelihood ratio above is monotone increasing in x_1, then the critical region becomes

$$\text{Reject } H_0 \text{ when } x_1 > k'$$

where k is such that

$$P_{\theta=1}\left[X_1 > k'\right] = 0.05$$

Hence, $e^{-k'} = 0.05$, namely $k' = -\log(0.05) \approx 2.996$.

The power of the test in (b) equals

$$p(0.5) = P_{\theta=0.5}\left[X_1 > 2.996\right] \approx e^{-1.5} \approx 0.2231.$$

On the other hand, the test in (a) has $p(0.5) = e^{-1.75} \approx 0.1737$ and, as expected, its power is lower than that of the Neyman–Pearson test in (b).

Exercise

Suppose X_1 ius a single observation recorded from the density function

$$f(x|\theta) = \{2(1-\theta)\,x + \theta\}\,\mathbb{1}_{(0,1)}(x).$$

and one wants to test $H_0 :\ \theta = 1/2$ vs $H_1 : \theta = 3/2$.

(a) Determine the most powerful test of size α.

(b) Evaluate the probability β of making a type II error for such a test.

Solution.

Notice that in this case the likelihood ratio, with a single observation x_1, is

$$\frac{\mathcal{L}(3/2)}{\mathcal{L}(1/2)} = \frac{\frac{3}{2} - x_1}{\frac{1}{2} + x_1} = \frac{3 - 2x_1}{1 + 2x_1}$$

and by the Neyman-Pearson lemma the optimal test is

$$\text{Reject}\ \ H_0\ \ \text{when}\ \ \frac{3 - 2x_1}{3 + 2x_1} > k$$

which amounts to rejecting H_0 when $x_1 < k'$ and k' is such that

$$P_{\theta=1/2}[X_1 < k'] = \alpha.$$

Using the fact that $f(x|1/2) = (1/2) + x\,\mathbb{1}_{(0,1)}(x)$, one has

$$\alpha = P_{\theta=1/2}[X_1 < k'] = \int_0^{k'} \{(1/2) + x\}\,\mathrm{d}x = \frac{k'}{2} + \frac{(k')^2}{2}$$

A (positive) solution of the equation is $k' = -1 + \sqrt{1 + 8\alpha}$ and this identifies the optimal test. As for the evaluation of the probability of type II error, note that

$$\beta = P_{\theta=1/2}[X_1 \geq k'] = \int_{k'}^{1} \left\{\frac{3}{2} - x\right\} \mathrm{d}x = \frac{3}{2} - \frac{1}{2} - \frac{3}{2}k' + \frac{(k')^2}{2}$$
$$= 1 + \alpha - \frac{1}{2}(-1 + \sqrt{1+8\alpha}) = \frac{3}{2} + \alpha - \frac{2}{2}\sqrt{1+8\alpha}$$

Exercise
Let X_1 br a single observation from the probability density function

$$f(x!\theta) = 2\theta x(1-x^2)^{\theta-1}\,\mathbb{1}_{(0,1)}(x).$$

(a) Determine the optimal test of size α for $H_0 : \theta = 1$ vs $H_1 : \theta = 10$.

(b) Evaluate the probability β of making a type II error for such a test.

(c) What is the probability of making the correct decision is $\theta = 10$ and $\alpha = 0.04$?

Solution.
The likelihood ratio

$$\frac{\mathcal{L}(10; x_1)}{\mathcal{L}(1; x_1)} = \frac{20x_1(1-x_1^2)^9}{2x_1} = 10(1-x_1^2)^9$$

is a decreasing function of x_1 in $(0,1)$. Hence, the most powerful test is of the type

$$\text{Reject } H_0 \text{ if } X_1 \leq k_\alpha$$

where k_α is such that $\mathbb{P}_{\theta=1}[X_1 \leq k_\alpha] = \alpha$. Hence, k_α is determined by noting that

$$\mathbb{P}_{\theta=1}[X_1 \leq k_\alpha] = \int_0^{k_\alpha} 2x\,\mathrm{d}x = k_\alpha^2$$

and, hence, the equation $k_\alpha^2 = \alpha$ yields $k_\alpha = \sqrt{\alpha}$. As for (b), one has

$$\beta = \mathbb{P}_{\theta=10}[X_1 > \sqrt{\alpha}] = \int_{\sqrt{\alpha}}^{1} 20x(1-x^2)^9\,\mathrm{d}x = (1-\alpha)^{10}.$$

Finally, as for (c), we see that $\alpha = 0.04$ yields the optimal test that rejects H_0 is $X_1 \leq 0.2$. Hence, the probability of taking the correct decision when $\theta = 10$ is

$$\mathbb{P}_{\theta=10}[X_1 \leq 0.2] = 1 - \beta \int_0^{0.2} 20x(1-x^2)^9\,\mathrm{d}x = 1 - \beta = 1 - (0.96)^{10}$$
$$\approx 0.335$$

and this is also known as the power of the test.

Exercise
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\,\cdot\,|\theta)$ where

$$f(x|\theta) = \frac{1}{\theta}\, x^{\frac{1}{\theta}-1}\, \mathbb{1}_{(0,1)}(x).$$

(a) Determine the density function of $-\sum_{j=1}^{n} \log X_j$

(b) Determine the optimal test for $H_0 : \theta = 2$ vs $H_1 : \theta = 0.2$

(c) Determine the probability of type II error a

Solution.
Note that the density of $Y = h(X) = -\log X$ is obtained through the standard change of variable technique and one has

$$f_Y(y) = f(h^{-1}(y)|\theta)\, |\mathrm{d}h^{-1}(y)/\mathrm{d}y| = \frac{1}{\theta}\, \mathrm{e}^{-\frac{y}{\theta}}\, \mathbb{1}_{(0,+\infty)}(y)$$

so that $-logX \sim$ neg-exp$(1/\theta)$. This implies that $-\sum_{j=1}^{n} \log X_j$ is the sum of iid negative exponential random variables with parameter $1/\theta$ and it is known that it is a Gamma distribution with parameters $(n,\, 1/\theta)$. Its probability density function is

$$f^*(x|\theta) = \frac{1}{\theta^n\, \Gamma(n)}\, x^{n-1}\, \mathrm{e}^{-\frac{x}{\theta}}\, \mathbb{1}_{(0,+\infty)}(x).$$

As far as (b) is concerned, note that for any $x_1, \ldots, x_n$ in $(0,1)$ the likelihood ratio is

$$\frac{\mathcal{L}(0.2; x_1 \ldots, x_n)}{\mathcal{L}(2; x_1, \ldots, x_n)} = \frac{5\, \prod_{i=1}^{n} x_i^4}{0.5\, \prod_{i=1}^{n} x_i^{-0.5}} = 10 \prod_{i=1}^{n} x_i^{3.5}$$

Note that this is an increasing function of $\prod_{i=1}^{n} x_i$ or, equivalently, a decreasing function of $-\sum_{i=1}^{n} \log X_i$. Hence, the most powerful test rejects H_0 when $-\sum_{i=1}^{n} \log X_i < k_\alpha$, where k_α is such that

$$\mathbb{P}_{\theta=2}[-\sum_{i=1}^{n} \log X_i \le k_\alpha] = \alpha$$

Since with $\theta = 2$ the distribution of $-\sum_{i=1}^{n} \log X_i$ is Gamma$(n, 1/2)$, this is also a χ^2_{2n} distribution and, hence, from the previous equation $k_\alpha = \chi^2_{2n,1-\alpha}$, namely is the quantile of order $1-\alpha$ of a χ^2 distribution with $2n$ degrees of freedom.

As for the probability of type II error in (c), for $\theta = 0.2$ the random variable $-\sum_{i=1}^{n} \log X_i$ has a Gamma distribution with parameters

$(n, 5)$. By virtue of the scaling property of the gamma distribution, $-10\sum_{i=1}^{n} \log X_i$ has a Gamma distribution with parmeters $(n, 1/2)$, namely it is χ^2 with $2n$ degrees of freedom. Hence,

$$\begin{aligned}\beta &= \mathbb{P}_{\theta=0.2}[-\sum_{i=1}^{n} \log X_i \geq \xi^2_{2n,1-\alpha}] \\ &= \mathbb{P}_{\theta=0.2}[-10\sum_{i=1}^{n} \log X_i \geq 10\xi^2_{2n,1-\alpha}] \\ &= 1 - F(10\,\chi^2_{2n,1-\alpha})\end{aligned}$$

where F is the cumulative distribution function of the χ^2 distribution with $2n$ degrees of freedom.

Exercise

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$ and suppose one wishes to test $H_0 : \theta = 1$ vs $H_1 : \theta = 2$.

(a) Determine the optimal test of size α.

(b) Evaluate the probability β of making a type II error for such a test.

Solution.

Note that the likelihood ratio is

$$\frac{\mathcal{L}(2; x_1, \ldots, x_n)}{\mathcal{L}(1; x_1, \ldots, x_n)} = \frac{\frac{1}{2^n}\, \mathbb{1}_{(0,+\infty)}(x_{(1)}), \mathbb{1}_{(-\infty,2)}(x_{(n)})}{\mathbb{1}_{(0,+\infty)}(x_{(1)}), \mathbb{1}_{(-\infty,1)}(x_{(n)})}$$

one has

$$\text{likelihood ratio } = \begin{cases} \frac{1}{2^n} & \text{if } 0 < x_{(n)} < 1 \\ +\infty & \text{if } 1 \leq x_{(n)} < 2 \end{cases}$$

This is an increasing function of the sufficient statistics $x_{(n)}$. It turns out that the UMP test of size α is the one that rejects H_0 whenever $x_{(k)} > k$ where k is such that

$$P_{\theta=1}[X_{(n)} > k] = \alpha.$$

Since in this case $F_{X_{(n)}}(x|\theta) = (x/\theta)^n\, \mathbb{1}_{(0,\theta)}(x) + \mathbb{1}_{[\theta,+\infty)}(x)$, k is determined by the following equation

$$1 - F_{X_{(n)}}(k|\theta = 1) = 1 - k^n = \alpha \quad \Longrightarrow \quad k = (1-\alpha)^{1/n}$$

From this one can deduce the probability of type II error, which is

$$\begin{aligned}\beta &= P_{\theta=2}[X_{(n)} \leq (1-\alpha)^{1/n}] = F_{X_{(n)}}((1-\alpha)^{1/n}|\theta = 2) \\ &= \frac{1-\alpha}{2^n}\end{aligned}$$

Exercise

Let $X \sim \text{Beta}(1, \theta)$ and suppose one wishes to test $H_0 : \theta = 2$ vs $H_1 : \theta = 3$. Show that the best test with size $\alpha = 0.0975$ is the one which rejects H_0 if $X < 0.05$.

Solution.

Note that for any the likelihood ratio

$$\frac{\mathcal{L}(3; x)}{\mathcal{L}(2; x)} = \frac{3(1-x)^2}{2(1-x)} = \frac{3}{2}(1-x)$$

is a decreasing function of x in $(0, 1)$. Hence, the optimal test of size α rejects H_0 if $X < k$, where k is such that

$$P_{\theta=2}[X < k] = \alpha.$$

Note that $F_X(x|\theta) = \int_0^x \theta(1-s)^{\theta-1}\,\mathrm{d}s = 1 - (1-x)^\theta$ for any $x \in [0, 1)$ and $F_X(x|\theta) = 1$ for any $x \geq 1$. Hence,

$$P_{\theta=2}[X < k] = F_X(k|\theta = 2) = 1 - (1-k)^2 = \alpha$$

and the solution is $k = 1 - (1-\alpha)^{1/2}$. Hence, the optimal test of size α rejects H_0 if $X < 1 - (1-\alpha)^{1/2}$. Hence, is $\alpha = 0.0975$, one has $k = 1 - (0.9025)^{1/2} = 0.05$.