Probability and Statistics

Exercise sheet 10 - Solutions

MC 10.1. Let $n \in \mathbb{N}$ and let $X_1, \dots, X_n$ be i.i.d. and standard normally distributed, i.e., $X_i \sim \mathcal{N}(0,1)$. Define

$$Y \coloneqq \sum_{i=1}^{n} X_i^2.$$

In particular, Y is a χ_n^2-distributed random variable. (Exactly one answer is correct in each question.)

1. What is the value of $\mathbb{E}[Y]$?
   (a) $\mathbb{E}[Y] = 0$.
   (b) $\mathbb{E}[Y] = n^2$.
   (c) $\mathbb{E}[Y] = n$.
   (d) $\mathbb{E}[Y] = \sqrt{n}$.
2. What is the value of $\mathrm{Var}[Y]$?
   (a) $\mathrm{Var}[Y] = n^2$.
   (b) $\mathrm{Var}[Y] = 2n$.
   (c) $\mathrm{Var}[Y] = n$.
   (d) $\mathrm{Var}[Y] = 2n^2$.
3. Let now $n = 12$. What is the approximation of the probability $\mathbb{P}\big[\,\big|\frac{Y}{n} - 1\big| \leq 0.75\big]$ using the CLT?
   (a) $\mathbb{P}\big[\,\big|\frac{Y}{n} - 1\big| \leq 0.75\big] \approx 2\Phi\left(\frac{3}{4}\sqrt{6}\right) - 1$.
   (b) $\mathbb{P}\big[\,\big|\frac{Y}{n} - 1\big| \leq 0.75\big] \approx 2\Phi\left(\frac{7}{4}\sqrt{6}\right)$.
   (c) $\mathbb{P}\big[\,\big|\frac{Y}{n} - 1\big| \leq 0.75\big] \approx \Phi\left(\sqrt{\frac{7}{4}}\right)$.
   (d) $\mathbb{P}\big[\,\big|\frac{Y}{n} - 1\big| \leq 0.75\big] \approx 1 - 2\Phi\left(\sqrt{6}\right)$.

Solution:

(i) (c). From the definition of Y and linearity of expectation we have

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} 1 = n.$$

(ii) (b). Using that $(X_n)_{n\in\mathbb{N}}$ are i.i.d., we have $\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] = (\mathbb{E}[X_1^2])^2$ for any $i \neq j$. It follows

$$\mathbb{E}[Y^2] = \mathbb{E}[(X_1^2 + X_2^2 + \cdots + X_n^2)^2] = \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i^2 X_j^2\Big] = n\mathbb{E}[X_1^4] + n(n-1)(\mathbb{E}[X_1^2])^2.$$

Using again $\mathbb{E}[X_1^2] = 1$ and

$$\begin{aligned}\mathbb{E}[X_1^4] &= \int_{-\infty}^{\infty} x^4 \varphi(x) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2} \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \big(- x^3 e^{-x^2/2} \big) \Big|_{x=-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \mathrm{d}x \\ &= 0 + 3 \times \mathbb{E}[X_1^2] \\ &= 3,\end{aligned}$$

where we have used integration by parts, we obtain

$$\mathbb{E}[Y^2] = 3n + n(n-1) = n^2 + 2n,$$

and thus

$$\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = (n^2 + 2n) - n^2 = 2n.$$

Alternative solution: Since $X_1, \ldots, X_n$ are independent, so are $X_1^2, \ldots, X_n^2$, and so we have

$$\mathrm{Var}[Y] = \sum_{i=1}^{n} \mathrm{Var}[X_i^2] = n \times \mathrm{Var}[X_1^2].$$

As above,

$$\mathrm{Var}[X_1^2] = \mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2 = 3 - 1 = 2,$$

and so

$$\mathrm{Var}[Y] = 2n.$$

(iii) (a). Let $Z_i \coloneqq X_i^2$ for $i \in \{1, \ldots, n\}$. Then $Y = \sum_{i=1}^{n} Z_i$ with $(Z_n)_{n \in \mathbb{N}}$ i.i.d. and $Z_i \sim \chi_1^2$, and so $\mathbb{E}[Z_i] = 1$ and $\mathrm{Var}[Z_i] = 2$. By the central limit theorem:

$$\begin{aligned}\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right] &= \mathbb{P}\left[\left|\frac{Y-n}{n}\right| \leq \frac{3}{4}\right] \\ &= \mathbb{P}\left[\left|\frac{Y-n}{\sqrt{2n}}\right| \leq \frac{3}{4}\sqrt{\frac{n}{2}}\right] \\ &= \mathbb{P}\left[-\frac{3}{4}\sqrt{\frac{n}{2}} \leq \frac{Y-n}{\sqrt{2n}} \leq \frac{3}{4}\sqrt{\frac{n}{2}}\right] \\ &\approx \Phi\left(\frac{3}{4}\sqrt{\frac{n}{2}}\right) - \Phi\left(-\frac{3}{4}\sqrt{\frac{n}{2}}\right) \\ &= 2\Phi\left(\frac{3}{4}\sqrt{\frac{n}{2}}\right) - 1.\end{aligned}$$

For $n = 12$, we thus have

$$\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right] \approx 2\Phi\left(\frac{3}{4}\sqrt{6}\right) - 1.$$

Exercise 10.2. Let U_1, U_2, U_3 be i.i.d. random variables uniformly distributed on $[0,1]$. We consider the random variables

$$L := \min\{U_1, U_2, U_3\} \quad \text{and} \quad M := \max\{U_1, U_2, U_3\}.$$

(a) Show that M and L have densities and find them.

(b) Show that for $\phi, \psi : \mathbb{R} \to \mathbb{R}$ piecewise continuous and bounded, the following holds:

$$\mathbb{E}[\phi(M)\psi(L)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(m) \cdot \psi(\ell) \cdot 6(m-\ell)\mathbf{1}_{\{0\le \ell \le m \le 1\}} \mathrm{d}\ell \mathrm{d}m.$$

(c) Use (b) to determine the joint distribution function and the joint density of (M, L).

Solution:

(a) We first note that $\mathbb{P}[M \in [0,1]] = 1$. For $m \in [0,1]$, using the independence and uniformity of U_1, U_2, and U_3, we have:

$$\mathbb{P}[M \le m] = \mathbb{P}[U_1 \le m, U_2 \le m, U_3 \le m] = \big(\mathbb{P}[U_1 \le m]\big)^3 = m^3.$$

So the distribution function F_M of M is given by

$$F_M(m) = \begin{cases} 0 & \text{for } m < 0, \\ m^3 & \text{for } 0 \le m \le 1, \\ 1 & \text{for } m > 1. \end{cases}$$

We see that F_M is continuous and piecewise continuously differentiable. Hence, the density exists and for $m \in (0,1)$:

$$f_M(m) = \frac{\mathrm{d}}{\mathrm{d}m} F_M(m) = 3m^2.$$

We thus obtain the density

$$f_M(m) = \begin{cases} 0 & \text{for } m < 0, \\ 3m^2 & \text{for } 0 \le m \le 1, \\ 0 & \text{for } m > 1. \end{cases}$$

To find the density of L, we first note that we have

$$\mathbb{P}[L \ge \ell] = \mathbb{P}[U_1 \ge \ell, U_2 \ge \ell, U_3 \ge \ell] = \big(\mathbb{P}[U_1 \ge \ell]\big)^3 = (1-\ell)^3.$$

Consequently, the distribution function F_L of L is given by

$$F_L(\ell) = \begin{cases} 0 & \text{for } \ell < 0, \\ 1-(1-\ell)^3 & \text{for } 0 \le \ell \le 1, \\ 1 & \text{for } \ell > 1. \end{cases}$$

Thus, analogously as before, we obtain that the density exists and is given by

$$f_L(\ell) = 3(1-\ell)^2 \mathbf{1}_{\{\ell \in [0,1]\}}.$$

(b) Let $\phi, \psi : \mathbb{R} \to \mathbb{R}$ be piecewise continuous and bounded. Using the joint density of U_1, U_2, U_3 we compute

$$\mathbb{E}[\phi(M) \cdot \psi(L)] = \int_0^1 \int_0^1 \int_0^1 \phi(\max\{u_1, u_2, u_3\})\psi(\min\{u_1, u_2, u_3\})\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3.$$

We distinguish cases depending on which variable is the maximum and which is the minimum. For the case $u_3 \le u_2 \le u_1$ (i.e. $m = u_1$ and $\ell = u_3$) we have:

$$\begin{aligned}&\int_0^1 \int_0^1 \int_0^1 \phi(u_1)\psi(u_3)\mathbf{1}_{\{u_3 \le u_2 \le u_1\}}\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3 \\ &= \int_0^1 \phi(u_1)\Big(\int_0^{u_1} \psi(u_3)\Big(\int_{u_3}^{u_1} \mathrm{d}u_2\Big)\mathrm{d}u_3\Big)\mathrm{d}u_1 \\ &= \int_0^1 \int_0^1 \phi(u_1)\psi(u_3)(u_1 - u_3)\mathbf{1}_{\{u_3 \le u_1\}}\mathrm{d}u_1\mathrm{d}u_3\end{aligned}$$

Since there are in total $3! = 6$ options how u_1, u_2, u_3 can be ordered ($u_3 \le u_2 \le u_1$, $u_2 \le u_3 \le u_1$, $u_2 \le u_2 \le u_3$, $\dots$), which are symmetric, we obtain

$$\mathbb{E}[\phi(M)\psi(L)] = 6 \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(m)\psi(\ell)(m - \ell)\mathbf{1}_{\{0 \le \ell \le m \le 1\}}\mathrm{d}m\mathrm{d}\ell,$$

as required.

Remark: Note that, due to continuity of the uniform distribution and independence, we have $\mathbb{P}[U_1 = U_2] = \mathbb{P}[U_1 = U_3] = \mathbb{P}[U_2 = U_3] = 0$, which can be verified by straightforward computations. We thus do not care that the cases $u_1 = u_2 = u_3, u_2 = u_3 < u_1$, etc..., are counted multiple times.

(c) For $a, b \in \mathbb{R}$, choose $\phi(x) \coloneqq \mathbf{1}_{\{x \le a\}}$ and $\psi(x) \coloneqq \mathbf{1}_{\{x \le b\}}$ to get

$$\begin{aligned}F_{M,L}(a, b) &= \mathbb{P}[M \le a, L \le b] = \mathbb{E}[\mathbf{1}_{\{M \le a\}}\mathbf{1}_{\{L \le b\}}] \\ &= \int_{-\infty}^{a} \int_{-\infty}^{b} 6(m - \ell)\mathbf{1}_{\{0 \le \ell \le m \le 1\}}\mathrm{d}m\mathrm{d}\ell\end{aligned}$$

So the joint density is

$$f_{M,L}(m, \ell) = 6(m - \ell)\mathbf{1}_{\{0 \le \ell \le m \le 1\}}.$$

To compute the joint distribution function, we evaluate the integral above.

- For $a \le 0$ or $b \le 0$, we have $F_{M,L}(a, b) = 0$.
- For $b \ge 1$ and $a \in [0, 1]$, $F_{M,L}(a, b) = F_M(a) = a^3$.
- For $a \ge 1$ and $b \in [0, 1]$, $F_{M,L}(a, b) = F_L(b) = 1 - (1 - b)^3$.
- For $0 \le a \le b \le 1$, $F_{M,L}(a, b) = \mathbb{P}[M \le a] = a^3$.

- For $0 \leq b \leq a \leq 1$,

$$\begin{aligned}
\mathbb{P}[M \leq a, L \leq b] &= \int_0^a \left(\int_0^{\min\{b,m\}} 6(m-\ell)\mathrm{d}\ell \right) \mathrm{d}m \\
&= \int_0^a 6m\ell - 3\ell^2 \Big|_{\ell=0}^{\min\{b,m\}} \mathrm{d}m \\
&= \int_0^a \left[6m \times \min\{b,m\} - 3\min\{b,m\}^2 \right] \mathrm{d}m \\
&= \int_0^b \left[6m \times \min\{b,m\} - 3\min\{b,m\}^2 \right] \mathrm{d}m \\
&\quad + \int_b^a \left[6m \times \min\{b,m\} - 3\min\{b,m\}^2 \right] \mathrm{d}m \\
&= \int_0^b \left[6m \times m - 3m^2 \right] \mathrm{d}m \\
&\quad + \int_b^a \left[6m \times b - 3b^2 \right] \mathrm{d}m \\
&= \int_0^b 3m^2 \mathrm{d}m + \int_b^a 3b(2m-b)\mathrm{d}m \\
&= b^3 + 3ab(a-b).
\end{aligned}$$

Exercise 10.3. Let $(X_i)_{i\in\mathbb{N}}$, $(Y_i)_{i\in\mathbb{N}}$, and $(Z_i)_{i\in\mathbb{N}}$ be sequences of i.i.d. random variables with

$$\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = 1/2,$$

and similarly $\mathbb{P}[Y_1 = 1] = \mathbb{P}[Y_1 = -1] = 1/2$ as well as $\mathbb{P}[Z_1 = 1] = \mathbb{P}[Z_1 = -1] = 1/2$, which are also independent of each other. Put differently, the sequence

$$(X_1, Y_1, Z_1, X_2, Y_2, Z_2, X_3, Y_3, Z_3, \dots)$$

is a sequence of i.i.d. random variables.

We define the partial sums

$$S_n^{(x)} \coloneqq \sum_{i=1}^n X_i, \quad S_n^{(y)} \coloneqq \sum_{i=1}^n Y_i, \quad \text{and} \quad S_n^{(z)} \coloneqq \sum_{i=1}^n Z_i.$$

The sequence $\big((S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\big)_{n\in\mathbb{N}}$ is called a random walk in $\mathbb{Z}^3$. Let $\alpha > 1/2$. Show that

$$\lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq n^\alpha \right] = 1,$$

where $\|(x,y,z)\| \coloneqq \sqrt{x^2+y^2+z^2}$ is the Euclidean norm.

Hint: First, apply the CLT to show that for all $\alpha > 1/2$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^\alpha \right] = \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(y)}| \leq n^\alpha \right] = \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(z)}| \leq n^\alpha \right] = 1.$$

Then, notice that for $\alpha' \in (1/2, \alpha)$ we have:

$$\left(\{|S_n^{(x)}| \leq n^{\alpha'}\} \cap \{|S_n^{(y)}| \leq n^{\alpha'}\} \cap \{|S_n^{(z)}| \leq n^{\alpha'}\}\right) \subseteq \left\{\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq \sqrt{3}n^{\alpha'}\right\}.$$

Use this to conclude.

Solution: We first show that for all $\alpha > 1/2$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^\alpha\right] = 1. \tag{1}$$

Since $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}[X_1] = 1$, we obtain by the CLT that for any $a \in \mathbb{R}$:

$$\mathbb{P}\left[S_n^{(x)} \leq a\sqrt{n}\right] = \mathbb{P}\left[\frac{S_n^{(x)}}{\sqrt{n}} \leq a\right] \xrightarrow{n\to\infty} \Phi(a),$$

and therefore also

$$\mathbb{P}\left[|S_n^{(x)}| \leq a\sqrt{n}\right] = \mathbb{P}\left[S_n^{(x)} \leq a\sqrt{n}\right] - \mathbb{P}\left[S_n^{(x)} \leq -a\sqrt{n}\right] \xrightarrow{n\to\infty} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1.$$

Since, for a fixed $a > 0$ and for sufficiently large n we have $a\sqrt{n} \leq n^\alpha$, monotonicity implies

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^\alpha\right] \geq \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq a\sqrt{n}\right] = 2\Phi(a) - 1, \quad a > 0.$$

As this inequality holds for every $a > 0$, we can write

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^\alpha\right] \geq \sup_{a>0}\{2\Phi(a) - 1\} = 1,$$

and so property (1) follows. Analogously, we can show that the same result holds for $S_n^{(y)}$ and $S_n^{(z)}$.

We now show that for all $\alpha > 1/2$, we have $\lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq n^\alpha\right] = 1$.

Choose $\alpha' \in (1/2, \alpha)$ and observe

$$\left(\{|S_n^{(x)}| \leq n^{\alpha'}\} \cap \{|S_n^{(y)}| \leq n^{\alpha'}\} \cap \{|S_n^{(z)}| \leq n^{\alpha'}\}\right) \subseteq \left\{\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq \sqrt{3}n^{\alpha'}\right\}.$$

Since $n^\alpha \geq \sqrt{3}n^{\alpha'}$ for large n, we get

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq n^\alpha\right] &\geq \lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\| \leq \sqrt{3} \cdot n^{\alpha'}\right] \\ &\geq \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^{\alpha'}, |S_n^{(y)}| \leq n^{\alpha'}, |S_n^{(z)}| \leq n^{\alpha'}\right] = 1,\end{aligned}$$

where the equality in the last step follows from the union bound, since we have

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \leq n^{\alpha'}, |S_n^{(y)}| \leq n^{\alpha'}, |S_n^{(z)}| \leq n^{\alpha'}\right] \\ &\quad = 1 - \lim_{n\to\infty} \mathbb{P}\left[\left\{|S_n^{(x)}| > n^{\alpha'}\right\} \cup \left\{|S_n^{(y)}| > n^{\alpha'}\right\} \cup \left\{|S_n^{(z)}| > n^{\alpha'}\right\}\right] \\ &\quad \geq 1 - \lim_{n\to\infty} \left(\mathbb{P}\left[|S_n^{(x)}| > n^{\alpha'}\right] + \mathbb{P}\left[|S_n^{(y)}| > n^{\alpha'}\right] + \mathbb{P}\left[|S_n^{(z)}| > n^{\alpha'}\right]\right) \\ &\quad = 1,\end{aligned}$$

where we have used (1) in the last step.

Exercise 10.4. The median m of a distribution F is defined by $m := F^{-1}(1/2) = \inf\{x \in \mathbb{R} : F(x) \geq 1/2\}$. Let $X_1, X_2, \dots$ be i.i.d. random variables with distribution function F and median $m = 0$. Let Z_n denote the sample median of $X_1, \dots, X_n$, that is, Z_n is the middle observation.

More formally $Z_n = X_{(k)}$ where $k = \left[\frac{n}{2} + 1\right]$ and $X_{(1)} \leq \dots \leq X_{(n)}$ denote the order statistics of $X_1, \dots, X_n$ (i.e. $X_{(1)} = \min\{X_i \,|\, i \in \{1, \dots, n\}\}$, $X_{(n)} = \max\{X_i \,|\, i \in \{1, \dots, n\}\}$, etc.), and $[x]$ denotes the integer part of x.

(a) Let $Y_i^x = 1_{\{X_i \leq x\}}$ and define $S_n^x := \sum_{i=1}^n Y_i^x$. Compute $\mathbb{E}[S_n^x]$ and $\mathrm{Var}[S_n^x]$.

(b) Express the event $\{Z_n \leq x\}$ using the random variable S_n^x.

(c) Using the CLT, give an approximation for $\mathbb{P}[Z_n \leq x]$ as $n \to \infty$.

(d) (*) Find the limit

$$\lim_{n\to\infty} \frac{1/2 - \alpha_n}{\sqrt{\frac{1}{n}\alpha_n(1-\alpha_n)}},$$

where $\alpha_n := F\left(\frac{x}{\sqrt{n}}\right)$.

Solution:

(a) We have $1 - \mathbb{P}[Y_i^x = 0] = \mathbb{P}[Y_i^x = 1] = \mathbb{P}[X_i \leq x] = F(x)$, and so $Y_i^x \sim \mathrm{Bernoulli}(F(x))$. Consequently $S_n^x \sim \mathrm{Binomial}(n, F(x))$, and so we have

$$\mathbb{E}[S_n^x] = nF(x), \quad \mathrm{Var}[S_n^x] = nF(x)(1 - F(x)).$$

(b) We observe that

$$\{Z_n \leq x\} = \{S_n^x \geq k\},$$

where $k = \left[\frac{n}{2} + 1\right]$.

(c) Using the central limit theorem, we obtain

$$\begin{aligned}
\mathbb{P}[Z_n \leq x] &= \mathbb{P}[S_n^x \geq k] \\
&= \mathbb{P}\left[\frac{S_n^x}{n} \geq \frac{k}{n}\right] \\
&= \mathbb{P}\left[\frac{\frac{S_n^x}{n} - F(x)}{\sqrt{\frac{1}{n}F(x)(1-F(x))}} \geq \frac{\frac{k}{n} - F(x)}{\sqrt{\frac{1}{n}F(x)(1-F(x))}}\right] \\
&\approx 1 - \Phi\left(\frac{k/n - F(x)}{\sqrt{\frac{1}{n}F(x)(1-F(x))}}\right) \\
&= \Phi\left(-\frac{\frac{k}{n} - F(x)}{\sqrt{\frac{1}{n}F(x)(1-F(x))}}\right).
\end{aligned}$$

(d) Define $\alpha_n := \frac{1}{n}\mathbb{E}\Big[S_n\Big(\frac{x}{\sqrt{n}}\Big)\Big] = F\left(\frac{x}{\sqrt{n}}\right)$ and note

$$\mathrm{Var}\Big[S_n\Big(\frac{x}{\sqrt{n}}\Big)\Big] = \frac{\alpha_n(1-\alpha_n)}{n}.$$

We examine the limit

$$\lim_{n\to\infty} \frac{1/2-\alpha_n}{\sqrt{\frac{1}{n}\alpha_n(1-\alpha_n)}} = \lim_{n\to\infty} -\frac{\alpha_n - 1/2}{\frac{x}{\sqrt{n}}} \frac{x}{\sqrt{\alpha_n(1-\alpha_n)}}.$$

Note that

$$\lim_{n\to\infty} \frac{\alpha_n - 1/2}{\frac{x}{\sqrt{n}}} = \lim_{n\to\infty} \frac{F\left(\frac{x}{\sqrt{n}}\right) - F(0)}{\frac{x}{\sqrt{n}}} = F'(0),$$

and

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} F\left(\frac{x}{\sqrt{n}}\right) = F(0) = 1/2.$$

Thus,

$$\lim_{n\to\infty} \frac{1/2-\alpha_n}{\sqrt{\frac{1}{n}\alpha_n(1-\alpha_n)}} = -2F'(0)x.$$

Exercise 10.5. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $X_i \sim \mathcal{U}([\theta-1,\theta])$ under $\mathbb{P}_\theta$, where $\theta \in \mathbb{R}$ is an unknown parameter. We consider the following estimators for θ:

$$T_1^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\Big(X_i + \frac{1}{2}\Big) \quad \text{and} \quad T_2^{(n)} = \max\{X_1,\ldots,X_n\}.$$

(a) Determine whether the estimators are unbiased.

(b) Compute the variances $\mathrm{Var}_\theta[T_1^{(n)}]$ and $\mathrm{Var}_\theta[T_2^{(n)}]$.

(c) Compute the mean squared error

$$\mathrm{MSE}_\theta[T_i^{(n)}] := \mathbb{E}_\theta[(T_i^{(n)} - \theta)^2], \quad i \in \{1,2\}.$$

Remark: Here, $\mathbb{E}_\theta$ and Var_θ denote the mean and the variance under probability measure $\mathbb{P}_\theta$.

Solution:

(a) Fix $\theta \in \mathbb{R}$. As $X_i \sim \mathcal{U}([\theta-1,\theta])$, we have

$$\mathbb{E}_\theta[T_1^{(n)}] = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta[X_i]\right) + \frac{1}{2} = \mathbb{E}_\theta[X_1] + \frac{1}{2} = \theta - \frac{1}{2} + \frac{1}{2} = \theta.$$

So, $T_1^{(n)}$ is unbiased.

Define $Y_i^\theta := X_i - (\theta - 1)$ so that $Y_i \sim \mathcal{U}([0,1])$ under $\mathbb{P}_\theta$. We further define

$$Y_\theta^{(n)} := \max\{Y_1^\theta, \dots, Y_n^\theta\} = T_2^{(n)} - (\theta - 1).$$

The distribution function and the density of $Y_\theta^{(n)}$ are respectively (see Exercise 10.2)

$$F_{Y_\theta^{(n)}}(a) = \begin{cases} 0 & \text{if } a < 0, \\ a^n & \text{if } a \in [0,1], \\ 1 & \text{if } a > 1, \end{cases} \quad \text{and} \quad f_{Y_\theta^{(n)}}(a) = na^{n-1}\mathbf{1}_{\{a \in [0,1]\}}. \tag{2}$$

Hence,

$$\mathbb{E}_\theta[Y_\theta^{(n)}] = \int_{-\infty}^{\infty} a f_{Y_\theta^{(n)}}(a)\mathrm{d}a = n\int_0^1 a^n \mathrm{d}a = \frac{n}{n+1},$$

and so

$$\mathbb{E}_\theta[T_2^{(n)}] = \mathbb{E}_\theta[Y_\theta^{(n)}] + (\theta - 1) = \theta - \frac{1}{n+1}.$$

Thus, $T_2^{(n)}$ is not unbiased.

(b) We have

$$\mathrm{Var}_\theta[T_1^{(n)}] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\theta[X_i] = \frac{1}{n}\mathrm{Var}_\theta[X_1] = \frac{1}{12n}.$$

Now, using (2), we compute

$$\mathbb{E}_\theta[(Y_\theta^{(n)})^2] = n\int_0^1 a^{n+1}\mathrm{d}a = \frac{n}{n+2}, \quad \text{so}$$

$$\mathrm{Var}_\theta[T_2^{(n)}] = \mathrm{Var}_\theta[Y_\theta^{(n)}] = \mathbb{E}_\theta[(Y_\theta^{(n)})^2] - (\mathbb{E}_\theta[Y_\theta^{(n)}])^2 = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)}.$$

(c) Finally, we can compute

$$\mathrm{MSE}_\theta[T_1^{(n)}] = \mathrm{Var}_\theta[T_1^{(n)}] + (\mathbb{E}_\theta[T_1^{(n)}] - \theta)^2 = \mathrm{Var}_\theta[T_1^{(n)}] = \frac{1}{12n}.$$

$$\mathrm{MSE}_\theta[T_2^{(n)}] = \mathrm{Var}_\theta[T_2^{(n)}] + (\mathbb{E}_\theta[T_2^{(n)}] - \theta)^2 = \frac{n}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2} = \frac{2}{(n+1)(n+2)}.$$

Exercise 10.6. We model the water level above the critical flood mark (140 cm above normal) in Lake Zurich. Let X denote the water height (in cm) above the critical mark. We use a generalized Pareto distribution:

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta}(1+x)^{-(1+\frac{1}{\theta})} & \text{for } x > 0, \\ 0 & \text{for } x \le 0, \end{cases}$$

where $\theta > 0$ is an unknown parameter to be estimated based on observations $x_1, \dots, x_n$. These are modeled as realizations of i.i.d. random variables $X_1, \dots, X_n$ with density $f_X(x;\theta)$. We define the estimator by

$$T^{(n)} = \frac{1}{n}\sum_{i=1}^n \log(1 + X_i).$$

(a) Compute the expectation and variance of $T^{(n)}$ under $\mathbb{P}_\theta$ for each $\theta > 0$.

Hint: Define $Y_i \coloneqq \log(1 + X_i)$. Then $Y_i \sim \operatorname{Exp}(1/\theta)$, i.e., the density of Y_i is $f_{Y_i}(y) = \frac{1}{\theta} e^{-y/\theta} \mathbf{1}_{\{y \geq 0\}}$.

(b) Is $T^{(n)}$ an unbiased estimator for θ?

(c) Compute the mean squared error $\operatorname{MSE}_\theta[T^{(n)}]$.

(d) Find the maximum likelihood estimator for θ.

Solution:

(a) Linearity of expectation gives

$$\mathbb{E}_\theta[T^{(n)}] = \mathbb{E}_\theta\left[\frac{1}{n}\sum_{i=1}^n \log(1 + X_i)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta[\log(1 + X_i)] = \mathbb{E}_\theta[\log(1 + X_1)].$$

By the hint, $Y_1 \coloneqq \log(1 + X_1) \sim \operatorname{Exp}(1/\theta)$. Therefore,

$$\mathbb{E}_\theta[Y_1] = \theta, \quad \text{and} \quad \operatorname{Var}_\theta[Y_1] = \theta^2.$$

We thus conclude

$$\mathbb{E}_\theta[T^{(n)}] = \mathbb{E}_\theta[Y_1] = \theta.$$

Similarly, we can compute thanks to independence

$$\begin{aligned}\operatorname{Var}_\theta[T^{(n)}] = \operatorname{Var}_\theta\left[\frac{1}{n}\sum_{i=1}^n \log(1 + X_i)\right] &= \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}_\theta[\log(1 + X_i)] \\ &= \frac{1}{n}\operatorname{Var}_\theta[\log(1 + X_1)] = \frac{1}{n}\theta^2.\end{aligned}$$

(b) From (a), we know that $\mathbb{E}_\theta[T^{(n)}] = \theta$, so $T^{(n)}$ is an unbiased estimator of θ.

(c) Since the estimator is unbiased (see part b), the mean squared error simplifies to:

$$\operatorname{MSE}_\theta[T^{(n)}] = \operatorname{Var}_\theta[T^{(n)}] = \frac{\theta^2}{n}.$$

(d) The likelihood function is

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta) = \frac{1}{\theta^n}\prod_{i=1}^n (1 + x_i)^{-(1+\frac{1}{\theta})} \mathbf{1}_{\{x_1 \geq 0, \ldots, x_n \geq 0\}}.$$

Thus, the log-likelihood function is

$$\begin{aligned}\log L(x_1, \ldots, x_n; \theta) &= \log\left(\frac{1}{\theta^n}\prod_{i=1}^n (1 + x_i)^{-(1+\frac{1}{\theta})}\right) \mathbf{1}_{\{x_1 \geq 0, \ldots, x_n \geq 0\}} \\ &= \left(-n\log\theta - \left(1 + \frac{1}{\theta}\right)\sum_{i=1}^n \log(1 + x_i)\right)\mathbf{1}_{\{x_1 \geq 0, \ldots, x_n \geq 0\}}.\end{aligned}$$

Differentiating with respect to θ gives

$$\frac{\partial}{\partial\theta}\log L(x_1,\dots,x_n;\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \log(1+x_i), \quad x_1 \geq 0,\dots,x_n \geq 0.$$

We further have

$$-\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \log(1+x_i) = 0 \iff n\theta = \sum_{i=1}^n \log(1+x_i) \iff \theta = \frac{1}{n}\sum_{i=1}^n \log(1+x_i).$$

Thus, the point

$$\theta^\star = \frac{1}{n}\sum_{i=1}^n \log(1+x_i)$$

is a unique stationary point.

We compute the second derivative:

$$\frac{\partial^2}{\partial\theta^2}\log L(x_1,\dots,x_n;\theta) = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n \log(1+x_i), \qquad x_1 \geq 0,\dots,x_n \geq 0.$$

Plugging in $\theta^\star$, we verify

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log L(x_1,\dots,x_n;\theta^\star) &= \frac{n}{(\theta^\star)^2} - \frac{2}{(\theta^\star)^3}\sum_{i=1}^n \log(1+x_i) \\ &= \frac{1}{\theta^\star}\Big(\frac{n}{\theta^\star} - \frac{2}{(\theta^\star)^2}\sum_{i=1}^n \log(1+x_i)\Big) \\ &= \frac{1}{\theta^\star}\Big(-\frac{1}{(\theta^\star)^2}\sum_{i=1}^n \log(1+x_i)\Big) \\ &= -\frac{1}{(\theta^\star)^3}\sum_{i=1}^n \log(1+x_i) \\ &< 0.\end{aligned}$$

This confirms that $\theta^\star$ is a maximum of the log-likelihood function.

We conclude that the maximum likelihood estimator for θ is

$$T_{ML} = \frac{1}{n}\sum_{i=1}^n \log(1+X_i) = T^{(n)}.$$