Probability and Statistics

Exercise sheet 10

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 $\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \le 0.75\right]$ using the CLT?
   (a) $\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \le 0.75\right] \approx 2\Phi\left(\frac{3}{4}\sqrt{6}\right) - 1$.
   (b) $\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \le 0.75\right] \approx 2\Phi\left(\frac{7}{4}\sqrt{6}\right)$.
   (c) $\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \le 0.75\right] \approx \Phi\left(\sqrt{\frac{7}{4}}\right)$.
   (d) $\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \le 0.75\right] \approx 1 - 2\Phi\left(\sqrt{6}\right)$.

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

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

(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).

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

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

The sequence $\left((S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\right)_{n\in\mathbb{N}}$ 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 $\|(x, y, z)\| := \sqrt{x^2 + y^2 + z^2}$ 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.

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

Exercise 10.5. Let $X_1, \dots, X_n$ 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:

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

(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

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

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

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 $Y_i \sim \mathrm{Exp}(1/\theta)$, i.e., the density of Y_i is $f_{Y_i}(y) = \frac{1}{\theta} e^{-y/\theta} \mathbf{1}_{\{y \ge 0\}}$.

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

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

(d) Find the maximum likelihood estimator for θ.