Probability and Statistics

Exercise sheet 4

Exercise 4.1 (On measurability)

(a) Consider $X_1, ..., X_m$ ($m \geq 1$) random variables defined on some $(\Omega, \mathcal{A}, P)$ and taking values in $(\mathbb{R}, \mathcal{B})$, where $\mathcal{A}$ is a σ-algebra on Ω, P is a probability measure on $\mathcal{A}$ and $\mathcal{B}$ is the Borel σ-field.

Show that

$$-X_1, \max_{1 \leq i \leq m} X_i, \min_{1 \leq i \leq m} X_i$$

are random variables.

(b) Consider now a sequence $(X_n)_{n \geq 1}$ of random variables $X_1, X_2, ...$ defined on $(\Omega, \mathcal{A}, P)$ and taking values in $(\mathbb{R}, \mathcal{B})$ as in (a).

Show that:

$$\sup_{n \geq 1} X_n, \inf_{n \geq 1} X_n, \limsup_{n \to \infty} X_n, \liminf_{n \to \infty} X_n, \lim_{n \to \infty} X_n$$

are all random variables, in the case of the limit assuming that it exists.

Here, recall the definition of

$$\limsup_{n \to \infty} := \inf_{n \geq 1} (\sup_{k \geq n} X_k)$$

and

$$\liminf_{n \to \infty} := \sup_{n \geq 1} (\inf_{k \geq n} X_k).$$

Exercise 4.2 (On the cdf of min and max of i.i.d random variables)

Let $X_1, ..., X_n$ be $\overset{\text{iid}}{\sim} F$.

(a) Let $S_n := \max_{1 \leq i \leq n} X_i$. Find the cdf of S_n as a function of F.

(b) Do the same but for $I_n := \min_{1 \leq i \leq n} X_i$.

(c) Fix $x \in \mathbb{R}$ such that $F(x) \in (0, 1)$. What is the limit of the cdf of S_n at x as $n \to \infty$? What about the cdf of I_n? How would you interpret these results? What does this mean if $X_1, ..., X_n$ take values in a finite set $\{\xi_1, ..., \xi_k\}$?

Exercise 4.3 (On expectation)

(a) For any cdf F show that

$$P(X \in (a, b]\,) = F(b) - F(a)$$

for any $a < b$ and where X is a random variable with cdf F.

(b) Let X be a nonnegative discrete random variable taking its values in the set $\{x_1, x_2, ...\}$ (possibly countably infinite), where we assume that the values are ordered by $x_1 < x_2 < ....$ Suppose $E(X)$ exists. Show that

$$E(X) = \sum_{j=0}^{\infty}(x_{j+1} - x_j)P(X > x_j)$$

with $x_0 := 0$.

Does this match with the tail sum seen in the lecture?

(c) Show that if F is the cdf of X (the same X as in (b)), then $E(X)$ can also be given by the formula

$$E(X) = \int_0^{\infty}(1 - F(x))dx. \tag{1}$$

(d) Show that for a general discrete random variable (possibly taking values in $(-\infty, 0)$),

$$E(X) = -\int_{-\infty}^{0} P(X < x)dx + \int_0^{\infty}(1 - F(x))dx \tag{2}$$

provided that $E(X)$ exists.

Remark: Actually, the formulas in 1 and 2 are true in general for any type of nonnegative and general random variables. Also, $\int_{-\infty}^{0} P(X < x)dx$ can be replaced by $\int_{-\infty}^{0} F(x)dx$.

Exercise 4.4 (Quantiles)

For a given $0 < \alpha < 1$, we call the α-quantile of F the quantity

$$q_\alpha = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\},$$

where F is a given cdf.

Remark: The function $\alpha \mapsto q_\alpha$ is also called the generalised inverse of the cdf F. If $\alpha = \frac{1}{2}$, $q_{\frac{1}{2}}$ is called the median.

(a) Show that $\alpha \mapsto q_\alpha$ is non-decreasing on $(0, 1)$.

(b) Toss a fair coin n times and record

$$X_i = \begin{cases} 1 & \text{if heads at the } i\text{th toss} \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$Y_n = \sum_{i=1}^{n} X_i$$

be the number of heads obtained in the n tosses.

Find the cdf of Y_n. Call it F_n.

(c) What is the median of F_n when n is even and when it is odd?

Exercise 4.5 (an interesting property of expectations)

(a) Suppose that X is a random variable. Show that $E(X^2) < \infty$ if and only if $\mathrm{var}(X) < \infty$.

(b) Suppose $\mathrm{var}(X) < \infty$. Show that $E(X)$ minimises the function

$$a \mapsto E[(X-a)^2] \quad (a \in \mathbb{R}).$$

Exercise 4.6 (Optional, for the more courageous)

Consider again the birthday problem from another perspective. Suppose that people are coming to a party and you are assigned the mission of writing down the birth date of each guest as they show up.

Let X be the number of people that showed up until you see for the first time a person born on the same day as somebody who showed up earlier.

(a) Find an expression for $E(X)$ (this can be done in two different ways).

(b) Find an expression for $\sigma = \sqrt{\mathrm{var}(X)}$.

(c) The numerical values are given as

$$E(X) \approx 24.62,$$
$$\sigma \approx 12.19.$$

Find an interval $[a, b]$ which satisfies

$$P(a \leq X \leq b) \geq 0.5.$$

Hint: Use Chebyshev's inequality.