Universität Bielefeld
Tanja Pasurek
Tutor: Ehsan Abedi WiSe 2023/24

Foundations of Statistics

Homework 4

Random variables. Expectation and variance.

Exercise 1 (Indicator random variable).
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Given an event $A \in \mathcal{A}$, define the indicator random variable

$$\mathbb{I}_A(\omega) := \begin{cases} 1, & \text{if } A \text{ occurs} \ \ (\text{i.e. } \omega \in A), \\ 0, & \text{if } A \text{ does not occur} \ \ (\text{i.e. } \omega \notin A). \end{cases}$$

(a) Prove that for any $A, B \in \mathcal{A}$

$$\mathbb{I}_A^2 = \mathbb{I}_A, \quad \mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B, \quad \mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_A \mathbb{I}_B.$$

(b) Show that $\mathbb{I}_A \sim \mathrm{Ber}(p)$ where $p = \mathbb{P}(A)$.

(c) Check the fundamental relation $\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A)$.

(d) Suppose that a random variable $U : \Omega \to [0, 1]$ has a uniform distribution, i.e. $U \sim \mathrm{Unif}(0, 1)$. For some $0 < p < 1$ define a discrete random variable

$$X(\omega) := \begin{cases} 1, & \text{if } U(\omega) < p, \\ 0, & \text{if } U(\omega) \geq p. \end{cases}$$

Show that $X \sim \mathrm{Ber}(p)$ and that it allows the representation $X = \mathbb{I}_A$.

Exercise 2 (Properties of the variance).

(a) Let X be a random variable with $\mathbb{E}(X^2) < \infty$. Show that for any constants $a, b \in \mathbb{R}$, we have

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

(b) Let X, Y be independent random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. Show that

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Exercise 3.
(a) Let $X : \Omega \to \mathbb{R}$ be a random variable with CDF F_X. Find the CDF of the random variable $Y := X^+ = \max\{0, X\}$.

(b) Let $X : \Omega \to \mathbb{R}$ be a continuous random variable with CDF F_X and PDF f_X. Find the CDF and PDF of the the random variable $Y := -X$.

Exercise 4 (Linear change-of-units transformation).
(a) A random variable U is uniformly distributed over $[0, 1]$, i.e. $U \sim \mathrm{Unif}(0, 1)$. Determine the distribution of $V := rU + s$ for any numbers $r > 0$ and $s \in \mathbb{R}$. What happens for $r = 0$ and $r < 0$?

(b) Let a continuous random variable $X : \Omega \to \mathbb{R}$ have CDF F_X and PDF f_X, and let us change units to $Y := rX + s$ for $r > 0$ and $s \in \mathbb{R}$. Show that

$$F_Y(y) = F_X\left(\tfrac{y-s}{r}\right) \quad \text{and} \quad f_Y(y) = \tfrac{1}{r} f_X\left(\tfrac{y-s}{r}\right).$$

(c) Determine the distribution of λX for $X \sim \mathrm{Exp}(\lambda)$ using **(b)**, where $\lambda > 0$. What kind of distribution does λX have?

Exercise 5. A hydrologist in Monsville models the maximum one-day rainfall X (inches) in a randomly selected year, following the density

$$f(x) := \frac{3}{2}\left(\frac{y}{2} + 1\right)^{-4} \quad \text{for } y \geq 0.$$

This is a special case of *Pareto density*, which is used to model phenomena where the values taken tend to be mostly "small" but occasionally "big".

(a) Plot the density function. Prove its normalization first by direct computation and then using the numerical integration in `R`.
(b) The severve flooding in Monsville happens if there is more than 7 inches of rain in one day. What is the proportion of years with severve flooding?
(c) Compute the average and variance of maximum-one day rainfall.

Exercise 6. The built-in-dataset `islands` in the package `stats` contains the size of the world's land masses that exceed 10,000 square miles. You can access that data for example by `require(stats);data(islands)`.
(a) How many are there? Print the 5 largest land masses using `sort()`.
(b) Find the smallest and the largest observations (i.e. its name and size) using the command `which`.
(c) Compute the 1st quantile and median based on their definitions. Check your result using the command `summary()`.
(d) Draw a histogram with a good choice of bin size. Try also
`dotchart(log(sort(islands), 10))`