ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

M.Sc. in Data Science

Course: Probability and Statistics for Data Analysis
Semester: Fall 2024
Instructor: Ioannis Vrontos (`vrontos@aueb.gr`)
Grader: Konstantinos Bourazas (`kbourazas@aueb.gr`)

Assignment 1

1. Suppose that events A and B in a sample space S satisfy $P(A) = 1/2$, $P(B'|A) = 3/4$, and $P(A \cup B) = 3/4$. Calculate:

(a) $P(A \cap B)$

(b) $P(B)$

(c) $P(A|B)$

(d) Are A and B independent?

2. Suppose we roll a fair six-sided die until we obtain a 6. Let X denote the number of rolls required.

(a) What is the probability that $X = n$, for any positive integer n?

(b) What is the probability that a 6 appears for the first time in the 10^{th} roll?

(c) Calculate the expected value $E(X)$.

3. Consider a communications channel where a symbol "0" is transmitted with probability 0.6 and a symbol "1" with probability 0.4. Due to interference, a "0" may be incorrectly received as a "1" with probability 0.1, and a "1" may be incorrectly received as a "0" with probability 0.2.

(a) If a "1" is received, what is the probability that a "1" was actually sent?

(b) If two "0" symbols are received consecutively, what is the probability that two "0" symbols were sent?

4. Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. For a fixed value x_0 with $F(x_0) < 1$, define:

$$g(x) = \begin{cases} \dfrac{f(x)}{1 - F(x_0)}, & \text{if } x \geq x_0, \\ 0, & \text{if } x < x_0. \end{cases}$$

Show that $g(x)$ is a valid probability density function (pdf).

5. Suppose that customer arrivals at a service desk follow a Poisson process with a rate of 3 per minute.

(a) What is the probability of no arrivals in the next minute?

(b) What is the probability of at least three arrivals in the next minute?

(c) What is the probability of at most five arrivals in the next three minutes?

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $Weibull(\kappa, \lambda)$ distribution with known shape parameter κ and unknown scale parameter λ. The pdf of X is given by:

$$f(x; \kappa, \lambda) = \frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{-(x/\lambda)^{\kappa}}, \quad x \geq 0,\, \lambda > 0.$$

Find a sufficient statistic for the parameter λ.

7. A pharmaceutical company collects two random samples to measure the concentration of a substance. For sample 1, $\bar{x}_1 = 45.3$ mg, $s_1^2 = 4.1$, $n_1 = 15$. For sample 2, $\bar{x}_2 = 47.8$ mg, $s_2^2 = 3.9$, $n_2 = 18$. Assume normal distribution for both samples.

(a) Construct a 95% confidence interval for the mean concentration of each sample.

(b) Test at the 5% significance level if the means of the two samples are statistically different.

(c) Test if the mean concentration of the first sample is significantly greater than 45 mg at the 1% significance level.