Introduction to Probability & Statistics

Assignment 4, 2024/25

Instructions

Submit your answers to all of the **Assignment Questions (AQ)**. You should upload your solutions to the VLE as a *pdf file*.
Marks will be awarded for clear, logical explanations, as well as for correctness of solutions.
You should also look at the other questions in preparation for your Week 9 seminar.

Practice Questions

PQ1. Let X be a random variable with $\mathbb{E}[X] = 5$. What is the expectation of $3X + 5$? If furthermore $\mathbb{E}[X^2] = 30$, what is the variance of X?

Answer

We can use the linearity of expectation to find that $\mathbb{E}[3X + 5] = 3\mathbb{E}[X] + 5 = 20$. The variance is $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 30 - 5^2 = 5$.

PQ2. I arrive at the train station at 12.00 exactly. My train departs at a time which follows a (continuous) uniform distribution on the interval [11.55, 12.15]. What is the probability that I miss my train?

Answer

Let X denote the random time after 11.55 at which the train leaves. The question tells us that $X \sim \mathrm{Uniform}[0, 20]$. I miss the train if $X < 5$, which has probability

$$\mathbb{P}(X < 5) = \int_0^5 \frac{1}{20} dx = \frac{1}{4}.$$

PQ3. Suppose that you have a lecture at 14.00, and that the time taken to travel from your room to the lecture theatre is normally distributed with mean 30 minutes and standard deviation 4 minutes. What is the latest time you should leave your room if you want to be 99% certain that you will not miss the start of the lecture? (Hint: if $Z \sim \mathrm{N}(0, 1)$ then the R function `qnorm(p)` returns the value $z \in \mathbb{R}$ such that $\mathbb{P}(Z \leq z) = p$.)

Answer

Let X denote the travel time to the lecture: $X \sim \mathrm{N}(30, 16)$. We wish to find x such that $\mathbb{P}(X \leq x) = 0.99$. Now,

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - 30}{4} \leq \frac{x - 30}{4}\right) = \mathbb{P}\left(Z \leq \frac{x - 30}{4}\right)$$

where $Z \sim \mathrm{N}(0, 1)$.

We can get hold of this value of x by using R (or by consulting statistical tables): `qnorm(0.99)` gives the value 2.326, meaning that $\mathbb{P}(Z \le 2.326) = 0.99$. Thus we require $(x - 30)/4 = 2.326 \iff x = 39.3$. Thus the latest you should leave your room is 39.3 minutes before the start of the lecture: i.e. at 13:20.

PQ4. A random variable Z has probability density function

$$f_Z(x) = \begin{cases} \frac{6}{5675}(5x^2 + 3x + 11) & \text{for } 3 \le x \le 8 \\ 0 & \text{otherwise.} \end{cases}$$

Would you expect $\mathbb{E}[Z]$ to lie closer to 3 or to 8? Calculate $\mathbb{E}[Z]$ and check whether your intuition was correct.

Answer

Since f_Z is increasing on the interval $[3, 8]$ we know from the interpretation of expectation as centre of mass that the expectation should lie closer to 8 than to 3. The computation:

$$\mathbb{E}[Z] = \int_3^8 x f_Z(x) dx = \frac{6}{5675} \int_3^8 \left(5x^3 + 3x^2 + 11x\right) dx = \frac{2787}{454} = 6.14.$$

PQ5. Give an example of a joint probability table for two discrete random variables X and Y, each having only two possible values, so that $F_{X,Y}(5, 6) = 0.4, F_X(5) = 0.5, F_Y(6) = 0.6$ and $\mathbb{E}[X] = 10, \mathbb{E}[Y] = 4$.

Answer

One possible example would be

$y \backslash x$	0	20	$p_Y(y)$
0	0.4	0.2	0.6
10	0.1	0.3	0.4
$p_X(x)$	0.5	0.5	1

PQ6. The joint probability mass function $p_{X,Y}(x, y)$ of two random variables X and Y is summarised by the following table:

$x \backslash y$	-1	0	1
4	$\eta - 1/16$	$1/4 - \eta$	0
5	$1/8$	$3/16$	$1/8$
6	$\eta + 1/16$	$1/16$	$1/4 - \eta$

where η is a real number.

a) Extend the table by including also the marginal probabilities, i.e., the values of the probability mass functions p_X and p_Y.

b) Which are the valid choices for η?

c) Is there a value of η for which X and Y are independent?

Answer

a) We extend the probability table to also include the marginal probability mass functions p_X and p_Y:

$x\backslash y$	-1	0	1	$p_X(x)$
4	$\eta - 1/16$	$1/4 - \eta$	0	$3/16$
5	$1/8$	$3/16$	$1/8$	$7/16$
6	$\eta + 1/16$	$1/16$	$1/4 - \eta$	$3/8$
$p_Y(y)$	$2\eta + 1/8$	$1/2 - \eta$	$3/8 - \eta$	1

b) All entries of the probability table must be non-negative and they must sum up to 1. In order for $p_{X,Y}(4,-1)$ to be non-negative we need $\eta \geq 1/16$. In order for $p_{X,Y}(4,0)$ and $p_{X,Y}(6,1)$ to be non-negative we need $\eta \leq 1/4$. The sum over all entries is not affected by the value of η, so does not give any additional constraints. Therefore any $\eta \in [1/16, 1/4]$ is a valid choice.

c) It is easy to find counterexamples to the factorisation of the joint probability mass function that would have to hold if X and Y were independent. For example

$$p_X(4)p_Y(1) = \frac{3}{16}\left(\frac{3}{8} - \eta\right) \neq 0 = p_{X,Y}(4,1)$$

unless $\eta = 3/8$. However the value $\eta = 3/8$ is not allowed, and hence X and Y can never be independent.

PQ7. Let X and Y be random variables. Show that $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Answer

We start from the definition of covariance, and use linearity of expectation:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

Assignment Questions – answers to be uploaded

AQ1. Buses leave campus for the train station every 15 minutes, at 0, 15, 30 and 45 minutes past the hour. If a student arrives at the bus stop at a time that follows a (continuous) uniform distribution on the interval between 10.00 and 10.30, find the probability that they wait

a) less than 5 minutes for a bus;
b) at least 8 minutes for a bus.

AQ2. Let $X \sim \mathrm{Geom}(p)$. Calculate $\mathbb{E}[h(X)]$, where $h(x) = e^{tx}$ for some $t > 0$. For what values of t is $\mathbb{E}[h(X)] < \infty$?

AQ3. Show that if Z is a standard normal random variable then, for $x > 0$,

a) $\mathbb{P}(Z > x) = \mathbb{P}(Z < -x)$;

b) $\mathbb{P}(|Z| > x) = 2\mathbb{P}(Z > x)$;
c) $\mathbb{P}(|Z| < x) = 2\mathbb{P}(Z < x) - 1$.

Hint: express the probabilities in terms of integrals over the density function ϕ, and use the fact that ϕ is an even function (i.e. $\phi(z) = \phi(-z)$).

AQ4. Let $X : \Omega \to \{1, 2\}$ and $Y : \Omega \to \{0, 1\}$ be two discrete random variables. The following is a partial table of their joint and their marginal mass functions:

$y \backslash x$	1	2	$p_Y(y)$
0	$1/6$	$1/2$	
1			
$p_X(x)$	$5/12$		1

a) Fill in the missing values.
b) Determine the joint distribution function of X and Y.
c) Calculate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.
d) Let $Z = XY$. Calculate $\mathbb{E}[Z]$.

Other Questions (for seminars / extra practice)

OQ1. A married couple decide to have children until they have at least one child of each sex: let X denote the total number of children that they have. The probability of any one child being a boy is $1/2$ (with the sex of each child being independent of all the others).

a) What is the mass function of X? (I.e. write down $\mathbb{P}(X = n)$ for all $n \in \mathbb{N}$.)

b) Show that
$$\mathbb{E}[X] = 3\,.$$
Hint: you may find it useful to refer to the result from lectures that if $Y \sim \text{Geom}(p)$ then $\mathbb{E}[Y] = 1/p$.

OQ2. Let X be a discrete random variable. Show that for all functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$,
$$\mathbb{E}[h_1(X) + h_2(X)] = \mathbb{E}[h_1(X)] + \mathbb{E}[h_2(X)]\ .$$

OQ3. Let X and Y be random variables and let $r, s, t, u \in \mathbb{R}$. Show that
$$\rho(rX + s, tY + u) = \begin{cases} \rho(X, Y) & \text{if } rt > 0 \\ 0 & \text{if } rt = 0 \\ -\rho(X, Y) & \text{if } rt < 0 \end{cases}$$
where $\rho(X, Y)$ denotes the correlation coefficient of X and Y.

OQ4. Let $X \sim \text{Uniform}(0, a)$ for some $a > 0$. Show that for any $n \in \mathbb{N}$,
$$\mathbb{E}[X^n] = \frac{a^n}{n+1}.$$
Use this to determine $\rho(X, X^2)$, and show that this does not depend upon the value of a.

OQ5. A bag contains 3 cubes, 4 pyramids and 7 spheres. An object is drawn randomly from the bag and its type is recorded. Then the object is replaced. This is repeated 20 times.

a. Let C_i be the indicator random variable for the event that the i-th draw gives a cube, for $i = 1, \ldots, 20$. Calculate $\mathbb{E}[C_i], \mathbb{E}[C_i^2]$ and $\mathbb{E}[C_i C_j]$ for $i \neq j$.

b. Let C be the number of times a cube was drawn, Use that $C = \sum_{i=1}^{20} C_i$ to calculate $\mathbb{E}[C]$ and $\text{Var}(C)$.

c. Let S_i be the indicator random variable for the event that the i-th draw gives a sphere. Calculate $\mathbb{E}[C_i S_i]$ and $\mathbb{E}[C_i S_j]$ for $i \neq j$.

d. Let S be the number of times a sphere was drawn. Use the above results to calculate $\mathbb{E}[CS]$, $\text{Cov}(C, S)$, $\rho(C, S)$.

OQ6. Prove that binomial coefficients satisfy the identity

$$n\binom{n-1}{r-1} = r\binom{n}{r}.$$

Use this to find $\mathbb{E}[X]$ and $\text{Var}(X)$, where $X \sim \text{Bin}(n, p)$.

OQ7. [Harder] Consider a random variable $X \sim \text{Uniform}[a, b]$, where a and b are unknown. You are told that

$$\mathbb{P}(X < 2) = 1/3 \quad \text{and} \quad \mathbb{P}(1 < X \leq 3) = 1/2\,.$$

Given this information, find a and b.

OQ8. [Harder] Let X and Y be two independent geometrically distributed random variables with parameter p, i.e., $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(p)$. For any natural numbers i and n with $i < n$ calculate the conditional probability $\mathbb{P}(X = i \mid X + Y = n)$. Describe in words the meaning in terms of Bernoulli trials of what you just calculated.

! Challenge question

A stick of length 1 is snapped into two at a point $U \sim \text{Uniform}(0, 1)$. What is the expected length of the piece containing the point s, where s is some fixed number between 0 and 1? For what values of s is this expected length maximised/minimised? How does the variance of this length depend upon s?