Introduction to Probability & Statistics

Assignment 4, 2025/26

Instructions

Submit your answers to the **four questions marked Hand-in**. You should upload your solutions to the VLE as a *single pdf file*. Marks will be awarded for clear, logical explanations, as well as for correctness of solutions.
Solutions to questions marked have been released at the same time as this assignment, in case you want to check your answers or need a hint.
You should also look at the other questions in preparation for your Week 9 seminar.

Starters

These questions should help you to gain confidence with the basics.

S1. 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?

S2. 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 $X \sim \text{Uniform}[0, 20]$. I miss the train if $X < 5$, which has probability

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

S3. Hand-in

Buses leave campus for the train station every 20 minutes, at 5, 25, 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 09.00 and 09.35, find the probability that they wait

a) less than 5 minutes for a bus;
b) at least 10 minutes for a bus.

S4. 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 $\mathbb{P}(Z \leq z) = p$.)

S5. 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

S6. 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 $\mathbb{E}[Z]$.

Answer

a) The missing entries in the probability table are determined by the requirement that summing the joint probabilities across a row or across a column in the table gives the corresponding marginal probability and by the requirement that the marginal probabilities for X as well as those for Y have to add up to 1. So first we determine $p_Y(0) = 1/6 + 1/2 = 2/3$. Then we can determine $p_Y(1) = 1 - p_Y(0) = 1 - 2/3 = 1/3$ and $p_X(2) = 1 - p_X(1) = 1 - 5/12 = 7/12$. Finally we determine $p_{X,Y}(1,1) = p_X(1) - p_{X,Y}(1,0) = 5/12 - 1/6 = 1/4$ and $p_{X,Y}(2,1) = p_X(2) - p_{X,Y}(2,0) = 7/12 - 1/2 = 1/12$.

$y\backslash x$	1	2	$p_Y(y)$
0	$1/6$	$1/2$	$2/3$
1	$1/4$	$1/12$	$1/3$
$p_X(x)$	$5/12$	$7/12$	1

b) The joint distribution function $F_{X,Y}(x,y)$ is by definition given by $\mathbb{P}(X \leq x, Y \leq y)$. So for example

$$F_{X,Y}(1.5,1.5) = p_{X,Y}(1,0) + p_{X,Y}(1,1) = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}.$$

By doing more such calculations we find that

$$F_{X,Y} = \begin{cases} 0 & \text{if } x < 1 \text{ or } y < 0 \\ 1/6 & \text{if } x \in [1,2) \text{ and } y \in [0,1) \\ 5/12 & \text{if } x \in [1,2) \text{ and } y \geq 1 \\ 2/3 & \text{if } x \geq 2 \text{ and } y \in [0,1) \\ 1 & \text{if } x \geq 2 \text{ and } y \geq 1. \end{cases}$$

c) For calculating the expectations of X and Y we can use their marginal mass functions:

$$\mathbb{E}[X] = 1 \cdot p_X(1) + 2 \cdot p_X(2) = 1 \cdot \frac{5}{12} + 2 \cdot \frac{7}{12} = \frac{19}{12}$$

and

$$\mathbb{E}[Y] = 0 \cdot p_Y(0) + 1 \cdot p_Y(1) = p_Y(1) = \frac{1}{3}.$$

d) The random variable $Z = XY$ can take the possible values 0, 1 and 2 with probabilities

$$p_Z(0) = p_{X,Y}(1,0) + p_{X,Y}(2,0) = p_Y(0) = \frac{2}{3}$$
$$p_Z(1) = p_{X,Y}(1,1) = \frac{1}{4}, \quad p_Z(2) = p_{X,Y}(2,1) = \frac{1}{12}.$$

Thus

$$\mathbb{E}[Z] = 1 \cdot p_Z(1) + 2 \cdot p_Z(2) = \frac{1}{4} + 2\frac{1}{12} = \frac{5}{12}.$$

S7. Hand-in

Let $X : \Omega \to \{0,1\}$ 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:

$y \backslash x$	0	1	$p_Y(y)$
0	$1/8$		$1/4$
1		$1/2$	
$p_X(x)$			

a) Fill in the missing values.
b) Calculate $\mathbb{E}[XY]$.

Mains

These are important, and cover some of the most substantial parts of the course.

M1. A random variable W has probability density function

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

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

Answer

Since f_W 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:

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

M2. Let $X \sim \text{Geom}(p)$. Calculate $\mathbb{E}[h(X)]$, where $h(x) = e^{tx}$ for some $t > 0$. For what values of t is $\mathbb{E}[h(X)] < \infty$?

M3. Hand-in

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

M4. Hand-in

Let $X \sim \text{Exp}(\lambda)$. Use proof by induction to show that

$$\mathbb{E}[X^m] = \frac{m!}{\lambda^m} \qquad \text{for all } m \in \mathbb{N} \cup \{0\}.$$

M5. Let X and Y be random variables. Show that $\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

M6. The joint probability mass function $p_{X,Y}(x,y)$ of two random variables X and Y is summarised by the following table, where η is some real number:

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

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.

M7. 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 $\mathbb{E}[Y] = 1/p$.

M8. 🛠 Let X be a discrete random variable. Show that for all functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[h_1(X) + h_2(X)] = \mathbb{E}[h_1(X)] + \mathbb{E}[h_2(X)]\,.$$

Answer

Let $h(x) = h_1(x) + h_2(x)$. From the formula for the expectation of a function of a discrete random variable it follows that

$$\begin{aligned}
\mathbb{E}[h(X)] &= \sum_{k\in X(\Omega)} h(k)p_X(k) \\
&= \sum_{k\in X(\Omega)} (h_1(k)+h_2(k))p_X(k) \\
&= \sum_{k\in X(\Omega)} h_1(k)p_X(k) + \sum_{k\in X(\Omega)} h_2(k)p_X(k) \\
&= \mathbb{E}[h_1(X)] + \mathbb{E}[h_2(X)]\,.
\end{aligned}$$

M9. 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.

Desserts

Still hungry for more? Try these if you want to push yourself further. (These are mostly harder than I'd expect you to answer in an exam, or involve non-examinable material.)

D1. 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 $\mathsf{Var}(X)$, where $X \sim \mathsf{Bin}(n, p)$.

D2. Let $X \sim \mathsf{Uniform}(0, a)$ 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.

D3. 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 $i = 1, \dots, 20$. Calculate $\mathbb{E}[C_i], \mathbb{E}[C_i^2]$ and $\mathbb{E}[C_iC_j]$ 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 $\mathsf{Var}(C)$.

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

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

D4. Consider a random variable $X \sim \mathsf{Uniform}[a, b]$, 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.

D5. Let X and Y be two independent geometrically distributed random variables with parameter p, i.e., $X \sim \mathsf{Geom}(p)$ and $Y \sim \mathsf{Geom}(p)$. 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 $U \sim \mathsf{Uniform}(0, 1)$. 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?