Question 1

Consider two discrete random variables X and Y with joint probability density function given by

$$f(x,y) = \begin{cases} c(2x+y), & \text{if } x \in \{0,1,2\} \text{ and } y \in \{0,1,2,3\}, \\ 0, & \text{otherwise.} \end{cases}$$

where c is an appropriately-chosen constant.

(a) Find the value of c.

(b) Find $\mathrm{P}(X=2, Y=1)$

(c) Find $\mathrm{P}(X \geq 1, Y=1)$

(d) Find $\mathrm{P}(X \geq 1, Y \leq 1)$

(e) Find the marginal probability (mass) function of X.

(f) Find the marginal probability (mass) function of Y.

(g) Are X and Y independent random variables?

(h) Find the the probability mass function of Y given $X = 2$.

(i) Compute $\mathrm{P}(Y=1|X=2)$.

(j) Compute $\mathrm{E}(Y|X=2)$.

Solution to Question 1

Part (a):

One can directly compute:

$$\begin{aligned}
& c\sum_{x=0}^{2}\sum_{y=0}^{3}(2x+y) = 1 \\
\Rightarrow & \sum_{x=0}^{2}\sum_{y=0}^{3}(2x+y) = \frac{1}{c} \\
\Rightarrow & \sum_{x=0}^{2}\sum_{y=0}^{3}2x + \sum_{x=0}^{2}\sum_{y=0}^{3}y = \frac{1}{c} \\
\Rightarrow & \sum_{x=0}^{2}4\cdot 2x + \sum_{x=0}^{2}6 = \frac{1}{c} \\
\Rightarrow & 8(3) + 6(3) = \frac{1}{c} \\
\Rightarrow & 42 = \frac{1}{c} \\
\Rightarrow & c = \frac{1}{42}
\end{aligned}$$

(Part (a) continued)

Or one can construct the following table for the joint probability mass function:

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	Totals $\downarrow$
$X = 0$	0	c	$2c$	$3c$	$6c$
$X = 1$	$2c$	$3c$	$4c$	$5c$	$14c$
$X = 2$	$4c$	$5c$	$6c$	$7c$	$22c$
Totals $\rightarrow$	$6c$	$9c$	$12c$	$15c$	$42c$

Table 1: Table of values for $f(x, y)$ for Question 1.

Looking at the bottom right corner, we see that the sum of the probabilities add up to $42c$. But this implies

$$\begin{aligned} 42c &= 1 \\ \Rightarrow c &= \frac{1}{42} \end{aligned}$$

However, Table 1 will help us in to quickly answer the next few questions.

Part (b):

Reading off Table 1,

$$\mathrm{P}\left(X = 2, Y = 1\right) = f(2, 1) = c(2 \cdot 2 + 1) = 5c = \frac{5}{42}.$$

Part (c):

Again, one can use Table 1,

$$\begin{aligned} \mathrm{P}\left(X \geq 1, Y = 1\right) &= \mathrm{P}\left(X = 1, Y = 1\right) + \mathrm{P}\left(X = 2, Y = 1\right) \\ &= 3c + 5c = 8c \\ &= \frac{8}{42} = \frac{4}{21} \end{aligned}$$

Part (d): One first computes (using Table 1):

$$\begin{aligned} \mathrm{P}\left(X \geq 1, Y = 0\right) &= \mathrm{P}\left(X = 1, Y = 0\right) + \mathrm{P}\left(X = 2, Y = 0\right) \\ &= 2c + 4c = 6c \\ &= \frac{6}{42} = \frac{1}{7} \end{aligned}$$

Then, one can compute:

$$\begin{aligned} \mathrm{P}\left(X \geq 1, Y \leq 1\right) &= \mathrm{P}\left(X \geq 1, Y = 0\right) + \mathrm{P}\left(X \geq 1, Y = 1\right) \\ &= \frac{6}{42} + \frac{8}{42} = \frac{14}{42} \\ &= \frac{1}{3} \end{aligned}$$

Part (e):

Denoting the probability function for the marginal distribution of X by f_1, we can simply read the marginal totals (in the right-hand "margin"):

$$f_1(x) = \begin{cases} 6c, & \text{if } x = 0, \\ 14c, & \text{if } x = 1, \\ 22c, & \text{if } x = 2, \\ 0, & \text{otherwise.} \end{cases}$$

$$\Rightarrow f_1(x) = \begin{cases} \frac{6}{42}, & \text{if } x = 0, \\ \frac{14}{42}, & \text{if } x = 1, \\ \frac{22}{42}, & \text{if } x = 2, \\ 0, & \text{otherwise.} \end{cases}$$

$$\Rightarrow f_1(x) = \begin{cases} \frac{1}{7}, & \text{if } x = 0, \\ \frac{1}{3}, & \text{if } x = 1, \\ \frac{11}{21}, & \text{if } x = 2, \\ 0, & \text{otherwise.} \end{cases}$$

Part (f):

Similarly to Part (e), we denote the probability function for the marginal distribution of Y by f_2, we can simply read the marginal totals (in the bottom "margin"):

$$f_2(y) = \begin{cases} 6c, & \text{if } y = 0, \\ 9c, & \text{if } y = 1, \\ 12c, & \text{if } y = 2, \\ 15c, & \text{if } y = 3, \\ 0, & \text{otherwise.} \end{cases}$$

$$\Rightarrow f_2(y) = \begin{cases} \frac{1}{7}, & \text{if } y = 0, \\ \frac{3}{14}, & \text{if } y = 1, \\ \frac{2}{7}, & \text{if } y = 2, \\ \frac{5}{14}, & \text{if } y = 3, \\ 0, & \text{otherwise.} \end{cases}$$

Part (g):

If the random variables X and Y are independent, then for all values of x and y:

$$\mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y).$$

Now, recall:

- In Part (b) we computed $\mathrm{P}(X = 2, Y = 1) = 5c = \frac{5}{42}$.
- From the marginal distribution of X in Part (e), one can compute $\mathrm{P}(X = 2) = \frac{11}{21}$.
- From the marginal distribution of Y in Part (f), $\mathrm{P}(Y = 1) = \frac{3}{14}$.

Therefore,

$$\mathrm{P}(X = 2)\,\mathrm{P}(Y = 1) = \frac{11}{\cancel{21}} \cdot \frac{\cancel{3}}{14} = \frac{11}{7} \cdot \frac{1}{14} = \frac{11}{98} \neq \frac{5}{42} = \mathrm{P}(X = 2, Y = 1)$$

and so X and Y are not independent.

Part (h):

Let us denote the probability mass function of Y given $X = 2$ as $f(y|2)$. Then, using the recalling the definition of a conditional probability (mass) function:

$$f(y|x) = \frac{f(x, y)}{f_1(x)},$$

So, with $x = 2$:

$$\begin{aligned} f(y|x = 2) &= \frac{c(2 \cdot 2 + y)}{11/21} \\ &= \frac{4 + y}{\cancel{42}} \cdot \frac{\cancel{21}}{11} \\ &= \frac{4 + y}{22} \end{aligned}$$

Part (i):

One uses $f(y|x = 2)$ from Part (h):

$$\mathrm{P}(Y = 1|X = 2) = f(y = 1|x = 2) = \frac{4 + 1}{22} = \frac{5}{22}$$

Part (j):

Again, one uses $f(y|x = 2)$ from Part (h):

$$\begin{aligned} \mathrm{E}(Y|X = 2) &= \sum_{y=0}^{3} y f(y|x = 2) \\ &= \sum_{y=0}^{3} y \left(\frac{4 + y}{22}\right) \\ &= (0)\left(\frac{4}{22}\right) + (1)\left(\frac{5}{22}\right) + (2)\left(\frac{6}{22}\right) + (3)\left(\frac{7}{22}\right) \\ &= \frac{5}{22} + \frac{12}{22} + \frac{21}{22} = \frac{38}{22} \\ &= \frac{19}{11} \end{aligned}$$

Question 2

Suppose X is uniformly distributed on the interval $[0,4]$, i.e. $X \sim \text{Unif}(0,4)$.

(a) Compute $\mathrm{P}\left(|X-2| \geq 1\right)$.

(b) Use Chebyshev's inequality to bound the probability that $|X-2| \geq 1$.

(c) Is the bound in (b) informative?

(d) For which values $\epsilon > 0$ can Chebyshev's inequality be used to obtain a nontrivial bound for $\mathrm{P}\left(|X-2| \geq \epsilon\right)$?

Solution to Question 2

Part (a):

The solution can be briefly written as

$$\mathrm{P}\left(|X-2| \geq 1\right) = \mathrm{P}\left(\{X \geq 3\} \text{ or } \{X \leq 1\}\right) = \mathrm{P}\left(X \geq 3\right) + \mathrm{P}\left(X \leq 1\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

When written out in detail, we first note that:

$$\begin{aligned}
&|X-2| \geq 1 \\
&\Rightarrow X-2 \geq 1 \text{ or } -(X-2) \geq 1 \\
&\Rightarrow X \geq 3 \text{ or } (X-2) \leq -1 \\
&\Rightarrow X \geq 3 \text{ or } X \leq 1
\end{aligned}$$

Now:

- Let A be the event $X \geq 3$.
- Let B be the event $X \leq 1$.
- Let C be the event $|X-2| \geq 1$.

Since an observed value of X cannot be simultaneously bigger than 3 and less than 1, $A \cap B = \varnothing$. Also, from the above, $C = A \cup B$. Therefore,

$$\begin{aligned}
\mathrm{P}\left(C\right) &= \mathrm{P}\left(A \cup B\right) \\
&= \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(A \cap B\right) \\
&= \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(\varnothing\right) \\
\Rightarrow \mathrm{P}\left(C\right) &= \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right)
\end{aligned}$$

since $\mathrm{P}\left(\varnothing\right) = 0$. So,

$$\mathrm{P}\left(|X-2| \geq 1\right) = \mathrm{P}\left(X \geq 3\right) + \mathrm{P}\left(X \leq 1\right).$$

The probability density function of $\text{Unif}(a,b)$ (from PBL Sheet 8) is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, the p.d.f. for $\text{Unif}(0,4)$ is

$$f_X(x) = \begin{cases} \frac{1}{4}, & \text{if } 0 < x < 4, \\ 0, & \text{otherwise.} \end{cases}$$

The cumulative distribution function for $X \sim \mathrm{Unif}(0,4)$ is then (for $x \in [0,4]$):

$$\begin{aligned} F_X(x) = \mathrm{P}(X \leq x) &= \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{0} (0)\,\mathrm{d}x + \int_{0}^{x} \frac{1}{4}\,\mathrm{d}x \\ &= \frac{x}{4} \end{aligned}$$

Then,

$$\mathrm{P}(X \leq 1) = \frac{1}{4}.$$

Since X is a continuous random variable, $\mathrm{P}(X = 3) = 0$, and therefore

$$\begin{aligned} \mathrm{P}(X \geq 3) &= 1 - \mathrm{P}(X < 3) \\ &= 1 - (\mathrm{P}(X < 3) + 0) = 1 - (\mathrm{P}(X < 3) + \mathrm{P}(X = 3)) \\ &= 1 - \mathrm{P}(X \leq 3) \\ &= 1 - \frac{3}{4} \\ &= \frac{1}{4} \end{aligned}$$

Therefore,

$$\mathrm{P}(|X-2| \geq 1) = \mathrm{P}(X \geq 3) + \mathrm{P}(X \leq 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Part (b):

We recall Chebyshev's inequality for a random variable X with mean μ and variance σ^2. For any constant $c > 0$,

$$\mathrm{P}(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}.$$

Here we have $X \sim \mathrm{Unif}(0,4)$. For a general $Y \sim \mathrm{Unif}(a,b)$, we computed in PBL Sheet 8, Question 1, that

$$\begin{aligned} \mathrm{E}(Y) &= \frac{a+b}{2} \\ \mathrm{Var}(Y) &= \frac{(b-a)^2}{12} \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mu &= \mathrm{E}(X) = 2 \\ \sigma^2 &= \mathrm{Var}(X) = \frac{4^2}{12} = \frac{16}{12} = \frac{4}{3} \end{aligned}$$

Taking $c = 1$, Chebyshev's inequality then gives us the bound

$$\mathrm{P}(|X-2| \geq 1) \leq \frac{4}{3}.$$

Part (c):

This is not informative, because $\mathrm{P}(|X-2| \geq 1) \in [0,1]$, and so we already had the bound

$$\mathrm{P}(|X-2| \geq 1) \leq 1 < \frac{4}{3}.$$

Part (d):

For any $\epsilon > 0$, Chebyshev's inequality gives

$$\mathrm{P}(|X-2| \geq \epsilon) \leq \frac{(4/3)}{\epsilon^2} = \frac{4}{3\epsilon^2},$$

and this bound is only nontrivial when

$$\begin{aligned} &\frac{4}{3\epsilon^2} < 1 \\ \Rightarrow &\epsilon^2 > \frac{4}{3} \\ \Rightarrow &\epsilon > \frac{2}{\sqrt{3}}. \end{aligned}$$

Question 3

Prove Proposition 1.8.8 from the lecture notes:

Proposition 1.8.8. Given a sample of observations $x_1, x_2, \ldots, x_n$, with sample median m. Then, for any real value a,

$$\min_a \left(\sum_{i=1}^{n} |x_i - a| \right) = \sum_{i=1}^{n} |x_i - m| \, .$$

Solution to Question 3

For a fixed $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we define the function $g_{\mathbf{x}} : \mathbb{R} \to \mathbb{R}$ for any $z \in \mathbb{R}$ by

$$g_{\mathbf{x}}(z) = \sum_{i=1}^{n} |x_i - z| \, .$$

Our goal is to find the value $a \in \mathbb{R}$ such that

$$g_{\mathbf{x}}(a) = \min_z g_{\mathbf{x}}(z) \, , \tag{1}$$

and show that this point a is the median of the sample $\mathbf{x}$. This solution is from [1].

Before proceeding to solve this problem in general, we look at the special case that $n = 2$, and $x_1 \leq x_2$, and define $g_{(x_1,x_2)}(z) = |x_1 - z| + |x_2 - z|$. In this special case, the value $a \in \mathbb{R}$ can be in one of three intervals:

$$\begin{aligned}
(1)\ a < x_1 (\leq x_2) \Rightarrow |x_1 - a| + |x_2 - a| &= (x_1 - a) + (x_2 - a) \\
&= x_1 + x_2 - 2a \\
&> x_1 + x_2 - 2x_1 \\
&= x_2 - x_1 \\
(2)\ a > x_2 (\geq x_1) \Rightarrow |x_1 - a| + |x_2 - a| &= -(x_1 - a) - [-(x_2 - a)] \\
&= 2a - x_1 - x_2 \\
&> 2x_2 - x_1 - x_2 \\
&= x_2 - x_1
\end{aligned}$$

$$\begin{aligned}
(3)\ a \in [x_1, x_2] \Rightarrow |x_1 - a| + |x_2 - a| &= -(x_1 - a) + (x_2 - a) \\
&= x_2 - x_1
\end{aligned}$$

This can be summarised as:

$$\begin{aligned}
z \notin [x_1, x_2] &\Rightarrow g_{(x_1,x_2)}(z) > x_1 - x_2, \\
z \in [x_1, x_2] &\Rightarrow g_{(x_1,x_2)}(z) = x_1 - x_2.
\end{aligned}$$

Therefore, $g_{(x_1,x_2)}(z)$ is minimised when $z \in [x_1, x_2]$.

We can summarise our result as

Lemma 1. If $(x_1, x_2) \in \mathbb{R}^2$ and $x_1 \leq x_2$, then the function $g_{(x_1,x_2)} : \mathbb{R} \to \mathbb{R}$ defined by

$$g_{(x_1,x_2)}(z) = |x_1 - z| + |x_2 - z|$$

is minimised when $z \in [x_1, x_2]$.

Let us now return to the main problem. First, notice that we can rewrite $g_{\mathbf{x}}$ as

$$g_{\mathbf{x}}(z) = \sum_{i=1}^{n} |x_i - z| = \sum_{i=1}^{n} \left| x_{(i)} - z \right|,$$

where $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are the reordered $x_1, x_2, \ldots, x_n$ values such that $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ (and $\left\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right\} = \{x_1, x_2, \ldots, x_n\}$). We will rewrite this sum in a different way, but will need to consider whether n is even or odd. Second, define c to be

$$c = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even,} \\ \frac{n+1}{2}, & \text{if } n \text{ is odd.} \end{cases} \tag{2}$$

Third, define for $j \in \{1, 2, \ldots, c-1\}$,

$$g_{\mathbf{x},j}(z) = \left| x_{(j)} - z \right| + \left| x_{(n+1-j)} - z \right|.$$

Using Lemma 1, $g_{\mathbf{x},j}(z)$ is minimised when $z \in \left[x_{(j)}, x_{(n+1-j)}\right]$. Fourth, we define $g_{\mathbf{x},c}(z)$:

$$n \text{ is even} : g_{\mathbf{x},c}(z) = g_{\mathbf{x},\frac{n}{2}}(z) = \left| x_{(\frac{n}{2})} - z \right| + \left| x_{(\frac{n}{2}+1)} - z \right|$$
$$n \text{ is odd} : g_{\mathbf{x},c}(z) = g_{\mathbf{x},\frac{n+1}{2}}(z) = \left| x_{(\frac{n+1}{2})} - z \right|$$

Notice again that $g_{\mathbf{x},c}(z)$ is minimised when $z \in \left[x_{(c)}, x_{(n+1-c)}\right]$, and that when n is odd this interval $\left[x_{(c)}, x_{(n+1-c)}\right]$ is the point $x_{(\frac{n+1}{2})} = \left[x_{(\frac{n+1}{2})}, x_{(\frac{n+1}{2})}\right]$, which is the sample median. Notice how for even n the sample median also minimises $g_{\mathbf{x},c}(z)$. The effort spent dealing with whether n is odd or even now allows us to write $g_{\mathbf{x}}(z)$ succinctly as

$$g_{\mathbf{x}}(z) = \sum_{j=1}^{c} g_{\mathbf{x},j}(z),$$

where c depends on n and is defined in Equation (2). Note that the intervals $\left[x_{(j)}, x_{(n+1-j)}\right]$ are nested, i.e.

$$\left[x_{(1)}, x_{(n)}\right] \supset \left[x_{(2)}, x_{(n-1)}\right] \supset \cdots \supset \left[x_{(j)}, x_{(n+1-j)}\right] \supset \cdots \supset \left[x_{(c)}, x_{(n+1-c)}\right].$$

and therefore,

$$\bigcap_{j=1}^{c} \left[x_{(j)}, x_{(n+1-j)}\right] = \left[x_{(c)}, x_{(n+1-c)}\right].$$

Let $a \in \left[x_{(c)}, x_{(n+1-c)}\right]$ (i.e. a is in the innermost interval). Then $g_{\mathbf{x},j}(a)$ is the minimum of $g_{\mathbf{x},j}(z)$, for all $j \in \{1, 2, \ldots, c\}$. For any value $a' \in \mathbb{R}$, $g_{\mathbf{x}}(a') \geq \min_z g_{\mathbf{x}}(z)$. In particular, this is true for $a \in \left[x_{(c)}, x_{(n+1-c)}\right]$, so $g_{\mathbf{x}}(a) \geq \min_z g_{\mathbf{x}}(z)$. But, we also have

$$g_{\mathbf{x}}(a) = \sum_{j=1}^{c} g_{\mathbf{x},j}(a) = \sum_{j=1}^{c} \min_z g_{\mathbf{x},j}(z) \leq \min_z \sum_{j=1}^{c} g_{\mathbf{x},j}(z) = \min_z g_{\mathbf{x}}(z),$$

showing $g_{\mathbf{x}}(a) \leq \min_z g_{\mathbf{x}}(z)$, and therefore $g_{\mathbf{x}}(a) = \min_z g_{\mathbf{x}}(z)$, which proves the result. The reason for the inequality

$$\sum_{j=1}^{c} \min_z g_{\mathbf{x},j}(z) \leq \min_z \sum_{j=1}^{c} g_{\mathbf{x},j}(z)$$

is that, in general, on the left-hand side there may be c different z values for minimizing the $g_{\mathbf{x},j}(z)$ functions, but on the right-hand side the same (minimising) z value is used for all $g_{\mathbf{x},j}(z)$ functions.

When n is odd, we must have $a = x_{(\frac{n+1}{2})}$, which is the sample median.
When n is even, any $a \in [x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}]$ minimises $g_{\mathbf{x}}(z)$; in particular the sample median $\frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}]$ is in this interval and will minimise $g_{\mathbf{x}}(z)$.

Comments: This proof is essentially an algebraic exercise. Contrast this proof with the proof of Theorem 1.8.5 in the notes (the proof was given in the Solution to Problem Sheet 9, Question 4), where that proof was dealing with expectations of random variables.

References

[1] N. C. Schwertman, A. J. Gilks, and J. Cameron. A simple noncalculus proof that the median minimizes the sum of the absolute deviations. *The American Statistician*, 44(1):38–39, 1990.