Exercise Sheet 4

The exercise sheet is subdivided into three parts:

- Part A contains warm-up questions you should do in your own time.
- You should solve all questions of Part B and put your solutions in the locker in the undergraduate common room in the Department of Statistical Science (Room number 117) assigned to your tutor group by the deadline stated on the exercise sheet. Don't forget to mark your solutions with your name and student number.

  Reminder: If you miss the deadline for reasons outside your control, for example illness or bereavement, you ***must*** submit a claim for extenuating circumstances, normally within a week of the deadline. Your home department will advise you of the appropriate procedures. For Statistical Science students, the relevant information is on the DOSSSH Moodle page.

  Your tutor will then mark one question selected at random after you handed in your solutions (mark counts towards ICA) and will return your marked solutions to you at the next tutorial at which the solutions to the part B questions will be discussed. Marks available for each part question are indicated in square brackets, e.g. [3]. Very succinct solutions to these questions will usually be made available on Moodle at 9:15am on Thursday mornings, so don't hand in late!
- Solutions to Part C questions will be made available in time for exam revision. These questions will not be discussed at the tutorials.

A: Warm-up questions

1. Write down the joint density of the $N_2\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 4\end{pmatrix}\right)$ distribution in component form. Obtain the marginal distributions of X_1 and X_2 and the conditional distributions of X_2 given $X_1 = x_1$ and X_1 given $X_2 = x_2$.

B: Answers to hand in by Thursday, 5 November, 9am

1. The Scrabble experiment carried out during the lectures to select the homework question to be marked used the following setup: A die is rolled to select the bag of Scrabble letters to sample from. If the die comes up one, bag $B = 1$ is sampled from, if the die comes up 2 or 3, bag $B = 2$ is sampled from, otherwise bag $B = 3$ is sampled from. Bag 1 contains all the Scrabble letters worth 1 point. The composition of Bag 2 is given by the following table:

Letter	Blank	G	D	B	C	M	P
Number of letters in bag	2	3	4	2	2	2	2
Points	0	2	2	3	3	3	3

The composition of Bag 3 is given by the following table:

Letter	F	H	V	W	Y	K	J	X	Q	Z
Number of letters in bag	2	2	2	2	2	1	1	1	1	1
Points	4	4	4	4	4	5	8	8	10	10

Instead of the probability of specific letters coming up, this question focuses on the number of points on the letter sampled. This is modelled as a random variable denoted[1] T with outcomes in $\{0, 1, 2, 3, 4, 5, 8, 10\}$. For each of the tasks below, carry out the required calculation and also give a clear explanation of what the result means in the context of the experiment.

(a) Compute the conditional expectation $\mathbb{E}[T|B = b]$ for each $b \in \{1, 2, 3\}$. [5]

[1] Denoting it by P would risk confusion with probabilities, so it's named T from точки.

(b) Compute the conditional variance $\mathrm{Var}(T|B = b)$ for each $b \in \{1, 2, 3\}$. [5]

(c) Compute the marginal expectation $\mathbb{E}[T]$ and variance $\mathrm{Var}(T)$ using the iterated conditional expectation and variance formulae. [5]

2. Let the joint distribution of the two continuous random variables X and Y be given by the pdf

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } x \geq -1 + (\alpha+1)y,\ x \leq 1 + y(\alpha-1),\ y \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha \in [-1, 1]$ is a parameter.

(a) For any value of $\alpha \in [-1, 1]$, the region where $f_{X,Y}(x, y)$ is positive forms a triangle. Make a sketch illustrating this triangle in the (x, y) plane for an arbitrary value of $\alpha \in [-1, 1]$. The co-ordinates of the vertices should be clearly labelled, as should the equations of the lines forming the sides of the triangle. Without carrying out any calculations of expected values, variances or covariances, decide which sign $\mathrm{Cov}(X, Y)$ will take in each of the three cases $\alpha \in \{-1, 0, 1\}$ and justify your decision. [6]

(b) Show that the marginal distributions of X and Y are given as

$$f_X(x) = \begin{cases} \frac{x+1}{\alpha+1} & \text{if } -1 \leq x < \alpha \\ 1 & \text{if } x = \alpha \\ \frac{x-1}{\alpha-1} & \text{if } \alpha < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = 2(1-y) \text{ on } y \in [0, 1].$$

and use these to compute the expected values $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ for any $\alpha \in [-1, 1]$. [6]

(c) Compute the covariance of X and Y. Express your result in terms of $\alpha \in [-1, 1]$ and compare your findings with the decisions made in part (a). `Hint:` You may find the integration easier to perform if you integrate with resepect to x first. [3]

C: Exam Practice Questions

1. Let $X_1 = I_1Y$, $X_2 = I_2Y$, where I_1, I_2 and Y are independent and I_1 and I_2 take values ± 1 each with probability $\frac{1}{2}$. Show that $\mathrm{E}(X_j) = 0$, $\mathrm{var}(X_j) = \mathrm{E}(Y^2)$ and $\mathrm{cov}(X_1, X_2) = 0$.

2. Suppose that $Y \sim N(0, 1)$ and $X|Y = y \sim N(\alpha + \beta y, \sigma^2)$. *Without initially assuming that* (X, Y) *are bivariate normal*, find the means, variances and correlation of X and Y.

3. Verify that $\mathrm{E}(X_1 + \cdots + X_p) = \mu_1 + \cdots + \mu_p$ and $\mathrm{var}(X_1 + \cdots + X_p) = \sum_{ij} \sigma_{ij}$, where $\mu_i = \mathrm{E}(X_i)$ and $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$. Suppose now that the X_i's are i.i.d.. Verify that $\bar{X}$ has mean μ and variance σ^2/p, where $\mu = \mathrm{E}(X_i)$ and $\sigma^2 = \mathrm{var}(X_i)$.