MAST20004 Probability – Assignment 4

Name: Lucas Fern

Student ID: 1080613

New completion process

- If you have a printer (or can access one), then you should print out the assignment template, handwrite your solutions into the answer spaces and then scan your assignment to a PDF file using a scanning app on your mobile phone for upload.

- If you do not have a printer (note the School strongly prefers you do have one for the final exam but you can use a second device to read the exam paper also), and you know how to annotate a PDF using an iPad/Android tablet/Graphics tablet, then annotate your answers directly onto the assignment PDF and save a copy for submission.

- If unfortunately you cannot manage to complete using either of the above methods, you may handwrite your answers as normal on blank paper and then scan for submission.

  In order to submit this way you will need to **mimic** the template by numbering your blank pages following the template and noting which answers to write on each of those pages. Remember to keep the cover page in your submission with just your name and student ID stated.

  Also you can note on pages where you run out of room that further working is appended and include additional working pages **but only at the end** after the template pages.

- Whether you complete on paper or by annotating, if you find you are unable to answer the whole question in the answer space provided then you can append additional handwritten solutions to the **end** of your templated assignment. If you do this you **MUST** make a note in the correct answer space or page for the question, warning the marker that you have appended additional remarks at the end.

- When finished, submit your assignment PDF to GradeScope below by first selecting your PDF file and then clicking on 'Upload PDF'.

Please turn over

- Also note the following

  - You can resubmit your assignment at any time up to the deadline. Please note that **ONLY** your last assignment submission will be marked.

  - Any one submission of an assignment **MUST** be in a single pdf file with pages in the correct order - other file types cannot be uploaded.

  - Please review your pdf to ensure that all pages are readable - material that cannot be read will not be marked.

- The **strict** submission deadline is **3 pm Melbourne time on Wednesday 3 June**. You have two weeks instead of the normal one week to complete this assignment. Consequently late assignments will **NOT** be accepted. We recommend you submit at least a day before the due date to avoid any technical delays.

  If there are extenuating, eg medical, circumstances, contact the Tutorial Coordinator.

- There are 5 questions, of which 2 randomly chosen questions will be marked. Note you are expected to submit answers to **all** questions, otherwise **mark penalties will apply**.

- Working and reasoning **must** be given to obtain full credit. Give clear and concise explanations. Clarity, neatness, and style count.

1. The price of a stock at the start of a trading day is \$100. The price of the stock one hour into the trading day is $100e^X$, and the price of the stock two hours into the trading day is $100e^Y$, where (X, Y) is a bivariate normal random variable with $\mathbb{E}(X) = 2$, $V(X) = 1$, $\mathbb{E}(Y) = 4$, $V(Y) = 2$, and $\text{Cov}(X, Y) = 1$.

   (a) Calculate the probability that the price of the stock one hour after the start of the day is higher than the starting price.

   $$X \sim N(2, 1)$$
   $$100\,e^X > 100 \text{ for } e^X > 1 ; \Rightarrow X > 0$$
   $$P(X > 0) = P\left(Z > \frac{0-2}{1}\right) \quad (Z \sim N(0,1))$$
   $$= P(Z > -2)$$
   $$= 0.977 \quad (\text{by tables})$$

   (b) If you buy one share of the stock at the start of the trading day and sell it after one hour, how much money would you expect to make (losses count as negative)?

   $$e^X \sim LN(2, 1)$$
   $$\therefore E(e^X) = e^{2 + \frac{1}{2}(1)}$$
   $$= e^{\frac{5}{2}}$$
   $$\Rightarrow \text{ expected value after 1h is } 100\,e^{\frac{5}{2}}$$
   $$\approx 1218.25$$
   and profit would be \$1118.25

3

(c) What is the correlation of X and Y?

$$\rho(X,Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1}{\sqrt{2}}$$

(d) Given the price of the stock two hours after the start of the day is \$100, what is the probability that the price of the stock one hour after the start of the day is higher than the starting price?

$$P(X > 0 \mid Y = 0) = \int_0^\infty p_{X\mid Y}(x \mid 0)\, dx$$

is the probability of having a positive profit after 1h given no profit after 2h.

$$(X \mid Y = y) \sim N\left(\mu_x + \rho\sigma_x \frac{(y - \mu_y)}{\sigma_y}, \sigma_x^2(1-\rho)^2\right)$$

$$(X \mid Y = 0) \sim N\left(2 + \frac{1}{\sqrt{2}}\left(\frac{-4}{\sqrt{2}}\right), \left(1 - \frac{1}{2}\right)\right)$$

$$\sim N(0, \tfrac{1}{2})$$

therefore $P(X > 0 \mid Y = 0) = \frac{1}{2}$ since the normal distribution is symmetric about its mean and $N(0, \frac{1}{2})$ has mean $\frac{1}{2}$.

2. Let X be uniform on the interval $(0,1)$ independent of Y which has probability density function $f_Y(y) = 2(1-y)$ for $0 < y < 1$. Find the probability density function of $Z = X + Y$ and sketch its graph.

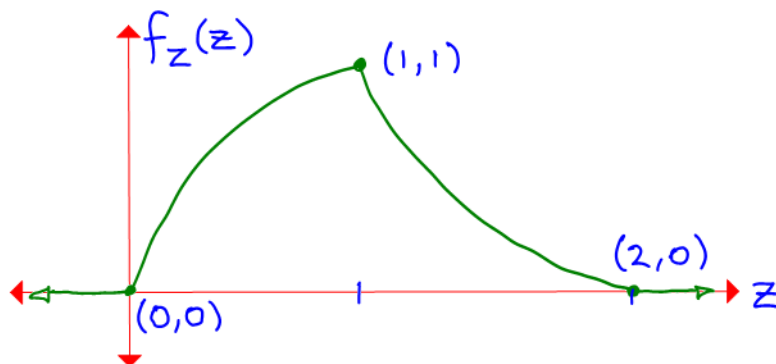$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) \, dy$$

$$= \int_0^1 f_X(z-y) f_Y(y) \, dy$$

$$f_X(z-y) = \begin{cases} 1, & 0 \le z-y \le 1 \text{ or } z-1 \le y \le z \\ 0, & \text{else} \end{cases}$$

$$\therefore f_Z(z) = \begin{cases} \int_0^z 2(1-y) \, dy, & 0 < z \le 1 \\ \int_{z-1}^1 2(1-y) \, dy, & 1 < z \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

$$\int 2(1-y) \, dy = 2(y - \tfrac{1}{2}y^2) + C$$

$$f_Z(z) = \begin{cases} 2z - z^2, & 0 < z \le 1 \\ z^2 - 4z + 4, & 1 < z < 2 \\ 0, & \text{elsewhere} \end{cases}$$

3. At a certain coffee cart, there are on average 400 paying customers per day with a standard deviation of 10. The mean amount a customer spends is \$5 with a standard deviation of \$2.

(a) If we assume the amount each customer spends is independent of the number of customers in a day, find the mean revenue per day for the cart.

let X be the number of customers per day and Y be the amount a customer spends.

$E(X) = 400$, $V(X) = 10^2$

$E(Y) = 5$, $V(Y) = 2^2$

$E(XY) = E(X)E(Y)$ when independent

$E(XY) = 2000$

so the expected revenue in a day is \$2000

(b) If in addition to the assumption in part (a), we also assume the amounts that the customers spend are all independent of each other, then what is the standard deviation of the revenue per day for the coffee cart?

let Y_i be the amount customer i spends, X is still customer count.
Let R be the revenue.
$R = Y_1 + Y_2 + \ldots + Y_X$

For $X = x$, $R = Y_1 + \ldots + Y_x$
and $E(Y_i) = 5$, $V(y) = 2^2$

$\Rightarrow E(R|X=x) = x \cdot 5$
and $V(R|X=x) = x \cdot 4$

$\therefore E(R|X) = 5X$; $V(R|X) = 4X$

$V(R) = E(V(R|X)) + V(E(R|X))$
$\qquad = E(4X) + V(5X)$
$\qquad = 4E(X) + 5^2 V(X)$
$\qquad = 4(400) + 25(100)$
$\qquad = 4100$
$\sigma_R = \sqrt{4100} = 10\sqrt{41}$

4. A box contains an unknown number of white and black balls. We wish to estimate the proportion p of white balls in the box. Let Z_n be the proportion of white balls obtained after n drawings.

   (a) Show that $\mathbb{E}(Z_n) = p$ and $V(Z_n) = \dfrac{p(1-p)}{n}$.

Let the amount of white balls picked in a sample of size n be X_n.

$X_n \sim Bi(n,p)$

$E(X_n) = np$, $V(X_n) = np(1-p)$

$Z_n = \dfrac{X_n}{n}$

$E(X_n \cdot \frac{1}{n}) = \frac{1}{n}(np) = p$

$V(X_n \cdot \frac{1}{n}) = \frac{1}{n^2}(np(1-p)) = \dfrac{p(1-p)}{n}$

(b) Use Chebyshev's Inequality to show that, for all $\epsilon > 0$,

$$P(|Z_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}.$$

$$P\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{1}{k^2}$$

$$\therefore \; P\left(\frac{|Z_n - p|}{\left(\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)} \geq k\right) \leq \frac{1}{k^2} \quad \rightarrow k^{-2} = \left(\frac{\epsilon \sqrt{n}}{\sqrt{p(1-p)}}\right)^{-2}$$

$$P\left(|Z_n - p| \geq k \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

$$\text{let } \epsilon = k \frac{\sqrt{p(1-p)}}{\sqrt{n}}, \; k = \frac{\epsilon \sqrt{n}}{\sqrt{p(1-p)}}$$

$$P(|Z_n - p| \geq \epsilon) \leq \frac{p(1-p)}{\epsilon^2 n}$$

the maximum value that the RHS of the inequality above can take is $\frac{1}{4\epsilon^2 n}$ as $p(1-p)$ is maximised at $p = \frac{1}{2}$.

Thus we can confidently say the value of the LHS must be less than $\frac{1}{4\epsilon^2 n}$ so :

$$P(|Z_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

(c) Using the result in part (b), find the smallest value of n such that, with probability greater than or equal to 0.9, the proportion Z_n in the sample will estimate p to within an accuracy of 0.05.

We want the minimum n such that

$$P(|Z_n - p| \geq 0.05) \leq 1 - 0.9$$

$$P(|Z_n - p| \geq 0.05) \leq \frac{1}{4(0.05)^2 n}$$

$$\leq \frac{100}{n}$$

$$\frac{100}{n} = 0.1 \implies n = 1000$$

we would have to sample at least 1000 balls to be 90% confident that our proportion Z_n was accurate ± 0.05.

5. Let $Y = X_1 + X_2 + \ldots + X_N$ where $N \overset{d}{=} \text{Bi}(n, p)$, $X_i \overset{d}{=} \text{Bi}(m, q)$, and $N, X_1, X_2, \ldots, X_N$ are independent.

(a) Find $P_{Y|N}(z)$, the conditional probability generating function of Y given N, and state the values of z for which it is defined.

$$P_{Y|N}(z) = E(z^Y | N)$$

$$\quad\quad\quad\quad\quad \curvearrowright Y \sim \text{Bi}(nm, q)$$

$$E(z^Y | N = n) = \sum_{y=0}^{nm} P_{Y|N}(y|n) z^y$$

$$= \sum_{y=0}^{nm} \binom{nm}{y} q^y (1-q)^{nm-y} z^y$$

$$= (qz + 1 - q)^{nm} \quad \text{(binomial formula)}$$

$$\implies E(z^Y | N) = (qz - q + 1)^{Nm}$$

$$P_{Y|N}(z) = (qz - q + 1)^{Nm}; \quad z \in \mathbb{R}$$

$$z \in \mathbb{R} \text{ since } N \in \mathbb{N} \cup \{0\}, \, p \in [0, 1]$$

(b) Find $P_Y(z)$, the probability generating function of Y, and state the values of z for which it is defined.

$$P_Y(z) = E(E(z^Y | N)) = E((qz - q + 1)^{mN})$$

$$= P_N((qz - q + 1)^m)$$

$$P_N(\alpha) = \sum_{\beta=0}^{\infty} P_N(\beta) \cdot \alpha^\beta = \sum_{\beta=0}^{n} \binom{n}{\beta} (\alpha p)^\beta (1-p)^{n-\beta}$$

$$= (p\alpha - p + 1)^n \quad \text{(binomial formula)}$$

$$\therefore P_N((qz - q + 1)^m) = (p(qz - q + 1)^m - p + 1)^n$$

$$P_Y(z) = (p(qz - q + 1)^m - p + 1)^n$$

$$\text{for } z \in \mathbb{R}$$

11

(c) Using $P_Y(z)$, evaluate $\mathbb{E}(Y)$

$$E(Y) = P_Y'(1)$$

$$P_Y'(z) = n\left(p(qz-q+1)^m - p + 1\right)^{n-1}\left(mp^2(qz-q+1)\right)$$

$$P_Y'(1) = n(p-p+1)^{n-1}(mp^2)$$

$$= nmp^2$$