MAST20004 Probability – Assignment 4

Name:

Student ID:

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 $\mathrm{Cov}(X, Y) = 1$.

$(X,Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ $X \sim N(\mu_X, \sigma_X^2)$ $Y \sim N(\mu_Y, \sigma_Y^2)$

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

let A be the event that the price of the stock one hour after the start of the day is higher than the starting price

$P(A) = P(100e^X > 100) = P(e^X > 1) = P(X > \log_e 1) = P(X > 0)$

$\mu_X = 2$ $\mu_Y = 4$ $\sigma_X = 1$ $\sigma_Y = 2$, $\rho = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1}{1 \cdot \sqrt{2}}$

$(X, Y) \stackrel{d}{=} N_2(2, 4, 1, 2, \frac{1}{\sqrt{2}})$

$X \stackrel{d}{=} N(2, 1^2)$.

so $P(X > 0) = P(\frac{X - \mu_X}{\sigma_X} > \frac{0 - \mu_X}{\sigma_X})$

$= P(Z > \frac{0-2}{1})$

$= P(Z > -2) = 1 - P(Z \le -2)$

use table

$= 1 - 0.0228$

$= 0.9772$

$X \sim N(\mu_X, \sigma_X^2)$

$Y = e^X \sim LN(\mu_X, \sigma_X^2)$

(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)?

$M_X(t) = E(e^{tX})$

$E(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0}$

the expected money to make is $E(100e^X) - 100$

$E(100e^X) = 100E(e^X)$.

if $X \stackrel{d}{=} N(2, 1)$

$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} = e^{2t + \frac{1}{2}t^2}$

$M_X(1) = E(e^X) = e^{2 + \frac{1}{2}} = e^{\frac{5}{2}}$

$E(100e^X) = 100M(1) = 100e^{\frac{5}{2}}$

so $E(100e^X) - 100 = 100(e^{\frac{5}{2}} - 1) = 1118.249$

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

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$= \frac{1}{\sqrt{1 \times 2}}$$

$$= \frac{1}{\sqrt{2}}$$

(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?

$$X \stackrel{d}{=} N(2,1)$$
$$Y \stackrel{d}{=} N(4,(\sqrt{2})^2)$$

$$P(100e^X > 100 \mid 100e^Y = 100)$$

$$= P(e^X > 1 \mid e^Y = 1)$$

$$= P(X > 0 \mid Y = 0)$$

$$= P\left(\frac{X-2}{1} > \frac{0-2}{1} \,\middle|\, \frac{Y-4}{\sqrt{2}} = \frac{0-4}{\sqrt{2}}\right)$$

Find conditional distr $X \mid Y=0$

$$= P(X_s > -2 \mid Y_s = -2\sqrt{2})$$

$$(X \mid Y=y) \stackrel{d}{=} N(\rho y, 1-\rho^2)$$

$$(X_s \mid Y_s = y) \stackrel{d}{=} N(\rho y, 1-\rho^2)$$

$$(X_s \mid Y_s = -2\sqrt{2}) \stackrel{d}{=} N\left(-2\sqrt{2}\cdot\frac{1}{\sqrt{2}}, 1-\frac{1}{2}\right)$$

let $V \stackrel{d}{=} N\left(-2, \frac{1}{2}\right)$

so $P(X_s > -2 \mid Y_s = -2\sqrt{2})$

$$= P(V > -2) = P\left(Z > \frac{-2-(-2)}{\sqrt{\frac{1}{2}}}\right)$$

$$= P(Z > 0)$$

$$= \frac{1}{2}$$

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.

$X \overset{d}{=} U(0,1)$ $\quad f_Y(y) = 2(1-y)$ for $0<y<1$

$f_X(x) = 1$ for $0<x<1$

$Z = X+Y \in (0,2)$.

$F_Z(z) = F_{X+Y}(z) = P(X+Y \le z)$

$= \int_y P(X+Y \le z \mid Y=y) \cdot P(Y=y)\,dy$

$= \int_y P(X \le z-y \mid Y=y) \cdot f_Y(y)\,dy$

since X, Y independent $= \int_y P(X \le z-y)\, f_Y(y)\,dy$

$= \int_y F_X(z-y)\, f_X(y)\,dy$

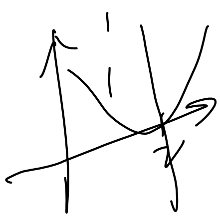
differentiate with respect to z

$f_Z(z) = \int_y f_X(z-y)\, f_X(y)\,dy$

for range of y:

since $f_X(x) = 1$ for $0<x<1$, $f_Y(y) = 2(1-y)$ for $0<y<1$. ①

so $f_X(z-y) = 1$ for $0 < z-y < 1 \Rightarrow z-1 < y < z$ ②



combine ① and ② $\max(0, z-1) < y < \min(z, 1)$

if $z-1 \ge 0$, $z \ge 1 \Rightarrow 1 \le z < 2$

$z-1 < y < 1$

$f_Z(z) = \int_{z-1}^{1} 1 \cdot 2(1-y)\,dy$

$= \int_{z-1}^{1} 2-2y\,dy$

$= \left[2y - y^2\right]_{y=z-1}^{y=1}$

$= 2-1-2(z-1)+(z-1)^2$

$= 1-2z+2+z^2-2z+1$

$= z^2-4z+4 = (z-2)^2$

if $z-1<0 \Rightarrow 0<z<1$

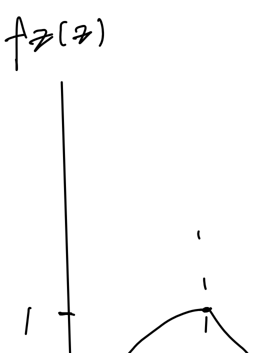
$0 < y < z$

$f_Z(z) = \int_0^z 2(1-y)\,dy$

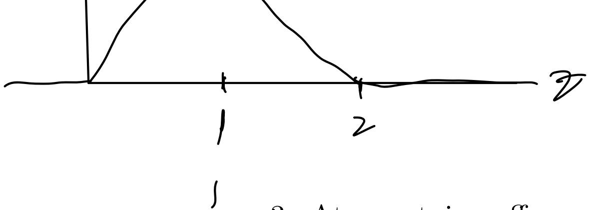
$= \int_0^z 2-2y\,dy$

$= \left[2y-y^2\right]_{y=0}^{y=z}$

$= 2z - z^2$



so. $f_Z(z) = \begin{cases} 2z - z^2 & 0<z<1 \\ (z-2)^2 & 1 \le z \le 2 \end{cases}$



0 else

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 amount per customer spend
let N be the customer per day
and R be the revenue per day

$$R = \sum_{i=1}^{N} X_i$$

X_i's are independent identical with mean 5, standard deviation 2
for conditional expectation $E(R) = E(E(R|N))$

$$= E\left(E\left(\sum_{i=1}^{N} X_i \middle| N\right)\right)$$

since for $N=n$, $R = X_1 + X_2 + \cdots + X_n$

$$E(R|N=n) = n \cdot E(X_i) = 5n$$

so $E(R|N) = 5N$.

$$E(E(R|N)) = E(5N) = 5E(N)$$

since for N, the mean is 400

so $E(R) = 5 \times 400 = 2000$

(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?

from the conditional variance.

$V(R) = V(E(R|N)) + E(V(R|N))$

given $N=n$ $\quad R = \sum_{i=1}^{N} X_i = X_1 + \cdots + X_n$

$E(R|N=n) = n \cdot E(X_i) = 5n$

$V(R|N=n) = n\,V(X_i) = 2^2 \cdot n = 4n$

Therefore

$E(R|N) = 5N \quad V(R|N) = 4N$

so $V(R) = V(E(R|N)) + E(V(R|N))$

$= V(5N) + E(4N)$

$SD_N = 10$

$V(N) = 10^2 = 100$

$= 5^2 V(N) + 4E(N)$

$= 5^2 \cdot 10^2 + 4 \cdot 400$

$= 4100$

$sd(R) = \sqrt{V(R)} = 64.031$

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. with replacement

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

let X be the number of white balls obtained after n drawings

$P(\text{draw a white ball}) = p$

$X \overset{d}{=} Bi(n, p)$. then $Z_n = \frac{X}{n}$

$E(Z_n) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$

$V(Z_n) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{\cancel{n}p(1-p)}{n^{\cancel{2}}} = \frac{p(1-p)}{n}$

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

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

~~for~~ Z_n has mean p and variance $\frac{p(1-p)}{n}$

so $sd(Z_n) = \sqrt{Var(Z_n)} = \sqrt{\frac{p(1-p)}{n}}$

from Chebyshev's Inequality

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

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

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

$$\epsilon^2 = k^2 \cdot \frac{p(1-p)}{n}$$

$$\frac{1}{k^2} = \frac{p(1-p)}{\epsilon^2 n}$$

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

since $p(1-p) = p - p^2 = -(p^2 - p + \frac{1}{4} - \frac{1}{4})$

$$= -(p - \tfrac{1}{2})^2 + \tfrac{1}{4}$$

$$\leq \tfrac{1}{4}$$

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

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.

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

$$\epsilon = 0.05$$

$$1 - \frac{1}{4n\epsilon^2} = 0.9$$

$$\frac{1}{4n\epsilon^2} = \frac{1}{10}$$

$$4n\epsilon^2 = 10$$

$$n = \frac{10}{4\epsilon^2} = \frac{10}{4 \cdot (0.05)^2}$$

$$= 1000$$

$\therefore$ smallest value of n is 1000

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.

$Y = X_1 + \cdots + X_N$ where $N \overset{d}{=} Bi(n,p)$

$Y | N=n \Rightarrow Bi(mn, q)$

$\therefore Y|N = Bi(mN, q)$

since if we fix $N=n$, sum of n independent, identical binomial distribution X_i, will also follow a binomial distribution $Bi(mn, q)$

$Y|N$ is a random variable that depend on random variable N.

$$P_{Y|N}(z) = E[z^Y | N=n) = \sum_{y=0}^{mN} z^y \binom{mN}{y} q^y (1-q)^{mN-y} = \sum_{y=0}^{mN} \binom{mN}{y} (zq)^y (1-q)^{mN-y}$$

$$= (1+zq-q)^{mN}$$

defined for all $z \in R$

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

$$P_Y(z) = E(z^Y) = E\Big(E(z^Y | N)\Big)$$

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

$$= E\Big[\big[(1+zq-q)^m\big]^N\Big]$$

let $(1+zq-q)^m = t$

$$= E(t^N)$$

$$= P_N\big((1+zq-q)^m\big)$$

since $N \overset{d}{=} Bi(n,p)$

$$= \sum_{r=0}^{n} t^r \binom{n}{r} p^r (1-p)^{n-r}$$

$$= (pt + 1-p)^n$$

$$= (1-p+p(1+zq-q)^m)^n$$

defined for $z \in R$.

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

$$P_Y(z) = \left(1-p+p(1+zq-q)^m\right)^n$$

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

$$E(Y) = P_Y'(1) = n\left(1-\cancel{p}+\cancel{p}(1)^m\right)^{n-1} \cdot mp(1+\cancel{q}-\cancel{q})^{m-1} \cdot q$$

$$= n \cdot 1 \cdot mp \cdot 1 \cdot q$$

$$= mnpq$$

$$\therefore \quad E(Y) = mnpq$$