MAST20004 Probability – Assignment 3

Name: Lucas Fern

Student ID: 1080613

- ## New completion process

  Note this assignment is being handled using a similar process to that now planned for the final exam so you can start to become familiar with it.

  To complete this assignment, you need to write your solutions into the blank answer spaces following each question in this assignment pdf.

  If you have a printer (or can access one), then you must print out the assignment template and handwrite your solutions into the answer spaces.

  If you do not have a printer (**NB remember to contact us as soon as possible to discuss the situation if you believe you may not be able to get one - see the Canvas exam page for more background**), but you can figure out how to annotate a PDF using an iPad/Android tablet/Graphics tablet or using Adobe Acrobat, then annotate your answers directly onto the assignment PDF and save a copy for submission.

  Failing both of these methods, you may handwrite your answers as normal on blank paper and then scan for submission (but note that you will thereby miss valuable practice for the exam process).

  Scan your assignment to a PDF file using your mobile phone then upload by going to the Assignments menu on Canvas and submit the PDF to the GradeScope tool by first selecting your PDF file and then clicking on 'Upload PDF'.

- The **strict** submission deadline is **3 pm Melbourne time on Friday 15 May**. 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 4 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. A company that manufactures screwdrivers sells them in cartons of 100. It is historically known that 2% of the screwdrivers manufactured by the company are defective.

   (a) Write an exact expression for the probability that a carton has more than 3 defective screwdrivers in it, and evaluate this probability.

   Let X be the amount of faulty screwdrivers in a carton.

   $X \sim Bi(100, 0.02)$

   $P(X>3) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3)$

   $P(X=0) = (0.98)^{100} = 0.133$

   $P(X=1) = \binom{100}{1}(0.02)(0.98)^{99} = 0.271$

   $P(X=2) = \binom{100}{2}(0.02)^2(0.98)^{98} = 0.273$

   $P(X=3) = \binom{100}{3}(0.02)^3(0.98)^{97} = 0.182$

   $\therefore P(X>3) \approx 0.141$

   (b) Approximate this same probability using the Poisson distribution.

   $Bi(100, 0.02) \approx Pn(2)$

   If $X \sim Pn(2)$

   $$p_X(x) = \frac{e^{-2}(2^x)}{x!}$$

   $\therefore P(X>3) = 1 - P(X=3) - P(X=2) - P(X=1) - P(X=0)$

   $= 1 - e^{-2}\left(\frac{2^3}{3!} + \frac{2^2}{2!} + \frac{2}{1} + 1\right)$

   $= 1 - e^{-2}\left(\frac{19}{3}\right)$

   ≈ 0.143

(c) Approximate this same probability using the normal distribution.

$$Bi(n,p) \approx N(np, np(1-p))$$

$$\Rightarrow Bi(100, 0.02) \approx N(2, 2(0.98))$$

For $X \sim N(2, 2(0.98))$

$$\begin{aligned} P(X>3) &= 1 - P(X \leqslant 3) \\ &= 1 - P\left(Z \leqslant \frac{3-2}{\sqrt{1.96}}\right) \quad (Z \sim N(0,1)) \\ &= 1 - P(Z \leqslant 0.714) \\ &\approx 1 - 0.761 \\ &\approx 0.238 \end{aligned}$$

(d) Which of the approximations in parts (b) and (c) is better? Why is this the case?

The poisson approximation is better since the normal distribution is not great for approximating when p is close to 0 or 1 and the Poisson distribution is.

2. Let $X \stackrel{d}{=} \text{Beta}(\alpha, \beta)$ where $\alpha, \beta > 0$.

(a) Let $Y = 1 - X$. Find the probability density function of Y and name the distribution.

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & , 0 \le x \le 1 \\ 0 & , \text{elsewhere} \end{cases}$$

$$F_X(a) = P(X \le a) = \int_0^a f(x)\,dx$$

$$F_Y(y) = P(1-X \le y) = P(X \ge 1-y) = 1 - P(X < 1-y)$$
$$= 1 - F_X(1-y)$$
$$= 1 - \int_0^{1-y} f(x)\,dx$$

$$\therefore f_Y(y) = -f(1-y) \cdot \frac{d}{dy}(1-y)$$
$$= \frac{(1-y)^{\alpha-1}\, y^{\beta-1}}{B(\alpha,\beta)}$$

This is the pmf of Beta(β, α)

$\therefore Y \sim \text{Beta}(\beta, \alpha)$

(b) Let $\alpha = 1$. Let $Z = -\log Y$. Find the probability density function of Z and name the distribution.

$Y \sim \text{Beta}(\beta, 1)$

$$f_Y(y) = \begin{cases} \dfrac{x^{\beta-1}}{B(\beta,1)} & , 0 \leqslant x \leqslant 1 \\ 0 & , \text{elsewhere} \end{cases}$$

$$B(\beta, 1) = \frac{\Gamma(\beta)}{\Gamma(\beta+1)}$$

$$F_Z(z) = P(Z \leqslant z) = P(-\log Y \leqslant z)$$
$$= P(Y \geqslant e^{-z})$$
$$= 1 - P(Y < e^{-z})$$

$$\therefore f_Z(z) = -\left(e^{-z(\beta-1)} \cdot \frac{\Gamma(\beta+1)}{\Gamma(\beta)}\right)(-e^{-z})$$
$$= e^{-z\beta} \cdot \frac{\Gamma(\beta+1)}{\Gamma(\beta)}$$

$$\Gamma(\beta+1) = \beta \cdot \Gamma(\beta) \implies \frac{\Gamma(\beta+1)}{\Gamma(\beta)} = \beta$$

$$\therefore f_Z(z) = \begin{cases} \beta e^{-\beta z} & , z \geqslant 0 \\ 0 & , z < 0 \end{cases}$$

$\boxed{Z \sim \exp(\beta)}$

(c) How can a realisation of Y (with $\alpha = 1$) be generated from a realisation of Z?

$Z = -\log Y$

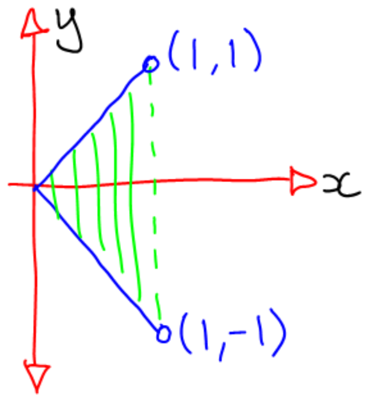
$\therefore Y = e^{-Z}$

is a realsation of Y where Z is a realisation of Z.

3. The joint probability density function of (X, Y) is given by

$$f_{(X,Y)}(x, y) = \begin{cases} c(x+y), & 0 < x < 1,\ -x < y < x \\ 0, & \text{otherwise} \end{cases}.$$

(a) Find the constant c. (*Hint: Draw a graph of the domain of* $f_{(X,Y)}(x, y)$.)



$$\int_0^1 \int_{-x}^{x} c(x+y)\,dy\,dx = 1$$

$$c\int_0^1 \left[xy + \frac{y^2}{2}\right]_{-x}^{x} dx = 1$$

$$c\int_0^1 \left(\frac{3}{2}x^2 - \left(-\frac{1}{2}x^2\right)\right) dx = 1$$

$$c\int_0^1 2x^2\,dx = 1$$

$$c\left[\frac{2}{3}x^3\right]_0^1 = 1$$

$$c\left(\frac{2}{3}\right) = 1$$

$$c = \frac{3}{2}$$

(b) Derive the marginal probability density function of X, f_X.

$$f_X(x) = \int_{-x}^{x} \frac{3}{2}(x+y)\,dy \quad ; 0 < x < 1$$

$$= \frac{3}{2}\left[xy + \frac{1}{2}y^2\right]_{-x}^{x}$$

$$= \frac{3}{2}\left(\left(\frac{3}{2}x^2\right) - \left(-\frac{1}{2}x^2\right)\right)$$

$$f_X(x) = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

(c) Derive the conditional probability density function of Y given $X = x$.

$$f_{(Y|X=x)}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

$$= \frac{\frac{3}{2}(x+y)}{3x^2}$$

$$= \begin{cases} \frac{x+y}{2x^2}, & 0<x<1, -x<y<x \\ 0, & \text{elsewhere} \end{cases}$$

(d) Are X and Y independent? Justify your answer.

$$f_Y(y) = \int_{|y|}^{1} \frac{3}{2}(x+y)\, dx$$

$$= \int_{-y}^{1} \frac{3}{2}(x+y)dx + \int_{y}^{1} \frac{3}{2}(x+y)dx$$

$$= \frac{3}{2}\left(\left(\frac{1}{2}+y\right)-\left(\frac{y^2}{2}+y^2\right)\right) + \frac{3}{2}\left(\left(\frac{1}{2}+y\right)-\left(\frac{y^2}{2}-y^2\right)\right)$$

$$= \frac{3}{2}\left(2\left(\frac{1}{2}+y\right)-y^2\right)$$

$$f_Y(y) = \begin{cases} \frac{3}{2}(1+2y-y^2), & -1<y<1 \\ 0, & \text{otherwise} \end{cases}$$

$f_Y(y) \neq f_{(Y|X)}(y|x)$ from part c, so X and Y are dependant.

(e) Calculate $\mathbb{P}\left(X \geq \frac{1}{2} \,|\, Y \leq 0\right)$.

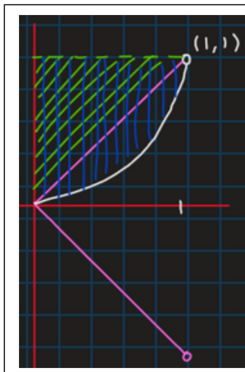
$$P(X \geq \tfrac{1}{2} \mid Y \leq 0) = \frac{P(X \geq \frac{1}{2}, Y \leq 0)}{P(Y \leq 0)}$$

$$= \frac{\int_{\frac{1}{2}}^{1} \int_{-x}^{0} \frac{3}{2}(x+y)\, dy\, dx}{\int_{0}^{1} \int_{-x}^{0} \frac{3}{2}(x+y)\, dy\, dy}$$

$$= \frac{\int_{\frac{1}{2}}^{1} \left(\frac{3}{4}x^2\right) dx}{\int_{0}^{1} \left(\frac{3}{4}x^2\right) dx}$$

$$= \left(\frac{7}{32}\right) \div \left(\frac{1}{4}\right) = \frac{7}{8}$$

(f) Calculate $\mathbb{P}(X^2 < Y)$.



$$P(X^2 < Y) = P(X^2 < Y < X)$$

$$= \int_0^1 \int_{x^2}^{x} \left(\frac{3}{2}(x+y)\right) dy\, dx$$

$$= \frac{3}{2} \int_0^1 \left[xy + \frac{1}{2}y^2\right]_{y=x^2}^{y=x} dx$$

$$= \frac{3}{2} \int_0^1 \left(-\frac{x^4}{2} - x^3 + \frac{3}{2}x^2\right) dx$$

$$= \frac{3}{2}\left[\frac{-x^5}{10} - \frac{x^4}{4} + \frac{x^3}{2}\right]_0^1$$

$$= \frac{3}{2}\left(\frac{-1}{10} - \frac{1}{4} + \frac{1}{2}\right)$$

$$= \frac{9}{40}$$

4. (a) Let X be a Cauchy random variable with $m = a = 2$, that is $X \stackrel{d}{=} \mathrm{C}(2,2)$. Write down the expression for the probability density function of X.

$$f_X(x) = \frac{1}{\pi} \frac{2}{4 + (x-2)^2} \quad ; x \in \mathbb{R}$$

(b) Derive the distribution function F_X.

$$F_X(x) = \int \frac{1}{\pi} \frac{2}{4+(x-2)^2}\, dx$$

let $u = x - 2$; $\frac{du}{dx} = 1$

$$F_X(x) = \frac{2}{\pi} \int \frac{1}{2^2 + u^2}\, du$$

$$= \frac{2}{\pi} \cdot \frac{1}{2} \arctan\left(\frac{u}{2}\right) + C$$

$$= \frac{1}{\pi} \arctan\left(\frac{x-2}{2}\right) + \frac{1}{2} \quad , x \in \mathbb{R}$$

$C = \frac{1}{2}$ since $\lim_{x \to \infty} \left(\frac{1}{\pi} \arctan\left(\frac{x-2}{2}\right)\right) = \frac{1}{2}$ and the cdf $\to 1$

(c) Find an appropriate function ψ so that if U is a uniform random variable on the interval $(0, 1)$, then $X = \psi(U)$.

$$F_X(x) = P(\psi(U) \le x)$$

$$= P(U \le \psi^{-1}(x)) = P_U(\psi^{-1}(x))$$

$$U \sim R(0,1) \Rightarrow P_U(\psi^{-1}(x)) = \psi^{-1}(x)$$

$$\psi^{-1}(x) = F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x-2}{2}\right) + \frac{1}{2}$$

$$\psi(x) = 2\tan\left(\pi\left(x - \frac{1}{2}\right)\right) + 2$$

(d) Use part (c) and the `rand` command in Matlab to simulate 10,000 independent realisations of X and compute the sample mean. Do this 19 more times so you have a total of 20 sample means.

Store the 20 sample means in a vector called `sample_mean_vec_Cauchy`. Evaluate the mean and standard deviation of the sample means using the `mean` and `std` commands in Matlab.

Write down the mean and standard deviation of the 20 sample means in the box below.

mean = 2.406 , std = 9.769

(e) The `randn` command in Matlab generates a realisation of the standard normal random variable $Z \stackrel{d}{=} \mathrm{N}(0, 1)$. Use the `randn` command to simulate 10,000 independent realisations of the normal random variable $Y \stackrel{d}{=} \mathrm{N}(2, 9)$ and compute the empirical mean. Do this 19 more times so you have a total of 20 sample means.

Store the 20 sample means in a vector called `sample_mean_vec_Normal`. Evaluate the mean and standard deviation of the sample means using the `mean` and `std` commands in Matlab.

Write down the mean and standard deviation of the 20 sample means in the box below.

mean = 1.994 , std = 0.0664

(f) Explain the difference between your observations in parts (d) and (e).

You should repeat the procedures in parts (d) and (e) a few times so you get an idea of what's going on.

The sample means of both distributions are quite similar, we expect N(2,9) to have mean 2, and although the cauchy distribution has no defined mean or variance the maximum probabilities of C(2,2) do occur around 2. The variance of the normal dist. is much lower as the tails are small relative to C(2,2), meaning occasionally a very large or small sample mean is recorded for C(2,2).