THE UNIVERSITY OF MELBOURNE

Student ID:

The University of Melbourne

Semester 1 Exam *Solutions* — *June, 2016*

School of Mathematics and Statistics

MAST20006 Probability for Statistics

Exam Duration: 3 Hours

Reading Time: 15 Minutes

This paper has 9 pages

Authorised materials:
This is a closed book exam.
Hand-held electronic calculators may be used, provided that all memories and programs are cleared.

Instructions to Invigilators:
Sixteen-page script books shall be supplied to each student.
Students may not take this paper with them at the end of the exam.

Instructions to Students:
This paper has **10** questions. Formula sheet is given at the end of this paper.
Attempt as many questions, or parts of questions, as you can.
Questions carry marks as shown in the brackets after the questions statements.
The total number of marks available for this examination is **100**.
Working and/or reasoning must be given to obtain full credit.

1. Among a class of students who enroll in an undergraduate probability subject, it is known that 40% of them (the first group) regularly attend the lectures and tackle all tutorial and lab questions, 35% of them (the second group) regularly attend lectures but rarely work on tutorial and lab questions, and the rest 25% (the third group) regularly miss lectures and do not undertake self-learning. By the end of semester it is known that 98% of the first group of students pass the subject, 70% of the second group pass the subject, while only 40% of the third group pass the subject.

   (a) What is the probability that a student randomly selected from the class will pass the subject? [3]

   - $P(\textit{passing the subject}) = 0.4 \times 0.98 + 0.35 \times 0.70 + 0.25 \times 0.40 = 0.737.$

   (b) A student from the class is known to have passed the subject, what is probability that he or she is from the third group? [3]

   - $P(\textit{the third group} \mid \textit{passed the subject}) = \dfrac{0.25 \times 0.40}{0.737} = 0.1357.$

   (c) Three students randomly selected from the class all have passed the subject. What is the conditional probability that none of them are from the first group? [3]

   - $P(\textit{none from the first group} \mid \textit{all passed the subject})$
   $= \dfrac{(0.35 \times 0.70 + 0.25 * 0.40)^3}{0.737^3} = \dfrac{0.345^3}{0.737^3} = 0.1026.$

2. A jar contains 5 green jelly beans and 5 purple jelly beans.

   (a) Suppose one jelly bean is to be selected at random. What is the value of p, the probability that a purple jelly bean is selected? [1]

   - $p = \frac{5}{5+5} = 0.5.$

   (b) Suppose 9 jelly beans are to be selected at random **with replacement**. Let X be the number of purple jelly beans in the selected. Calculate $P(X = 1)$, the probability that there is exactly one purple jelly bean in the 9 selected. [2]

   - $P(X = 1) = \binom{9}{1} \times 0.5^1 \times 0.5^8 = \frac{9}{512} = 0.0176.$

   (c) Now suppose 5 beans are to be selected at random **without replacement**. Let Y be the number of purple jelly beans in the selected. Calculate $P(Y = 1)$, the probability that there is only one purple jelly bean in the 5 selected. [3]

   - $P(Y = 1) = \frac{\binom{5}{1}\binom{5}{4}}{\binom{10}{5}} = \frac{25}{252} = 0.0992.$

3. The probability of suffering a side effect from a certain flu vaccine is 0.003. Suppose 1000 people are to be inoculated. Let X be the number of people who will suffer the side effect. Note that X would follow a binomial distribution.

   (a) Find $P(X \leq 1)$. [2]

   - $P(X \leq 1) = P(X = 0) + P(X = 1)$
     $= (1 - 0.003)^{1000} + 1000 \times 0.003 \times 0.997^{999} = 0.1986997.$

   (b) A binomial distribution $b(n, p)$ can be approximated by a Poisson($\lambda = np$) distribution if p is small and n large. Use this result to approximate the probability in part (a) by a Poisson probability. [2]

   - $X \overset{d}{\approx}$ *Poisson*$(\lambda = 3)$*, thus* $P(X \leq 1) \approx e^{-3} + 3e^{-3} = 0.1991483.$

   (c) The probability in part (a) may also be approximated by a normal probability based on the central limit theorem. Give a normal approximation (using the continuity correction, and expressed in terms of the standard normal cdf $\Phi(\cdot)$ only) to $P(X \leq 1)$. [3]

   - *By CLT,* $X \overset{d}{\approx} N(\mu = 3, \sigma^2 = 2.991).$
   - *So* $P(X \leq 1) \approx P\left(Z \leq \frac{1+0.5-3}{\sqrt{2.991}}\right) = P(Z \leq -0.8673) = \Phi(-0.8673).$
     (=0.1928813, a poor approximation; but not required for this question.)

4. Police are to conduct random breath testing on drivers on a busy road one Friday evening. Suppose 3% of the drivers drink and drive at the time. Let X be the number of drivers that police need to test to get the first case of drinking and driving. Let Y be the number of drivers tested to get 3 such cases.

   (a) Name the probability distribution and specify the value of any parameter(s) for each of the two random variables X and Y. [2]

   - $X \overset{d}{=}$ *Geometric*$(p = 0.03)$, $\quad Y \overset{d}{=}$ *Negative binomial*$(r = 3, p = 0.03)$.

   (b) What is the probability that at least 4 drivers are to be tested to get the first drinking and driving case? [2]

   - $P(X \geq 4) = \sum_{k=4}^{\infty} 0.03 \times 0.97^{k-1} = 1 - 0.03(1 + 0.97 + 0.97^2) = 0.97^3 = 0.9127.$

   (c) What is the probability that exactly 30 drivers are to be tested to get 3 drinking and driving cases? [2]

   - $P(Y = 30) = \binom{29}{2} \times 0.03^3 \times 0.97^{27} = 0.004816445.$

   (d) On average, how many drivers do police need to test to get 3 cases of drinking and driving? [1]

   - $E(Y) = \frac{r}{p} = \frac{3}{0.03} = 100.$

   (e) Find $P(Y > 50)$. [3]

   - *Let* $Z \overset{d}{=} b(50, 0.03)$*. Then using the relation between* $b(n,p)$ *and* $Nb(r,p)$*,*
   - $P(Y > 50) = P(Z \leq 2) = 0.97^{50} + 50 \cdot 0.03 \cdot 0.97^{49} + \binom{50}{2} \cdot 0.03^2 \cdot 0.97^{48} = 0.2180654 + 0.3372145 + 0.2555182 = 0.8107981.$

5. Let X be a continuous random variable with probability density function (pdf)

$$f(x) = \frac{1}{24}(x+5), \quad -4 < x < 2.$$

(a) Find the cumulative distribution function (cdf) of X. [2]

- *For* $-4 < x < 2$, $P(X \le x) = \int_{-4}^{x} \frac{1}{24}(t+5)dt = \frac{1}{48}(t+5)^2 \mid_{-4}^{x} = \frac{(x+5)^2-1}{48}$.
- *Hence the cdf of* X *is* $F(x) = \begin{cases} 0 & \text{if } x \le -4 \\ \frac{(x+5)^2-1}{48} & \text{if } -4 < x < 2 \\ 1 & \text{if } x \ge 2 \end{cases}$

(b) Let X_1 and X_2 be two independent random variables each having the pdf $f(x)$ given above. Define $W = \max\{X_1, X_2\}$. Find the 25-th percentile of W. [3]

- $P(W \le w) = P(\max\{X_1, X_2\} \le w) = P(X_1 \le w)P(X_2 \le w)$. *Using this property the cdf of* W *can be found to be*

$$F_W(w) = \begin{cases} 0 & \text{if } w \le -4 \\ \left[\frac{(w+5)^2-1}{48}\right]^2 & \text{if } -4 < w < 2 \\ 1 & \text{if } w \ge 2 \end{cases}$$

- *Solve* $\frac{1}{4} = F_w(\pi_{0.25}) = \left[\frac{1}{48}(\pi_{0.25}+5)^2 - \frac{1}{48}\right]^2$, *and it follows* $\pi_{0.25} = 0$.

(c) Consider the transformation $Y = X^2$ of X.

i. Is this transformation one-to-one? Find the support of Y. [2]

- *This is not one-to-one. The support of* Y *is* $0 \le y < 16$.

ii. Derive the cdf of Y. [3]

- *For* $0 \le y \le 4$,
  $G(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$
  $= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{24}(x+5)dx = \frac{1}{48}[(5+\sqrt{y})^2 - (5-\sqrt{y})^2] = \frac{5}{12}\sqrt{y}$.
- *For* $4 < y < 16$,
  $G(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le 2)$
  $= \int_{-\sqrt{y}}^{2} \frac{1}{24}(x+5)dx = \frac{49}{48} - \frac{1}{48}(5-\sqrt{y})^2 = \frac{5}{24}\sqrt{y} - \frac{1}{48}y + \frac{1}{2}$.
- *So the cdf of* Y *is*

$$G(y) = \begin{cases} 0, & y < 0, \\ \frac{5}{12}\sqrt{y}, & 0 \le y \le 4, \\ \frac{5}{24}\sqrt{y} - \frac{1}{48}y + \frac{1}{2}, & 4 < y < 16, \\ 1, & y \ge 16. \end{cases}$$

iii. Find the pdf of Y. [2]

-

$$g(y) = G'(y) = \begin{cases} 0, & y < 0, \\ \frac{5}{24\sqrt{y}}, & 0 \le y \le 4, \\ \frac{5}{48\sqrt{y}} - \frac{1}{48} & 4 < y < 16, \\ 0, & y \ge 16. \end{cases}$$

6. Let X_1, X_2, X_3 be independent Binomial$(n = 2, p = \frac{1}{2})$ random variables. Define $Y_1 = X_1 + X_3$ and $Y_2 = X_2 + X_3$.

(a) Find the value of $\text{Cov}(Y_1, Y_2)$. [2]

- $\text{Cov}(Y_1, Y_2) = \text{Cov}(X_1 + X_3, X_2 + X_3) = \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \text{Cov}(X_3, X_2) + \text{Cov}(X_3, X_3) = \text{Var}(X_3) = np(1-p) = \frac{1}{2}$.

(b) Use Chebyshev's inequality $P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$ to find a lower bound for $P(|Y_1 - 2| < 2)$. [1]

- *$E(Y_1) = 2E(X_1) = 2np = 2$ and $Var(Y_1) = 2\,Var(X_1) = 2np(1-p) = 1$.*
- $P(|Y_1 - 2| < 2) = P(|Y_1 - 2| < 2\sqrt{1}) \geq 1 - \frac{1}{4} = \frac{3}{4} = 0.75$.

(c) Find the exact value of $P(|Y_1 - 2| < 2)$. [2]

- *Y_1 has a Binomial$(n = 4, p = \frac{1}{2})$ distribution because X_1 and X_2 are iid Bin$(2, \frac{1}{2})$. Thus*

$$\begin{aligned} P(|Y_1 - 2| < 2) &= P(0 < Y_1 < 4) = P(1 \leq Y_1 \leq 3) \\ &= P(Y_1 = 1, 2, 3) = \frac{4}{2^4} + \frac{6}{2^4} + \frac{4}{2^4} = \frac{7}{8} = 0.875. \end{aligned}$$

(d) Define

$$Z_1 = \begin{cases} 1 & \text{if } Y_1 = 0, \\ 0 & \text{otherwise;} \end{cases} \quad \text{and} \quad Z_2 = \begin{cases} 1 & \text{if } Y_2 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

i. Find the marginal pmf (probability mass function) of Z_1 and Z_2 respectively. [2]

- *Both Z_1 and Z_2 are Bernoulli r.v.s with $P(Z_1 = 1) = P(Z_2 = 1) = P(X_1 = X_3 = 0) = P(X_2 = X_3 = 0) = 2^{-4}$.*

ii. Find the joint pmf (joint probability mass function) of (Z_1, Z_2). [3]

- $P(Z_1 = 1, Z_2 = 1) = P(X_1 = X_2 = X_3 = 0) = \dfrac{1}{64}$.
- $P(Z_1 = 1, Z_2 = 0) = P(X_1 = X_3 = 0, X_2 > 0) = \dfrac{1}{2^4}(1 - \dfrac{1}{2^2}) = \dfrac{3}{64}$.
- $P(Z_1 = 0, Z_2 = 1) = P(X_1 > 0, X_2 = X_3 = 0) = (1 - \dfrac{1}{2^2})\dfrac{1}{2^4} = \dfrac{3}{64}$.
- $P(Z_1 = 0, Z_2 = 0) = 1 - \dfrac{1+3+3}{2^6} = \dfrac{57}{64}$.

iii. Find the correlation coefficient between Z_1 and Z_2. [3]

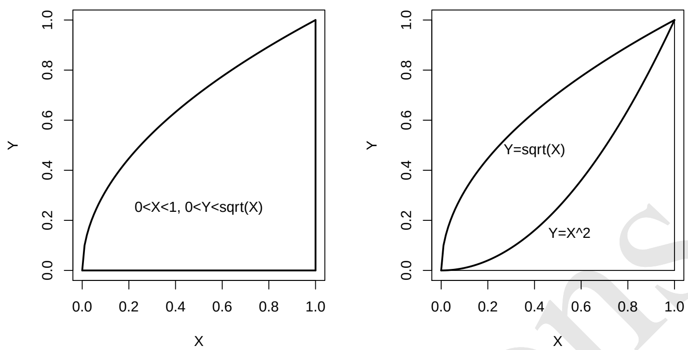
- *$E(Z_1) = P(Z_1 = 1) = \frac{1}{16}$ and $Var(Z_1) = P(Z_1 = 1)P(Z_1 = 0) = \frac{15}{256}$.*
- *Z_1 has the same distribution as Z_2. So $E(Z_2) = \frac{1}{16}$ and $Var(Z_2) = \frac{15}{256}$.*
- *$E(Z_1 Z_2) = P(Z_1 = Z_2 = 1) = \frac{1}{64}$. So $Cov(Z_1, Z_2) = \frac{1}{64} - \frac{1}{256} = \frac{3}{256}$.*
- *Therefore,*

$$\rho(Z_1, Z_2) = \frac{Cov(Z_1, Z_2)}{\sqrt{Var(Z_1)}\sqrt{Var(Z_2)}} = \frac{\frac{3}{256}}{\frac{15}{256}} = \frac{1}{5} = 0.2.$$

7. The marginal distribution of a random variable X is uniform $U(0, 1)$. The conditional distribution of another random variable Y, given $X = x$, is uniform $U(0, \sqrt{x})$.

(a) Find the joint pdf of X and Y. Sketch the region where $f(x, y) > 0$. [2]

- $f(x,y) = f_1(x)h(y|x) = 1 \times \dfrac{1}{\sqrt{x}} = \dfrac{1}{\sqrt{x}}, \quad 0 < x < 1, \quad 0 < y < \sqrt{x}.$



(b) Find $f_2(y)$, the marginal pdf of Y. [2]

- $f_2(y) = \displaystyle\int_{-\infty}^{\infty} f(x,y)dx = \int_{y^2}^{1} \frac{1}{\sqrt{x}}dx = 2\sqrt{x}\,\Big|_{y^2}^{1} = 2(1-y), \quad 0 < y < 1.$

(c) Find $g(x|y)$, the conditional pdf of X, given $Y = y$. [2]

- $g(x|y) = \dfrac{f(x,y)}{f_2(y)} = \dfrac{1/\sqrt{x}}{2(1-y)} = \dfrac{1}{2(1-y)\sqrt{x}}, \quad y^2 < x < 1.$

(d) Compute the probability $P(X < \frac{4}{9} \mid Y = \frac{1}{2})$. [2]

- $P(X < \dfrac{4}{9} \mid Y = \dfrac{1}{2}) = \displaystyle\int_{1/4}^{4/9} \frac{1}{\sqrt{x}}dx = 2\sqrt{x}\,\Big|_{1/4}^{4/9} = \frac{4}{3} - 1 = \frac{1}{3}.$

(e) Compute the probability $P(Y > X^2)$. [2]

- $P(Y > X^2) = \displaystyle\int_0^1 \left(\int_{x^2}^{\sqrt{x}} \frac{1}{\sqrt{x}}dy \right) dx = \int_0^1 (1 - x^{3/2})dx = \frac{3}{5}.$
- *Or* $P(Y > X^2) = \displaystyle\int_0^1 \left(\int_{y^2}^{\sqrt{y}} \frac{1}{\sqrt{x}}dx \right) dy = \int_0^1 (2y^{1/4} - 2y)dy = \frac{3}{5}.$

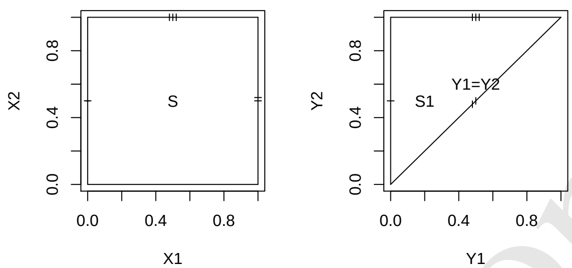
8. Consider two random variables X_1 and X_2 with the joint pdf

$$f(x_1, x_2) = \begin{cases} 6x_1^2 x_2, & 0 < x_1 < 1, \quad 0 < x_2 < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = X_1 X_2$ and $Y_2 = X_2$ be a joint transformation of (X_1, X_2).

(a) Find the support of (Y_1, Y_2) and sketch it. [3]

- *The support of* (Y_1, Y_2) *is* $\{0 < Y_1 < Y_2 < 1\}$.



(b) Find the inverse transformation. [2]

- $X_1 = Y_1/Y_2, X_2 = Y_2.$

(c) Compute the Jacobian of the inverse transformation. [2]

- $J = \begin{vmatrix} 1/y_2 & -y_1/y_2^2 \\ 0 & 1 \end{vmatrix} = \dfrac{1}{y_2}.$

(d) Compute the joint pdf of (Y_1, Y_2). [2]

- $g(y_1, y_2) = |J| \cdot f(\dfrac{y_1}{y_2}, y_2) = \dfrac{1}{y_2} \cdot 6 \left(\dfrac{y_1}{y_2}\right)^2 y_2 = \dfrac{6y_1^2}{y_2^2}, \quad 0 < y_1 < y_2 < 1.$

(e) Find the marginal pdf of Y_1 from the joint pdf of (Y_1, Y_2). [3]

- $g_1(y_1) = \displaystyle\int_{y_1}^{1} \frac{6y_1^2}{y_2^2} dy_2 = -6y_1^2 + 6y_1 = 6y_1(1 - y_1), \quad 0 < y_1 < 1.$

9. Let $X_1, X_2, \cdots, X_n$ be independent random variables each having the moment-generating function (mgf)

$$M(t) = \frac{e^{5t}}{1-2t}, \quad t < \frac{1}{2}.$$

(a) Find the mgf $M_{Y_n}(t)$ of the sum $Y_n = X_1 + X_2 + \cdots + X_n$. [2]

- $M_{Y_n}(t) = E(e^{tY_n}) = [M_{X_1}(t)]^n = \left[\frac{e^{5t}}{1-2t}\right]^n = \frac{e^{5nt}}{(1-2t)^n}, \quad t < \frac{1}{2}.$

(b) Find the mgf $M_{\bar{Y}_n}(t)$ of the sample mean $\bar{Y}_n = \frac{Y_n}{n}$. [2]

- $M_{\bar{Y}_n}(t) = E(e^{t\bar{Y}_n}) = M_{Y_n}\left(\frac{t}{n}\right) = \frac{e^{5n(t/n)}}{(1-2t/n)^n} = \frac{e^{5t}}{(1-2t/n)^n}, \quad t < \frac{n}{2}.$

(c) Find the limiting mgf $\lim_{n\to\infty} M_{\bar{Y}_n}(t)$. What distribution does the limiting mgf correspond to? What is the implication of this result? [2]

- *For any given* t, $\lim_{n\to\infty} M_{\bar{Y}_n}(t) = \lim_{n\to\infty} \frac{e^{5t}}{(1-2t/n)^n} = \frac{e^{5t}}{e^{-2t}} = e^{7t}$
- *The limit is the mgf of a degenerate distribution having probability 1 at 7.*
- *This implies that* $\bar{Y}_n \xrightarrow{p} 7 = E(X_1)$ *as* $n \to \infty$.
- *Note it is also a correct answer if it is based on applying the WLLN. Then the WLLN must be correctly stated, and* $E(X_1) = 7$ *be proved.*

(d) Let $Z_n = \sqrt{n}(\bar{Y}_n - 7)$. Find $M_{Z_n}(t)$, the mgf of Z_n. Then find $\lim_{n\to\infty} M_{Z_n}(t)$. Finally explain what is the limiting distribution of Z_n when $n \to \infty$. [5]

- $$\begin{aligned} M_{Z_n}(t) &= E(e^{t\sqrt{n}(\bar{Y}_n - 7)}) = e^{-7t\sqrt{n}} M_{\bar{Y}_n}(\sqrt{n}t) \\ &= e^{-7t\sqrt{n}} \cdot \frac{e^{5\sqrt{n}t}}{(1-2\sqrt{n}t/n)^n} = \left[\frac{e^{-2t/\sqrt{n}}}{1-2t/\sqrt{n}}\right]^n, \quad t < \frac{\sqrt{n}}{2}. \end{aligned}$$
- *By Taylor's series expansion,* $e^u \approx 1 + u + \frac{1}{2}u^2$ *when* $|u|$ *is sufficiently small. Using this result, for any given* t,

$$\begin{aligned} \lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty} \left[\frac{e^{-\frac{2t}{\sqrt{n}}}}{1-2\frac{t}{\sqrt{n}}}\right]^n = \lim_{n\to\infty} \left[\frac{1 - \frac{2t}{\sqrt{n}} + \frac{2t^2}{n}}{1 - \frac{2t}{\sqrt{n}}}\right]^n \\ &= \lim_{n\to\infty} \left[1 + \frac{\frac{2t^2}{n}}{1 - \frac{2t}{\sqrt{n}}}\right]^n = \lim_{n\to\infty} \left(1 + \frac{2t^2}{n}\right)^n = e^{4t^2/2} \end{aligned}$$

- *The limit mgf is that of* $N(\mu = 0, \sigma^2 = 4)$.
- *This implies that* $Z_n \xrightarrow{d} N(0,4)$ *as* $n \to \infty$.
- *Note it is also a correct answer if it is based on applying the CLT. Then the CLT must be correctly stated, and* $Var(X_1) = 4$ *be proved.*

10. A random variable X has the following mgf

$$M(t) = \frac{2}{5(1-t)} + \frac{3}{5(1-2t)}, \quad t < \frac{1}{2}.$$

(a) Find the value of $E(X)$. [2]

- $M'(t) = \dfrac{2}{5(1-t)^2} + \dfrac{6}{5(1-2t)^2}$.
- *Thus* $E(X) = M'(0) = \frac{8}{5} = 1.6$.

(b) Find the value of $\mathrm{Var}(2^{-X})$. [4]

-
$$E(2^{-X}) = E(e^{-X\ln 2}) = M(-\ln 2) = \frac{2}{5(1+\ln 2)} + \frac{3}{5(1+2\ln 2)}$$
$$= 0.2362464 + 0.2514359 = 0.4876823.$$

-
$$E((2^{-X})^2) = E(e^{-X\ln 4}) = M(-\ln 4) = \frac{2}{5(1+\ln 4)} + \frac{3}{5(1+2\ln 4)}$$
$$= 0.1676239 + 0.159042 = 0.3266659.$$

- *Hence* $Var(2^{-X}) = 0.3266659 - 0.4876823^2 = 0.08883187$.

(c) Find the probability $P(X > 4)$. [4]

- $\dfrac{1}{1-t}$ *is the mgf of* $Y \stackrel{d}{=} Exp(1)$ *and* $\dfrac{1}{1-2t}$ *is the mgf of* $W \stackrel{d}{=} Exp(2)$.
- *It follows that*

$$P(X > 4) = \frac{2}{5}P(Y > 4) + \frac{3}{5}P(W > 4) = \frac{2}{5}e^{-4} + \frac{3}{5}e^{-4/2} = 0.08852743.$$

`Total marks = 100`

End of the exam questions.
Formulas are on the next page.