The University of Melbourne

Mid Semester 1 Assessment — 2019 (Solutions)

School of Mathematics and Statistics

MAST90105 Methods of Mathematical Statistics

Exam duration: **3 hours**
Reading time: **15 minutes**
This paper has **5 pages** including this page

Authorised materials:
The calculator authorised at the University of Melbourne is the CASIO FX82 and this is permitted.
Two A4 double-sided handwritten sheets of notes.

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:
There are 8 questions. All questions may be attempted.
The number of marks for each question is indicated after the question.
The total number of marks available is 100.
Your raw mark of this exam will be multiplied with 0.35 before being added to your final subject mark.

1. A mutation in a certain gene can occur with probability 0.02. The probability of a rear disease in a person with mutation in this gene is 0.10, and this probability is 0.002 otherwise.

   (a) Find the probability that a random person has this disease.

   - Consider the following events: $A = \{person\ has\ a\ mutation\ in\ the\ gene\}$, $B = \{person\ has\ no\ mutation\ in\ the\ gene\}$ and $C = \{person\ has\ the\ disease\}$.
   - $\Pr(A) = 0.02$, $\Pr(B) = 1 - \Pr(A) = 0.98$, $\Pr(C|A) = 0.10$ and $\Pr(C|B) = 0.002$. We use the law of total probability to find:

   $$\Pr(C) = \Pr(C|A)\Pr(A) + \Pr(C|B)\Pr(B) = 0.10 \cdot 0.02 + 0.002 \cdot 0.98 = \mathbf{0.00396}.$$

   (b) Find the probability that a person with the disease has a mutation in this gene.

   - Using Bayes' theorem or conditional probability formula, we get:

   $$\Pr(A|C) = \Pr(C|A)\Pr(A)/\Pr(C) = 0.10 \cdot 0.02/0.00396 = \mathbf{0.505}.$$

   (c) Consider a group of three patients with the disease. Find the probability that two of them have mutations in this gene.

   - Let X be the number of patients who have mutations in the gene. X follows a binomial distribution with parameter $p^* = \Pr(A|C) = 0.505$ and $\Pr(X = 2) = \binom{3}{2}(p^*)^2(1 - p^*) = 3 \cdot 0.505^2 \cdot (1 - 0.505) = \mathbf{0.379}$.

   [4 + 2 + 4 = 10 marks]

2. In Sydney area, the number of earthquakes during next t years, X_t, follows a Poisson process with rate 1 per year.

   (a) Find the probability that there will be no earthquakes in Sydney next year.

   - X_1 is the number of earthquakes in the next year. X_1 follows a Poisson distribution with $\lambda = 1$.
   - $\Pr(X_1 = 0) = e^{-\lambda} = e^{-1} = \mathbf{0.368}$.

   (b) Let T_0 be the time (in years) until first year without an earthquake. Find $\Pr(2 \le T_0 \le 5)$.

   - T_0 has a geometric distribution with parameter $p = \Pr(X_1 = 0) = 0.368$ and, with $q = 1 - p$, $\Pr(2 \le T_0 \le 5) = \sum_{k=2}^{5} q^{k-1}p = q(1 - q^4) = \mathbf{0.531}$.

   (c) Find $\Pr(X_3 \ge 3|X_3 \ge 2)$.

   - X_3 follows a Poisson distribution with parameter $3\lambda = 3$.
   - We use the formula of conditional probability to find:

   $$\Pr(X_3 \ge 3|X_3 \ge 2) = \frac{\Pr(X_3 \ge 3, X_3 \ge 2)}{\Pr(X_3 \ge 2)} = \frac{\Pr(X_3 \ge 3)}{\Pr(X_3 \ge 2)} = \frac{1 - \sum_{k=0}^{2}\Pr(X_3 = k)}{1 - \sum_{k=0}^{1}\Pr(X_3 = k)}$$

   $$= \frac{1 - e^{-3} - 3e^{-3} - 3^2e^{-3}/2}{1 - e^{-3} - 3e^{-3}} = \frac{1 - 8.5e^{-3}}{1 - 4e^{-3}} = \mathbf{0.72}.$$

   (d) Find $\Pr(2 \le X_2 \le 4|X_5 = 7)$.

   - Again, let $X_{3:5}$ be the number of earthquakes in years 3–5. X_2 and $X_{3:5}$ are independent random variables and they follow a Poisson distribution with parameters $2\lambda = 2$ and $3\lambda = 3$, respectively. X_5 follows a Poisson distribution with parameter $5\lambda = 5$.
   - For $k = 2, 3, 4$ we use the formula of conditional probability to find:

   $$\Pr(X_2 = k|X_5 = 7) = \frac{\Pr(X_2 = k, X_5 = 7)}{\Pr(X_5 = 7)} = \frac{\Pr(X_2 = k, X_{3:5} = 7 - k)}{\Pr(X_5 = 7)}$$

   $$= \frac{\Pr(X_2 = k)\Pr(X_{3:5} = 7 - k)}{\Pr(X_5 = 7)} = \frac{e^{-2}2^k e^{-3}3^{7-k}7!}{k!(7-k)!e^{-5}5^7} = \binom{7}{k}\left(\frac{2}{5}\right)^k\left(\frac{3}{5}\right)^{7-k}.$$

- $\Pr(2 \leq X_2 \leq 4 | X_5 = 7) = \sum_{k=2}^{4} \binom{7}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{7-k} = \mathbf{0.745}$.

[2 + 3 + 5 + 5 = 15 marks]

3. There are three coins: one coin is fair, and the other two are biased. The probability that the first biased coin shows tail and head is 1/3 and 2/3, respectively, and the probability that the second biased coin shows tail and head is 2/3 and 1/3, respectively. We randomly select one coin and flip it two times. Let X be the number of tails in the two flips.

(a) Find the range and probability mass function (pmf) of X.

- X can take values $0, 1, 2$. Let $C_1 = \{\text{fair coin is selected}\}$, $C_2 = \{\text{the first biased coin is selected}\}$ and $C_3 = \{\text{the second biased coin is selected}\}$. We have $\Pr(C_1) = \Pr(C_2) = \Pr(C_3) = 1/3$.
- Number of tails for a selected coin follows a binomial distribution with $p = 1/2$ and $p = 1/3, p = 2/3$ for the fair coin and two biased coins, respectively.
- We use the law of total probability to find

$$\begin{aligned}
\Pr(X = 0) &= \Pr(X = 0|C_1)\Pr(C_1) + \Pr(X = 0|C_2)\Pr(C_2) + \Pr(X = 0|C_3)\Pr(C_3) \\
&= \left(\frac{1}{2}\right)^2 \frac{1}{3} + \left(\frac{2}{3}\right)^2 \frac{1}{3} + \left(\frac{1}{3}\right)^2 \frac{1}{3} = \frac{29}{108}, \\
\Pr(X = 1) &= \Pr(X = 1|C_1)\Pr(C_1) + \Pr(X = 1|C_2)\Pr(C_2) + \Pr(X = 1|C_3)\Pr(C_3) \\
&= 2\left(\frac{1}{2}\right)^2 \frac{1}{3} + \left(2 \cdot \frac{2}{3} \cdot \frac{1}{3}\right)\frac{1}{3} + \left(2 \cdot \frac{1}{3} \cdot \frac{2}{3}\right)\frac{1}{3} = \frac{50}{108}, \\
\Pr(X = 2) &= \Pr(X = 2|C_1)\Pr(C_1) + \Pr(X = 2|C_2)\Pr(C_2) + \Pr(X = 2|C_3)\Pr(C_3) \\
&= \left(\frac{1}{2}\right)^2 \frac{1}{3} + \left(\frac{1}{3}\right)^2 \frac{1}{3} + \left(\frac{2}{3}\right)^2 \frac{1}{3} = \frac{29}{108},
\end{aligned}$$

(b) Find $\mathrm{E}(X)$, $\mathrm{Var}(X)$ and the moment generating function of X.

- $\mathrm{E}(X) = 0 \cdot \Pr(X = 0) + 1 \cdot \Pr(X = 1) + 2 \cdot \Pr(X = 2) = 0 \cdot \frac{29}{108} + 1 \cdot \frac{50}{108} + 2 \cdot \frac{29}{108} = 1$,
- $\mathrm{E}(X^2) = 0^2 \cdot \Pr(X = 0) + 1^2 \cdot \Pr(X = 1) + 2^2 \cdot \Pr(X = 2) = 0 \cdot \frac{29}{108} + 1 \cdot \frac{50}{108} + 4 \cdot \frac{29}{108} = \frac{83}{54}$,
- $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = \frac{83}{54} - 1 = \frac{29}{54}$.
- The MGF of X is $M(t) = \frac{29}{108} + \frac{50}{108}e^t + \frac{29}{108}e^{2t}$.

[5+5 = 10 marks]

4. The moment generating function of a random variable X is

$$M(t) = C\left(\frac{e^t + e^{-t}}{3}\right)^2 + \left(\frac{e^{t/2} + e^{-t/2}}{3}\right)^2.$$

(a) Find the constant C and the probability mass function of X.

- $M(0) = 4C/9 + 4/9 = 1 \Rightarrow \mathbf{C = 1.25}$.
- Let $f(x) = \Pr(X = x)$. We can write $M(t) = \frac{C}{9}e^{2t} + \frac{1}{9}e^t + \frac{2C+2}{9} + \frac{1}{9}e^{-t} + \frac{C}{9}e^{-2t}$.
- It implies that $f(2) = f(-2) = \frac{C}{9} = \frac{5}{36}$, $f(1) = f(-1) = \frac{1}{9}$, $f(0) = \frac{2C+2}{9} = \frac{1}{2}$.

(b) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

- $\mathrm{E}(X) = \mathbf{0}$ because $f(x)$ is symmetric about zero.
- $\mathrm{Var}(X) = \mathrm{E}(X^2) = 4f(-2) + f(-1) + f(1) + 4f(2) = 4 \cdot \frac{5}{36} + 1 \cdot \frac{1}{9} + 1 \cdot \frac{1}{9} + 4 \cdot \frac{5}{36} = \mathbf{\frac{4}{3}}$.

[5 + 5 = 10 marks]

5. A device has two components that work independently of each other. This device fails if at least one of these components fail. The lifetimes (times to failure, measured in years) of these two components, T_1, T_2, follow an exponential distribution with $\Pr(T_k \leq t) = 1 - e^{-tk}$, $k = 1, 2$.

   (a) Find the probability that exactly one of the two components fails in one year.

   - Denote the event $A_k = \{\text{component } k \text{ fails in one year}\}$, $k = 1, 2$. It follows that $\Pr(A_k) = \Pr(T_k \leq 1) = 1 - e^{-k}$.
   - Denote the event $B = \{\text{exactly one component fails in one year}\}$. Using independence of A_1 and A_2, we find:

   $$B = A_1 A_2^c \sqcup A_1^c A_2, \quad \Pr(B) = (1 - e^{-1})e^{-2} + e^{-1}(1 - e^{-2}) = \mathbf{0.4036}.$$

   (b) Let T_0 be the lifetime of the device. Find the cumulative distribution function of T_0, $\Pr(T_0 \leq t)$.

   - Note that the two events $\{T_0 > t\}$ and $\{T_1 > t, T_2 > t\}$ are the same.
   - We use independence of T_1 and T_2 to find

   $$\Pr(T_0 \leq t) = 1 - \Pr(T_0 > t) = 1 - \Pr(T_1 > t, T_2 > t)$$

   $$= 1 - \Pr(T_1 > t) \Pr(T_2 > t) = 1 - e^{-3t}.$$

   (c) What is the expected lifetime of the device?

   - We found in (b) that T_0 follows an exponential distribution with rate $\lambda = 3$ and hence $\mathrm{E}(T_0) = \mathbf{1/3}$ **years**.

   [5 + 5 + 2 = 12 marks]

6. Let X be a continuous random variable with the probability density function

   $$f(x) = \begin{cases} C, & \text{if } -1 \leq x \leq 0, \\ 2C, & \text{if } \quad 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Find the constant C.

   - $\int_{-1}^{1} f(x) = \int_{-1}^{0} C dx + \int_{0}^{1} 2C dx = C + 2C = 3C = 1 \Rightarrow \mathbf{C = 1/3}$.

   (b) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

   - $\mathrm{E}(X) = \int_{-1}^{1} x f(x) dx = \frac{1}{3}\int_{-1}^{0} x dx + \frac{2}{3}\int_{-1}^{0} x dx = \frac{1}{3}\frac{x^2}{2}\Big|_{-1}^{0} + \frac{2}{3}\frac{x^2}{2}\Big|_{0}^{1} = -\frac{1}{3}\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{1}{2} = \mathbf{\frac{1}{6}}$.
   - $\mathrm{E}(X^2) = \int_{-1}^{1} x^2 f(x) dx = \frac{1}{3}\int_{-1}^{0} x^2 dx + \frac{2}{3}\int_{-1}^{0} x^2 dx = \frac{1}{3}\frac{x^3}{3}\Big|_{-1}^{0} + \frac{2}{3}\frac{x^3}{3}\Big|_{0}^{1} = \frac{1}{3}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{1}{3} = \mathbf{\frac{1}{3}}$,
   - $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = \frac{1}{3} - \frac{1}{36} = \mathbf{\frac{11}{36}}$.

   (c) Find the cumulative distribution function of X, $F(x)$.

   - $F(x) = 0$ for $x < -1$ and $F(x) = 1$ for $x > 1$.
   - If $-1 \leq x \leq 0$: $F(x) = \int_{-1}^{x} f(t) dt = \frac{1}{3}t\Big|_{-1}^{x} = \frac{x+1}{3}$.
   - If $0 \leq x \leq 1$: $F(x) = \int_{-1}^{x} f(t) dt = \int_{0}^{1} f(t) dt + \int_{1}^{x} f(t) dt = \frac{1}{3}t\Big|_{-1}^{0} + \frac{2}{3}t\Big|_{0}^{x} = \frac{1+2x}{3}$.

   (d) Find the median of X.

   - Let x_m be the median of X. $F(x) < 1/3$ for $x < 0$ and therefore $x_m > 0$:
   - $F(x_m) = (1 + 2x_m)/3 = 0.5 \Rightarrow \mathbf{x_m = 0.25}$.

   (e) Let $Y = X^2$. Find the cumulative distribution function of Y, $G(y)$.

- Y can take values in $(0, 1)$ interval and therefore $G(y) = 0$ for $y < 0$ and $G(y) = 1$ for $y > 1$.
- If $0 \leq y \leq 1$, then

$$G(y) = \Pr(Y \leq y) = \Pr(X^2 \leq y) = \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$$
$$= \frac{1 + 2\sqrt{y}}{3} - \frac{1 - \sqrt{y}}{3} = \sqrt{y}$$

(f) Find the probability density function of Y, $g(y)$.

- $g(y) = 0$ for $y < 0$ or $y > 1$.
- If $0 \leq y \leq 1$: $g(y) = G'(y) = 0.5/\sqrt{y}$.

[2 + 5 + 4 + 4 + 5 + 2 = 22 marks]

7. Let X_1, X_2 be two independent Bernoulli random variables with probability of success $p = 1/2$. Define two new random variables $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$.

(a) Find the joint probability mass function of Y_1, Y_2.

- $f(1, 1) = \Pr(Y_1 = Y_2 = 1) = \Pr(X_1 = X_2 = 1) = \Pr(X_1 = 1)\Pr(X_2 = 1) = 1/4$,
- $f(1, 0) = \Pr(Y_1 = 1, Y_2 = 0) = \Pr(X_1 = X_2 = 1, X_1 = X_2 = 0) = 0$,
- $f(0, 1) = \Pr(Y_1 = 0, Y_2 = 1) = \Pr(X_1 = 0, X_2 = 1) + \Pr(X_1 = 1, X_2 = 0) = 1/2$,
- $f(0, 0) = \Pr(Y_1 = Y_2 = 0) = \Pr(X_1 = X_2 = 0) = 1/4$.

(b) Find $\mathrm{E}(Y_1), \mathrm{E}(Y_2), \mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$.

- $\mathrm{E}(Y_1) = \mathrm{E}(Y_1^2) = \Pr(Y_1 = 1) = f(1, 0) + f(1, 1) = \mathbf{1/4}$, $\mathrm{E}(Y_2) = \mathrm{E}(Y_2^2) = \Pr(Y_2 = 1) = f(0, 1) + f(1, 1) = \mathbf{3/4}$,
- $\mathrm{Var}(Y_1) = \mathrm{E}(Y_1^2) - \{\mathrm{E}(Y_1)\}^2 = \mathbf{3/16}$, $\mathrm{Var}(Y_2) = \mathrm{E}(Y_2^2) - \{\mathrm{E}(Y_2)\}^2 = \mathbf{3/16}$.

(c) Find $\mathrm{Cov}(Y_1, Y_2)$. Are the variables Y_1, Y_2 independent? Why or why not?

- $\mathrm{Cov}(Y_1, Y_2) = \mathrm{E}(Y_1 Y_2) - \mathrm{E}(Y_1)\mathrm{E}(Y_2) = f(1, 1) - (1/4) \cdot (3/4) = \mathbf{1/16}$.
- $\mathrm{Cov}(Y_1, Y_2) \neq 0$ and therefore Y_1 and Y_2 are **not independent**.

[4 + 4 + 3 = 11 marks]

8. Let Z_0, Z_1 and Z_2 be three independent standard normal random variables.

(a) Define two random variables $W_1 = Z_1 + \rho Z_0$ and $W_2 = Z_2 + \rho Z_0$. Find ρ such that the correlation $\mathrm{Corr}(W_1, W_2) = 0.2$.

- $\mathrm{Var}(W_1) = \mathrm{Var}(W_2) = 1 + \rho^2$
- $\mathrm{Cov}(W_1, W_2) = \mathrm{Cov}(Z_1, Z_2) + \rho\mathrm{Cov}(Z_1, Z_0) + \rho\mathrm{Cov}(Z_2, Z_0) + \rho^2\mathrm{Cov}(Z_0, Z_0) = \rho^2$
- $\mathrm{Corr}(W_1, W_2) = \rho^2/(1 + \rho^2) = 0.2 \Rightarrow \rho = \pm\mathbf{0.5}$.

(b) Define two random variables $Y_1 = Z_1 + 2Z_2$, $Y_2 = 2Z_1 - Z_2$. Find $\mathrm{E}(Y_1|Y_2 = 1)$.

- $\mathrm{Cov}(Y_1, Y_2) = 2\mathrm{Var}(Z_1) + 3\mathrm{Cov}(Z_1, Z_2) - 2\mathrm{Var}(Z_2) = 0$
- $\mathrm{E}(Y_1|Y_2 = 1) = \mathrm{E}(Y_1) = \mathbf{0}$ because Y_1 and Y_2 are independent: they follow a bivariate normal distribution with zero covariance.

[5 + 5 = 10 marks]

Total marks = 100

End of the Questions