Imperial College London

MATH50010

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2021

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Probability for Statistics

Date: Tuesday, 18 May 2021

Time: 09:00 to 11:00

Time Allowed: 2 hours

Upload Time Allowed: 30 minutes

This paper has 4 Questions.

Candidates should start their solutions to each question on a new sheet of paper.

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS AS SEPARATE PDFs TO THE RELEVANT DROPBOXES ON BLACKBOARD INCLUDING A COMPLETED COVERSHEET WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. (a) Suppose $(\Omega, \mathcal{F}, \Pr)$ is a probability space. Determine whether or not each of the following statements is true or false in general. Give proofs or counterexamples as appropriate.

(i) $\mathcal{F}$ is closed under symmetric differences: if $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \triangle B \in \mathcal{F}$, where $A \triangle B = (A \cup B) \backslash (A \cap B)$.

(2 marks)

(ii) If $A_1, A_2, \ldots A_n \in \mathcal{F}$ then

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \Pr(A_i) - (n-1).$$

(4 marks)

(iii) If $A \in \mathcal{F}$ is such that $\Pr(A) = 1$, then $A = \Omega$.

(2 marks)

(iv) $\mathcal{F}$ is closed under arbitrary unions: if $A_i \in \mathcal{F}$ for all $i \in I$, then $\bigcup_{i \in I} A_i \in \mathcal{F}$.

(3 marks)

(v) If $A_1, A_2, \ldots$ is a sequence of events such that $\Pr(A_n) = 0$ for all $n \geq 1$, then $\Pr\left(\bigcup_{n \geq 1} A_n\right) = 0$.

(3 marks)

(b) Suppose a coin is flipped twice, so that the sample space for the experiment is $\Omega = \{HH, HT, TH, TT\}$. Let $X, Y : \Omega \to \mathbf{R}$ be defined as follows

$$X(\omega) = \begin{cases} 1 & \omega \in \{HH, HT\} \\ 0 & \omega \in \{TH, TT\} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \omega \in \{HH, TT\} \\ 0 & \omega \in \{HT, TH\} \end{cases}$$

(i) State whether or not each of the functions X and Y is a random variable with respect to the algebra

$$\mathcal{F} = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$$

of subsets of Ω. Justify your answers.

(4 marks)

(ii) Determine $\mathcal{F}_Z$, the smallest algebra of subsets of Ω with respect to which $Z = XY$ is a random variable.

(2 marks)

(Total: 20 marks)

2. The random variable X has probability density function given by

$$f_X(x|\theta) = \theta(1-x)^{\theta-1} \qquad 0 < x < 1,$$

and zero otherwise, where $\theta > 0$.

(a) Find the cumulative distribution function $F_X(\cdot|\theta)$.

(2 marks)

(b) With reference to F_X, explain why X is a continuous random variable.

(1 mark)

(c) Give examples of functions $g : \mathbf{R} \to \mathbf{R}$ such that $Y = g(X)$ is i) a discrete random variable ii) a random variable that is neither discrete nor continuous. Justify your answers.

(4 marks)

(d) Given a random sample $U_1, U_2, \ldots U_n \sim \textsc{Uniform}(0,1)$, explain how to generate a random sample of size n from the distribution of X. State clearly any results that you use.

(4 marks)

(e) Show that the random variable $X_n = \min\{U_1, \ldots, U_n\}$ has the same distribution as X for a particular value of θ, which you should specify.

(4 marks)

(f) Show that $X_n \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

(2 marks).

(g) Find a deterministic (i.e. non-random) sequence $(a_n)_{n\geq 1}$ such that $a_n X_n \xrightarrow{\mathcal{D}} Z$, where Z is a non-degenerate random variable, whose distribution you should specify.

(3 marks)

(Total: 20 marks)

3. (a) Consider the standard bivariate Normal random variable $\boldsymbol{Z} = (X, Y)$ with probability density function

$$f_{XY}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right), \qquad (x, y) \in \mathbf{R}^2.$$

(i) Determine the joint distribution of $(S, T) = \left(\frac{1}{2}(X_1 + X_2), \frac{1}{2}(X_1 - X_2)\right)$. State whether or not S and T are independent, with brief justification. (4 marks)

(ii) Find the joint distribution of (R, Θ) where $X = R\cos\Theta$ and $Y = R\sin\Theta$. State whether or not R and Θ are independent . (4 marks)

(iii) Find the distribution of $V = \frac{Y}{X}$. (3 marks)

(b) Suppose the random variable Z has $E(Z) = \mu$ and $\mathrm{Var}(Z) = \sigma^2$.

(i) Show that for all $\alpha > 0$,

$$\Pr(Z - \mu \geq \alpha) \leq \Pr\left((Z - \mu + y)^2 \geq (\alpha + y)^2\right)$$

for all $y > 0$.

(2 marks)

(ii) Deduce that for any $\alpha > 0$,

$$\Pr(Z - \mu \geq \alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}.$$

(3 marks)

(iii) Suppose now that $\mu = 0$. For given values of σ^2 and α, show that the upper bound for $\Pr(Z \geq \alpha)$ established in (ii) is the sharpest possible, by constructing an example in which equality holds.

(4 marks)

(Total: 20 marks)

4. The sequence $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

(a) Draw the transition diagram for (X_n).

(3 marks)

(b) Determine the communicating classes of the chain, stating whether each class is recurrent or transient.

(3 marks)

(c) Determine the period of each communicating class. (3 marks)

(d) Suppose that the initial distribution of the chain is uniform on the set $\{1, 2, 5, 6\}$. Find $\Pr(X_1 = 2)$.

(3 marks)

(e) If the initial distribution of the chain is $\pi_0 = (0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0)$, determine the limiting probability distribution of the chain.

(3 marks)

(f) Consider now the Markov chain Y_n on the state space $\{1, 2\}$, with transition matrix

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $0 < \alpha, \beta < 1$.

Write down the log likelihood of the parameters α and β given a realization $(y_0, y_1, y_2, \ldots, y_n)$ of the chain, where each $y_k \in \{1, 2\}$.

Derive the maximum likelihood estimators of α and β. Give your answer in terms of n_{ij}, the number of jumps from state i to state j for $i, j \in \{1, 2\}$. Assume each $n_{ij} > 0$.

(5 marks)

(Total: 20 marks)

1. (a) (i) By Axiom 3 (closure under unions), $A \cup B \in \mathcal{F}$. By Axiom 2 (closure under complements) and then Axiom 3, sim. seen ⇓ 2, A

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{F}.$$

A further appeal to Axiom 2 gives $(A \cup B) \backslash (A \cap B) \in \mathcal{F}$. meth seen ⇓

(ii) Note that for any $A, B \in \mathcal{F}$ we have 4, A

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Since $\Pr(\cdot)$ is bounded above by 1, this gives

$$\Pr(A \cap B) = \Pr(A) + \Pr(B) - \Pr(A \cup B) \geq \Pr(A) + \Pr(B) - 1.$$

Applying this result with $A = \bigcap_{i=1}^{n} A_i$ and $B = A_{n+1}$ gives

$$\Pr\left(\bigcap_{i=1}^{n+1} A_i\right) = \Pr\left(\bigcap_{i=1}^{n} A_i \cap A_{n+1}\right) \geq \Pr\left(\bigcap_{i=1}^{n} A_i\right) + \Pr(A_{n+1}) - 1.$$

Supposing for induction that the desired result holds for some $n \geq 1$, we see that

$$\Pr\left(\bigcap_{i=1}^{n+1} A_i\right) \geq \sum_{i=1}^{n} \Pr(A_i) - (n-1) + \Pr(A_{n+1}) - 1 = \sum_{i=1}^{n+1} \Pr(A_i) - n.$$

Since the result clearly holds for $n = 1$, the result is established for all $n \geq 1$. sim. seen ⇓

(iii) Take e.g. $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}$, the Borel sigma algebra on Ω and $\Pr$ Lebesgue measure on Ω. Then $\Pr((0, 1)) = 1$ but $(0, 1) \neq \Omega$. 2, A

(iv) Let $\Omega = [0, 1]$ and let $\mathcal{F}$ be the sigma algebra consisting of all countable or co-countable subsets of Ω. Then define $A_i = \{i\}$ for $i \in I = \left[0, \frac{1}{2}\right]$. Then each $A_i \in \mathcal{F}$, but the uncountable union $\bigcup_{i \in I} A_i = [0, \frac{1}{2}] \notin \mathcal{F}$. unseen ⇓ 3, D

(v) Define the increasing sequence of events $B_n = \bigcup_{i=1}^{n} A_i$. Then (by a union bound) for each $n \geq 1$, sim. seen ⇓ 3, C

$$\Pr(B_n) \leq \sum_{i=1}^{n} \Pr(A_i) = 0$$

and so by the continuity property of $\Pr$ on increasing sequences,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \Pr\left(\lim_{n \to \infty} B_n\right) = \lim_{n \to \infty} \Pr(B_n) = 0.$$

(b) (i) A function $Z : \Omega \to \mathbf{R}$ is a random variable if $Z^{-1}(B) \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}$. meth seen ⇓ 2, A

2, B

Let $B \in \mathcal{B}$ be a Borel set. Then

$$X^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ \{TH, TT\} & 0 \in B, 1 \notin B \\ \{HH, TH\} & 0 \notin B, 1 \in B \\ \Omega & 0 \in B, 1 \in B \end{cases}$$

Since the pre-image of each Borel set is a set in $\mathcal{F}$, X is a random variable.
For Y, note that $Y^{-1}(0) = \{HT, TH\} \notin \mathcal{F}$, so Y is not a random variable.

unseen ⇓

(ii)

2, C

$$Z(\omega) = \begin{cases} 1 & \omega = HH \\ 0 & \omega \in \{HT, TH, TT\} \end{cases}$$

The desired algebra is $\mathcal{F}_Z = \{Z^{-1}(B) : B \in \mathcal{B}\}$. This then gives

$$\mathcal{F}_Z = \{\emptyset, \{HH\}, \{HT, TH, TT\}, \Omega\}.$$

2. (a)

meth seen ⇓

2, A

$$F_X(x|\theta) = \int_{-\infty}^{x} f_X(t|\theta)\,dt = \int_0^x \theta(1-t)^{\theta-1}\,dt$$
$$= \left[(1-t)^\theta\right]_0^x = 1-(1-x)^\theta \qquad x \in (0,1).$$

Hence

$$F_X(x|\theta) = \begin{cases} 0 & x < 0 \\ 1-(1-x)^\theta & 0 \le x < 1 \\ 1 & x \ge 1. \end{cases}$$

sim. seen ⇓

(b) X is a continuous random variable because $F_X(x|\theta)$ is a continuous function of x.

1, A

unseen ⇓

(c) (i) e.g.

2, B

$$g(x) = \begin{cases} 0 & x \le \frac{1}{2} \\ 1 & x > \frac{1}{2} \end{cases}$$

Then

$$\Pr(Y=0) = \Pr\left(X \le \frac{1}{2}\right) = 1 - \left(\frac{1}{2}\right)^\theta$$

and

$$\Pr(Y=1) = \left(\frac{1}{2}\right)^\theta,$$

so that Y has a discrete distribution.

unseen ⇓

(ii) e.g.

2, B

$$g(x) = \begin{cases} 0 & x \le \frac{1}{2} \\ x & x > \frac{1}{2} \end{cases}$$

Now

$$\Pr(Y \le 0) = \Pr(Y=0) = \Pr\left(X \le \frac{1}{2}\right) = 1 - \left(\frac{1}{2}\right)^\theta > 0,$$

since $\theta > 0$. But for any $y < 0$, $\Pr(Y \le y) = 0$, so $F_Y(y)$ is not continuous at zero, and so Y is not a continuous random variable.

sim. seen ⇓

(d) The probability integral transform states that if F is a strictly increasing CDF and $U_1, U_2, \ldots, U_n \sim \text{UNIFORM}[0,1]$ is a random sample then $F^{-1}(U_1), \ldots, F^{-1}(U_n)$ is a random sample from the distribution with CDF F.

4, A

In this case, $F^{-1}(u) = 1-(1-u)^{\frac{1}{\theta}}$ for $u \in (0,1)$

meth seen ⇓

(e) Note that $\{X_n > x\} = \bigcap_{i=1}^n \{U_i > x\}$. So then, since the U_i are independent, we have

4, C

$$\Pr(X_n > x) = \Pr\left(\bigcap_{i=1}^n \{U_i > x\}\right) = \prod_{i=1}^n \Pr(U_i > x).$$

Now, for $x \in (0,1)$, this gives

$$\Pr(X_n > x) = (1-x)^n,$$

so that

$$\Pr(X_n \le x) = \begin{cases} 0 & x < 0 \\ 1-(1-x)^n & x \in (0,1) \\ 1 & x \ge 1. \end{cases}$$

and X_n has the same distribution as X, with $\theta = n$. meth seen ⇓

(f) Let $\epsilon > 0$. Then (at least for $\epsilon < 1$), 2, B

$$\Pr(|X_n| > \epsilon) = \Pr(X_n > \epsilon) = (1-\epsilon)^n \to 0$$

as $n \to \infty$, so that $X_n \to 0$ as $n \to \infty$. sim. seen ⇓

(g) Inspecting the behaviour in the previous part suggests $a_n = n$ is a suitable rescaling. Then for $x \in (0, n)$, 3, D

$$\Pr(nX_n > x) = \Pr\left(X_n > \frac{x}{n}\right) = \left(1 - \frac{x}{n}\right)^n,$$

so that for any $x > 0$,

$$\lim_{n\to\infty} \Pr(nX_n \le x) = \lim_{n\to\infty} 1 - \left(1 - \frac{x}{n}\right)^n = 1 - \exp(-x).$$

Hence $nX_n \xrightarrow{\mathcal{D}} Z \sim \text{Exp}(1)$.

3. (a) (i) Let $\boldsymbol{W} = (R, S)^t$. Then $\boldsymbol{W} = A\boldsymbol{Z}$ where

seen ⇓

4, A

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

By a result from lectures, $\boldsymbol{W}$ results from an invertible linear transformation of a multivariate normal vector, and so is multivariate normal.
Its distribution is completely determined by its mean vector and variance-covariance matrix, which we find below.
$\mathrm{E}(\boldsymbol{W}) = \mathrm{E}(A\boldsymbol{Z}) = A\mathrm{E}(\boldsymbol{Z}) = A\boldsymbol{0} = 0$.

$$\mathrm{Var}(\boldsymbol{W}) = \mathrm{Var}(A\boldsymbol{Z}) = A\mathrm{Var}(\boldsymbol{Z})A^t = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Since S and T are uncorrelated and multivariate normal, they are independent. (Joint density factorizes.)

meth seen ⇓

(ii) The inverse transformation is given by $X = R\cos\Theta$, $Y = R\sin\Theta$.

4, B

$$\frac{\partial x}{\partial r} = \cos\theta \qquad \frac{\partial y}{\partial r} = \sin\theta$$
$$\frac{\partial x}{\partial \theta} = -r\sin\theta \qquad \frac{\partial y}{\partial \theta} = r\cos\theta$$

The Jacobian of the inverse transformation is therefore $r(\cos^2\theta + \sin^2\theta) = r$.
By the multivariate transformation theorem,

$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right), \qquad r > 0, 0 < \theta < 2\pi.$$

By the factorization theorem, since the joint density factorizes, R and Θ are independent random variables.

meth seen ⇓

(iii) $V = \tan\Theta$. We find the marginal density of Θ from the previous part

3, D

$$f_\Theta(\theta) = \int_0^\infty \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right)\, dr = \frac{1}{2\pi}, 0 < \theta < 2\pi,$$

Away from the origin, $\tan$ is a 2-1 mapping since (X, Y) and $(-X, -Y)$ correspond to the same value of $\tan\Theta$. Considering the symmetries of $\tan$, we have

$$\Pr(V \le v) = \Pr(\tan\Theta \le v) = \frac{1}{2} + 2\Pr(0 \le \Theta \le \arctan v) = \frac{1}{2} + \frac{2}{2\pi}\arctan(v) = \frac{1}{2} + \frac{1}{\pi}\arctan v.$$

Differentiating gives the density of V:

$$f_V(v) = \frac{1}{\pi(1+v^2)}, \qquad -\infty < v < \infty$$

unseen ⇓

(b) (i) For any $y > 0$ we have

2, B

$$\{Z - \mu \ge \alpha\} = \{Z - \mu + y \ge \alpha + y\}$$

Note that for any $r, s \in \mathbf{R}$, $r^2 \ge s^2$ iff $r^2 - s^2 \ge 0$ iff $(r+s)(r-s) \ge 0$. So for $\alpha > 0$ and $y > 0$ we have

$$\{Z - \mu + y \geq \alpha + y\} \subseteq \{(Z - \mu + y)^2 \geq (\alpha + y)^2\},$$

so that

$$\Pr(Z - \mu \geq \alpha) \leq \Pr\left((Z - \mu + y)^2 \geq (\alpha + y)^2\right),$$

unseen ⇓

3, D

(ii) Applying Markov's inequality to the non-negative random variable $(Z-\mu+y)^2$ gives

$$\Pr\left((Z - \mu + y)^2 \geq (\alpha + y)^2\right) \leq \frac{\mathrm{E}\left((Z - \mu + y)^2\right)}{(\alpha + y)^2} = \frac{\mathrm{E}\left((Z - \mu)^2 + 2y(Z - \mu) + y^2\right)}{(\alpha + y)^2}.$$

Now $\mathrm{E}\left((Z - \mu)^2\right) = \mathrm{Var}(Z) = \sigma^2$ and $\mathrm{E}\left((Z - \mu)\right) = 0$, so that we have for all $y > 0$,

$$\Pr(Z - \mu \geq \alpha) \leq \frac{\sigma^2 + y^2}{(\alpha + y)^2}.$$

We now seek the best upper bound by differentiating the right hand expression with respect to y:

$$\frac{d}{dy}\left(\frac{\sigma^2 + y^2}{(\alpha + y)^2}\right) = \frac{2y}{(\alpha + y)^2} - \frac{2(\sigma^2 + y^2)}{(\alpha + y)^3}.$$

For the derivative to be equal to zero, we need

$$\frac{y}{(\alpha + y)^2} = \frac{(\sigma^2 + y^2)}{(\alpha + y)^3},$$

so that $y = \frac{\sigma^2}{\alpha}$. This gives a lowest upper bound of

$$\frac{\sigma^2 + \left(\frac{\sigma^2}{\alpha}\right)^2}{\left(\alpha + \frac{\sigma^2}{\alpha}\right)^2} = \frac{\sigma^2}{\sigma^2 + \alpha^2}$$

as required.

unseen ⇓

4, D

(iii) We seek a random variable Z supported on the two point set $\{a, b\}$ where $a < b$. We require $\mathrm{E}(Z) = 0$, $\mathrm{Var}(Z) = \sigma^2$ and

$$\Pr(Z \geq \alpha) = \frac{\sigma^2}{\sigma^2 + \alpha^2},$$

so that the inequality is sharp. Call right-hand probability p. Then

$$\mathrm{E}(Z) = a(1 - p) + bp = 0$$

$$\mathrm{Var}(Z) = a^2(1 - p) + b^2 p = \sigma^2.$$

Substituting for a gives $a = -\frac{bp}{1-p}$ and so

$$\frac{b^2 p^2}{1 - p} + b^2 p = \sigma^2,$$

so that

$$b^2 p(p + (1-p)) = \sigma^2(1-p)$$

and $b^2 = \sigma^2 \frac{1-p}{p} = \alpha^2$.
So then $b = \alpha$ (as $b > 0$) and $\alpha = -\frac{bp}{1-p} = -\frac{\sigma^2}{\alpha}$.
For this choice of Z, the upper bound is attained, so the bound is sharp: no better bound is possible in general.

4. (a) See figure

meth seen ⇓

3, A

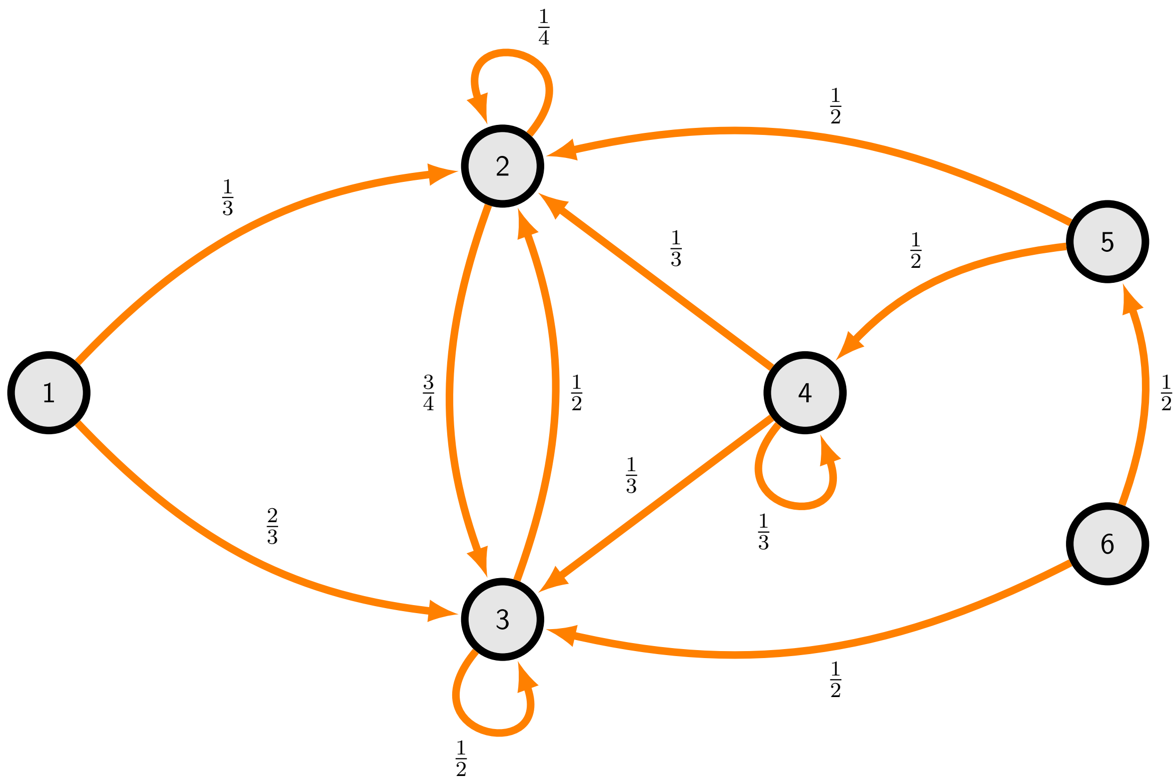

Figure 1: The six-state Markov chain in (a).

meth seen ⇓

3, A

(b) By inspecting the connectivity of the diagram, the communicating classes are $\{1\}$ (transient), $\{2, 3\}$ (recurrent), $\{4\}$ (transient), $\{5\}$ (transient) $\{6\}$ (transient).

meth seen ⇓

3, A

(c) The period of a state i is $d(i) = \gcd\{n \geq 1 : p_{ii}^n > 0\}$. Note that period is a class property, so it is enough to determine d for one member of each class.

$d(1) = d(5) = d(6) = \gcd(\emptyset) = 0$ (trivially).

$d(2) = d(3) = 1.$

$d(4) = 1.$

meth seen ⇓

3, B

(d)

$$\begin{aligned}
\Pr(X_1 = 2) &= \sum_{i \in \mathcal{E}} \Pr(X_1 = 2 | X_0 = i) \Pr(X_0 = i) \\
&= \frac{1}{3} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4} + 0 \times \frac{1}{4} \\
&= \frac{1}{4}\left(\frac{1}{3} + \frac{1}{4} + \frac{1}{2}\right) = \frac{1}{4} \times \frac{13}{12} = \frac{13}{48}.
\end{aligned}$$

unseen ⇓

3, C

(e) We start in $\{2, 3\}$, which is an isolated set. Hence can consider a reduced two-state chain with transition matrix

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

From p79 of the notes, this has stationary distribution

$$\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) = \left(\frac{\frac{1}{2}}{\frac{3}{4}+\frac{1}{2}}, \frac{\frac{3}{4}}{\frac{3}{4}+\frac{1}{2}}\right) = \left(\frac{2}{5}, \frac{3}{5}\right),$$

so that the limiting distribution on the full state space starting from π_0 is

$$\left(0, \frac{2}{5}, \frac{3}{5}, 0, 0, 0\right)$$

meth seen ⇓

(f)

2, A

$$\begin{aligned} L(P) &= \Pr(Y_0 = y_0, \ldots, Y_n = y_n) \\ &= \Pr(Y_n = y_n | Y_{n-1} = y_{n-1}, \ldots, Y_0 = y_0) \Pr(Y_{n-1} = y_{n-1}, \ldots, Y_0 = y_0). \end{aligned}$$

Using the Markov property, this is

$$\begin{aligned} L(P) = \Pr(Y_0 = y_0, \ldots, Y_n = y_n) &= \Pr(Y_n = y_n | Y_{n-1} = y_{n-1}) \Pr(Y_{n-1} = y_{n-1}, \ldots, Y_0 = y_0) \\ &= P(y_{n-1}, y_n) \Pr(Y_{n-1} = y_{n-1}, \ldots, Y_0 = y_0), \end{aligned}$$

where $P = (p_{ij})$ is the transition matrix. Applying the same argument to the rightmost expression, we get

$$\begin{aligned} L(P) &= P(y_{n-1}, y_n) P(y_{n-2}, y_{n-1}) \ldots P(y_0, y_1) \Pr(Y_0 = y_0) \\ &= \Pr(Y_0 = y_0) \prod_{i,j \in \{1,2\}} p_{ij}^{n_{ij}} \end{aligned}$$

So then

$$l(\alpha, \beta) = \sum_{i,j \in \{1,2\}} n_{ij} \log(p_{ij}) = n_{11} \log(1-\alpha) + n_{12} \log \alpha + n_{21} \log \beta + n_{22} \log(1-\beta).$$

Differentiating to find the maximum

$$\frac{\partial l}{\partial \alpha} = \frac{-n_{11}}{1-\alpha} + \frac{n_{12}}{\alpha}.$$

$$\frac{\partial l}{\partial \beta} = \frac{-n_{22}}{1-\beta} + \frac{n_{12}}{\beta}.$$

3, B

Setting these partial derivatives to zero, we see that $\widehat{\alpha} = \frac{n_{12}}{n_{11}+n_{12}}$ and $\widehat{\beta} = \frac{n_{21}}{n_{21}+n_{22}}$ are critical points of the log-likelihood function. Note that inspecting second derivatives guarantees that these critical points are indeed maxima, for

$$\frac{\partial^2 l}{\partial \alpha^2} = \frac{-n_{11}}{(1-\alpha)^2} - \frac{n_{12}}{\alpha^2} < 0,$$

and symmetrically for β (the mixed second partial derivative is zero).

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.
Please record below, some brief but non-trivial comments for students about how well (or otherwise) the questions were answered. For example, you may wish to comment on common errors and misconceptions, or areas where students have done well. These comments should note any errors in and corrections to the paper. These comments will be made available to students via the MathsCentral Blackboard site and should not contain any information which identifies individual candidates. Any comments which should be kept confidential should be included as confidential comments for the Exam Board and Externals. If you would like to add formulas, please include a sperate pdf file with your email.

ExamModuleCode	QuestionNumber	Comments for Students
MATH50010	1	Question generally well-answered. Some proofs by induction in a (ii) were unsound. Counterexamples in a (iii) and a (iv) were often found, though a surprising number of candidates thought these parts were true. Part b was well done by almost all candidates
MATH50010	2	This question was addressed reasonably well by most candidates. Parts (a) and (b) were generally done well, with few issues. Part (c) proved challenging for many, with the majority of candidates dropping marks for failing to sufficiently justify their choice of g(x). Another common mistake here, particularly in part (c)(ii), was confusing continuity of g(x) with continuity of the resulting RV, Y=g(X); the former does not guarantee the latter. Parts (d) and (e) both went reasonably well; candidates that achieved full marks in these sections were able to fully justify each step of their derivations. Part (f) was approached in one of two ways: the more concise approach was to directly show that the definition of convergence in probability is satisfied (as in the provided solutions); the alternative approach was to show that the cdf converges to that of a constant random variable (i.e. discrete, with degenerate distribution), equal to zero, and to use the fact that convergence in distribution to a constant RV implies convergence in probability. A common mistake with the latter approach was to state that convergence in distribution implies convergence in probability; this is not true in general. Part (f) was generally done well, with most candidates spotting the link between the cdf of X_n, and that of an Exponential distribution.

MATH50010	3	Question 3(a). i was well answered, only a small part of the students did not provide any comments about independence. A great part of the students managed to calculate the Jacobian correctly in Question 3(a). ii. Question 3 was the one the students struggled the most. The first question, b.i was well answered, with details about the proof. However, in question 3(b). ii a substantial number of students didn't provide any argument for the proof or used the wrong inequality. Question 3(b). iii was the most challenging one, where just two students managed to provide a compelling example.
MATH50010	4	In general, some key details and/or justifications were left out (causing the loss of marks - without a proper explanation, the marker cannot follow the arguments to, at least, try to give partial marks to the answers). More specific issues: question 4)c) asked about the period of the classes, therefore only saying that they were (or were not) aperiodic did not answer the question. 4e) Many students did not justify why the stationary distribution equals the limiting distribution. 4 f) The log likelihood needed to be derived as well as the MLE of alpha and beta and so just copying the final results from coursework was not allowed.