Imperial College
London

M2S1

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2017

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

Probability and Statistics II

Date: Monday 22 May 2017

Time: 14:00 - 16:00

Time Allowed: 2 Hours

This paper has 4 Questions.

Candidates should start their solutions to each question in a new main answer book.

Supplementary books may only be used after the relevant main book(s) are full.

All required additional material will be provided.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.
- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.
- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

Raw Mark	Up to 12	13	14	15	16	17	18	19	20
Extra Credit	0	½	1	1 ½	2	2 ½	3	3 ½	4

- Each question carries equal weight.
- Calculators may not be used.

1. Suppose that, measured in centimeters, heights of adult men in the UK follow a normal distribution with mean 175cm and standard deviation 5cm, while heights of adult women in the UK follow a normal distribution with mean 165cm and standard deviation 5cm. Suppose also that 50% of UK adults are women. Let $X_1, \dots, X_n$ be the heights (in cm) and $Y_1, \dots, Y_n$ be the sexes of n adults selected at random (with replacement) from the UK population, where $Y_i = 1$ if selected adult i is a man and $Y_i = 0$ if she is a woman.

(a) Suppose that the first randomly selected adult is a man. What is the probability that his height is greater than 175cm? Now suppose that the first randomly selected adult is a woman. What is the probability that her height is greater than 175cm? In both cases, report an (approximate) numeric answer.

(b) Compute the marginal probability, $p = \Pr(X_1 > 175)$. Again report an (approximate) numeric answer.

(c) Suppose X_1 is found to be 165. What is the probability that this person is a man? Your simplified answer may involve transcendental or irrational numbers (e.g., π, e, $\sqrt{2}$).

(d) In terms of p, k, and n, give an exact expression for the probability that more than k of the n selected adults are taller than 175cm. Also, assuming n is large, give an approximation to this probability. Your approximation should make use of the continuity correction and be given in terms of p, k, n, and Φ, the standard normal cumulative distribution function.

(e) Derive the mean and variance of X_1.

(f) Derive the mean and variance of $X_1 - X_2$.

2. Let $Z_1, \dots, Z_K$ be normal random variables such that $\boldsymbol{Z} \sim \mathsf{N}_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{Z} = (Z_1, \dots, Z_K)^\top$, $\boldsymbol{\mu}$ is a $(K \times 1)$ vector of zeros, and $\boldsymbol{\Sigma}$ is a $(K \times K)$ variance-covariance matrix.

(a) Suppose $K \geq 2$ and $\boldsymbol{\Sigma}$ is a $(K \times K)$ identity matrix.
You may state your answer to parts (i) – (iv) without justification.

(i) What is the distribution of $5Z_1 - Z_2$?

(ii) What is the distribution of $Z_1^2 + Z_2^2$?

(iii) What is the distribution of $5(Z_1^2 + Z_2^2)$?

(iv) What is the *joint* distribution of $\bar{Z} = \dfrac{1}{K}\sum_{i=1}^{K} Z_i$ and $T^2 = \sum_{i=1}^{K}(Z_i - \bar{Z})^2$?

(b) Suppose $K = 3$ and $\boldsymbol{\Sigma}$ is a $(K \times K)$ identity matrix.

(i) What is the *joint* distribution of $Y_1 = Z_1 - Z_2$, $Y_2 = Z_1 + Z_2$, and $Y_3 = Z_1 + Z_2 + Z_3$?

(ii) What is the correlation between Y_2 and Y_3 ?

(c) In this part, suppose $K = 2$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

(i) Suppose $\rho = 0$. Are Z_1 and Z_2 independent? Justify your answer.

(ii) Show that $Z_2 | Z_1 = z_1 \sim \mathsf{N}(\rho z_1, 1 - \rho^2)$ for any $\rho \in (-1, 1)$. You may use the fact that $Z_1 \sim \mathsf{N}(0, 1)$.

3. (a) Carefully state the central limit theorem for independent and identically distributed random variables as given in the notes. What additional assumption did we require in our proof of the central limit theorem?

(b) Let $(X_1, \ldots, X_n)$ be a random sample from a distribution with cumulative distribution function given by,

$$F_X(x) = 1 - \frac{1}{1+\lambda x}, \quad \text{for} \quad x > 0,$$

where λ is a positive parameter. Let

$$U_n = \max\{X_1, \ldots, X_n\}/n \ \text{ and } \ V_n = n \min\{X_1, \ldots, X_n\}.$$

(i) Derive expressions for the cumulative distribution functions of U_n and V_n.

(ii) Find the limiting distribution of V_n as $n \longrightarrow \infty$.

(c) State and prove the weak law of large numbers.

4. Suppose that $(X_1, \ldots, X_n)$ is a random sample from the probability density function given by

$$f_X(x) = \frac{4\alpha^{-3/2}}{\sqrt{\pi}} x^2 e^{-x^2/\alpha} \qquad \text{for} \quad x > 0,$$

where α is a positive parameter. This problem considers estimators of α and their properties. You may find it useful to recall that for positive integer, n, $\Gamma(n) = (n-1)!$ and $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$ so that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \qquad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}, \qquad \Gamma\left(\frac{5}{2}\right) = \frac{3\sqrt{\pi}}{4}, \quad \text{and} \quad \Gamma\left(\frac{7}{2}\right) = \frac{15\sqrt{\pi}}{8}.$$

(a) Define the terms *unbiased estimator* and *consistent estimator*.

(b) Show that $\mathrm{E}(X_1^n) = 2\Gamma(\frac{n+3}{2})\alpha^{n/2}/\sqrt{\pi}$ and derive $\mathrm{Var}(X_1)$.

(c) Derive the method of moments and maximum likelihood estimators of α.

(d) Derive an expression for the mean square error of the maximum likelihood estimator of α.

(e) Derive an explicit formula for the bias of the method of moments estimator of α. Is this estimator unbiased? Is it asymptotically unbiased?

DISCRETE DISTRIBUTIONS

	range $\mathbb{X}$	*parameters*	*pmf* f_X	*cdf* F_X	$\mathrm{E}[X]$	$\mathrm{Var}[X]$	*mgf* M_X
$Bernoulli(\theta)$	$\{0,1\}$	$\theta \in (0,1)$	$\theta^x(1-\theta)^{1-x}$		θ	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
or	$\{0,1,2,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n+x-1}{x}\theta^n(1-\theta)^x$		$\dfrac{n(1-\theta)}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta}{1-e^t(1-\theta)}\right)^n$

The PDF of the *multivariate normal distribution* is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{K/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big\},$$

for $\boldsymbol{x} \in \mathbb{R}^K$ with $\boldsymbol{\Sigma}$ a $(K \times K)$ variance-covariance matrix and $\boldsymbol{\mu}$ a $(K \times 1)$ mean vector.

The location/scale transformation $Y = \mu + \sigma X$ gives

$$f_Y(y) = \frac{1}{\sigma}f_X\left(\frac{y-\mu}{\sigma}\right) \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right)$$

$$M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathrm{E}[Y] = \mu + \sigma\mathrm{E}[X] \qquad \mathrm{Var}[Y] = \sigma^2\mathrm{Var}[X]$$

The *gamma function* is given by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$.

CONTINUOUS DISTRIBUTIONS

		parameters	*pdf*	*cdf*	$\mathrm{E}[X]$	$\mathrm{Var}[X]$	*mgf*
$Uniform(\alpha,\beta)$ (stand. model $\alpha=0,\beta=1$)	(α,β)	$\alpha<\beta\in\mathbb{R}$	$\dfrac{1}{\beta-\alpha}$	$\dfrac{x-\alpha}{\beta-\alpha}$	$\dfrac{(\alpha+\beta)}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	$\dfrac{e^{\beta t}-e^{\alpha t}}{t\,(\beta-\alpha)}$
$Exponential(\lambda)$ (stand. model $\lambda=1$)	$\mathbb{R}^+$	$\lambda\in\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda-t}\right)$
$Gamma(\alpha,\beta)$ (stand. model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$		$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta-t}\right)^\alpha$
$Weibull(\alpha,\beta)$ (stand. model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$	$1-e^{-\beta x^\alpha}$	$\dfrac{\Gamma\left(1+1/\alpha\right)}{\beta^{1/\alpha}}$	$\dfrac{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2}{\beta^{2/\alpha}}$	
$Normal(\mu,\sigma^2)$ (stand. model $\mu=0,\sigma=1$)	$\mathbb{R}$	$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\left\{1+\dfrac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		$0\quad(\text{if } \nu>1)$	$\dfrac{\nu}{\nu-2}\quad(\text{if } \nu>2)$	
$Pareto(\theta,\alpha)$	$\mathbb{R}^+$	$\theta,\alpha\in\mathbb{R}^+$	$\dfrac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$	$1-\left(\dfrac{\theta}{\theta+x}\right)^\alpha$	$\dfrac{\theta}{\alpha-1}$ (if $\alpha>1$)	$\dfrac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$ (if $\alpha>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

Imperial College London

Course: M2S1
Setter:
Checker:
Editor:
External: external
Date: January 7, 2017

MSc EXAMINATIONS (MATHEMATICS)

May 2017

M2S1

Probability and Statistics II

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. (a) *[Seen Method]* First, given $Y_1 = 1$, we have $X_1 \sim \mathrm{Norm}(175, \sigma = 5)$ so

$$\Pr(X_1 > 175 \mid Y_1 = 1) = \Pr\left(Z > \frac{175 - 175}{5}\right) = \Pr(Z > 0) = 0.50,$$

where Z is a standard normal random variable. Similarly, if $Y_1 = 0$, $X_1 \sim \mathrm{Norm}(165, \sigma = 5)$ so

$$\Pr(X_1 > 175 \mid Y_1 = 0) = \Pr\left(Z > \frac{175 - 165}{5}\right) = \Pr(Z > 2) \approx 0.025.$$

(b) *[Seen Method]* Using the results from part (a) and the law of total probability,

$$\begin{aligned} p = \Pr(X_1 > 175) &= \Pr(X_1 > 175 \mid Y_1 = 1)\Pr(Y_1 = 1) + \Pr(X_1 > 175 \mid Y_1 = 0)\Pr(Y_1 = 0) \\ &\approx \frac{1}{2}(0.50 + 0.025) \approx 0.263. \end{aligned}$$

(c) *[Seen Similar]* The conditional probability that $Y_1 = 1$ (i.e., that the first selected adult is a man) given a height of $x_1 = 165$cm is

$$\begin{aligned} f_{Y|X}(1 \mid 165) &= \frac{f_{X|Y}(165 \mid 1) f_Y(1)}{f_{X|Y}(165 \mid 1) f_Y(1) + f_{X|Y}(165 \mid 0) f_Y(0)} \\ &= \frac{\exp\left\{-\frac{(165-175)^2}{50}\right\}}{\exp\left\{-\frac{(165-175)^2}{50}\right\} + \exp\left\{-\frac{(165-165)^2}{50}\right\}} = \frac{e^{-2}}{e^{-2}+1} = \frac{1}{1+e^2}. \end{aligned}$$

(d) *[Seen Less Abstract Form]* The number of the selected adults who are taller than 175cm is a random variable with distribution $\textsc{Binomial}(n, p)$, where p was derived in part (b). Thus, the probability that more than k of the selected adults is taller than 175cm is

$$\sum_{i=k+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \approx \Pr\left(Z \geq \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) = 1 - \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

(e) *[Seen Similar]* We can use the laws of iterated expectations and total variance:

$$\mathrm{E}(X_1) = \mathrm{E}\Big[\mathrm{E}(X_1 \mid Y_1)\Big] = \mathrm{E}\Big[175 Y_1 + 165(1 - Y_1)\Big] = 170.$$

and

$$\begin{aligned} \mathrm{Var}(X_1) &= \mathrm{E}\Big[\mathrm{Var}(X_1|Y_1)\Big] + \mathrm{Var}\Big[\mathrm{E}(X_1|Y_1)\Big] = \mathrm{E}\Big[25Y_1 + 25(1-Y_1)\Big] + \mathrm{Var}\Big[175Y_1 + 165(1-Y_1)\Big] \\ &= 25 + \mathrm{Var}\Big[10Y_1 + 165\Big] = 5 + \frac{100}{4} = 50. \end{aligned}$$

(f) *[Seen Method]* Because X_1 and X_2 are independent and identically distributed,

$$\mathrm{E}(X_1 - X_2) = \mathrm{E}(X_1) - \mathrm{E}(X_2) = 0$$

and

$$\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 50 + 50 = 100.$$

2. (a) *[Seen]* The distributions are

   (i) $\text{Norm}(0, \sigma^2 = 26)$.

   (ii) χ^2 with 2 degrees of freedom or, equivalently, $\text{Gamma}(1, \frac{1}{2})$, using the parameterization in the formula sheet.

   (iii) $\text{Gamma}(1, \frac{1}{10})$, using the parameterization in the formula sheet.

   (iv) $\bar{Z}$ and T^2 are independent with marginal distributions, $\bar{Z} \sim \text{Norm}(0, \sigma^2 = \frac{1}{K})$ and $T^2 \sim \chi^2$ with $K-1$ degrees of freedom.

(b) (i) *[Seen Method]* Setting $\boldsymbol{M} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, we can write $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{Z}$, where $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^\top$. Using a theorem in the notes, we have that $\boldsymbol{Y} \sim \mathsf{N}_K(\boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^\top)$, i.e.,

$$\boldsymbol{Y} \sim \mathsf{N}_K\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 3 \end{pmatrix}\right),$$

with $K = 3$.

   (ii) *[Seen Similar]* Using the result from part (i), the correlation between Y_2 and Y_3 is

$$\text{Correlation}(Y_2, Y_3) = \frac{\text{Cov}(Y_2, Y_3)}{\sqrt{\text{Var}(Y_2)\text{Var}(Y_3)}} = \frac{2}{\sqrt{6}}.$$

(c) (i) *[Seen Similar]* The joint probability density function of Z_1 and Z_2 is

$$f_{Z_1Z_2}(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right\},$$

for $-\infty < z_1, z_2 < \infty$. Setting $\rho = 0$,

$$f_{Z_1Z_2}(z_1, z_2) = \frac{1}{2\pi} \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\} = \frac{1}{2\pi} e^{-z_1^2/2} \times e^{-z_2^2/2}.$$

Thus, by the factorization theorem Z_1 and Z_2 are independent.

   (ii) *[Seen]* By the definition of a conditional probability density function,

$$\begin{aligned} f_{Z_2|Z_1}(z_2|z_1) &= \frac{f_{Z_1Z_2}(z_1, z_2)}{f_{Z_1}(z_1)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-z_1^2/2\right\}} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \frac{\exp\left\{-\frac{z_1^2}{2} - \frac{(z_2 - \rho z_1)^2}{2(1-\rho^2)}\right\}}{\exp\left\{-z_1^2/2\right\}} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(z_2 - \rho z_1)^2}{2(1-\rho^2)}\right\}, \end{aligned}$$

for $-\infty < z_2 < \infty$, so $Z_2 | Z_1 = z_1 \sim \text{N}(\rho z_1, 1 - \rho^2)$. (The third equality comes from completing the square in z_2.) For an alternative solution simply note

$$f_{Z_2|Z_1}(z_2|z_1) \propto f_{Z_1Z_2}(z_1, z_2) \propto \exp\left\{-\frac{1}{2(1-\rho^2)}(z_2^2 - 2\rho z_1 z_2)\right\} \propto \exp\left\{-\frac{(z_2 - \rho z_1)^2}{2(1-\rho^2)}\right\},$$

for $-\infty < z_2 < \infty$, which again implies the result.

3. (a) *[Seen]* Central Limit Theorem: Let $X_1, X_2, \ldots$ be a countable sequence of iid random variables with finite mean, $\mathrm{E}(X_i) = \mu$, and finite variance, $\mathrm{Var}(X_i) = \sigma^2$. Also let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, and $G_n(x)$ be the CDF of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y = \Phi(x),$$

i.e., $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{D}} \mathrm{Norm}(0,1)$.

Our proof of the Central Limit Theorem also requires the moment generating function of X_i to exist.

(b) (i) *[Seen]* The cumulative distribution functions of U_n and V_n are

$$\begin{aligned} F_{U_n}(u) &= \Pr(U_n \le u) = \Pr(\max\{X_1,\ldots,X_n\}/n \le u) = \Pr(\max\{X_1,\ldots,X_n\} \le nu) \\ &= \Pr(\text{each of the } X_i \text{ is } \le nu) = \prod_{i=1}^{n} F_X(nu) = \left(\frac{\lambda nu}{1+\lambda nu}\right)^n, \quad \text{for } u > 0. \end{aligned}$$

and

$$\begin{aligned} F_{V_n}(v) &= \Pr(V_n \le v) = \Pr(\min\{X_1,\ldots,X_n\} \le v/n) = 1 - \Pr(\text{each } X_i \text{ is } > v/n) \\ &= 1 - \prod_{i=1}^{n}\Big(1 - F_X(v/n)\Big) = 1 - \left(\frac{1}{1+\lambda v/n}\right)^n, \quad \text{for } v > 0. \end{aligned}$$

(ii) *[Seen]* To find the limiting distribution of V_n, note that

$$\lim_{n\to\infty} F_{V_n}(v) = \lim_{n\to\infty} 1 - \left(1 + \frac{\lambda v}{n}\right)^{-n} = 1 - \exp\{-\lambda v\}$$

so that $V_n \xrightarrow{\mathcal{D}} \textsc{Exponential}(\lambda)$.

(c) *[Seen]* Weak Law of Large Numbers: Let $X_1, X_2, \ldots$ be an countable sequence of independent and identically distributed random variables with finite first and second moments, $\mathrm{E}(X_i) = \mu$ and $\mathrm{E}(X_i^2)$, also let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Then

$$\lim_{n\to\infty} \Pr\Big(\big|\bar{X}_n - \mu\big| < \epsilon\Big) = 1, \quad \forall \epsilon > 0,$$

i.e., $\bar{X}_n \xrightarrow{\mathcal{P}} \mu$.

Proof: Because $\mathrm{Var}(\bar{X}) = \sigma^2/n$, by Chebychev's inequality,

$$0 \le \Pr\left[(\bar{X}_n - \mu)^2 \ge t^2\right] \le \frac{\sigma^2}{nt^2} \longrightarrow 0 \text{ as } n \to \infty$$

for any $t > 0$ where $\sigma^2 = \mathrm{Var}(X_i)$. Thus $\lim_{n\to\infty} \Pr\Big(\big|\bar{X}_n - \mu\big| < \epsilon\Big) = 1$.

4. (a) *[Seen]* We say T_n is a consistent estimator of θ if T_n converges in probability to θ, i.e., $T_n \xrightarrow{\mathcal{P}} \theta$.

We say T is an unbiased estimator of θ if $\mathrm{E}(T) = \theta$.

(b) *[Seen Part]* To derive $\mathrm{E}(X_1^n)$,

$$\begin{aligned}
\mathrm{E}(X_1^n) &= \int_0^\infty x^n f_X(x)\mathrm{d}x = \frac{4\alpha^{-3/2}}{\sqrt{\pi}} \int_0^\infty x^{n+2} e^{-x^2/\alpha}\mathrm{d}x \\
&= \frac{2\alpha^{-3/2}}{\sqrt{\pi}} \int_0^\infty z^{(n+1)/2} e^{-z/\alpha}\mathrm{d}z = \frac{2\alpha^{-3/2}}{\sqrt{\pi}} \Gamma\left(\frac{n+3}{2}\right) \alpha^{(n+3)/2} = \frac{2\alpha^{n/2}}{\sqrt{\pi}} \Gamma\left(\frac{n+3}{2}\right),
\end{aligned}$$

where the third equality follows from the change of variable $z = x^2$ and the fourth from identifying the integrand as an unnormaizlied gamma probability density function. Finally,

$$\mathrm{Var}(X_1) = \mathrm{E}(X_1^2) - \Big(\mathrm{E}(X)\Big)^2 = \frac{3\alpha}{2} - \frac{4\alpha}{\pi} = \alpha\left(\frac{3}{2} - \frac{4}{\pi}\right).$$

(c) *[Seen]* To compute the Method of Moment estimator we set $\bar{X} = \mathrm{E}(X)$, where $\mathrm{E}(X)$ is a function of α, and solve. Using the result from part (b) with $n = 1$, we solve $\bar{X} = 2\sqrt{\frac{\alpha}{\pi}}$ to obtain the estimator $\hat{\alpha}_{\mathrm{MoM}} = \pi\bar{X}^2/4$.

To compute the Maximum Likelihood estimator, we derive the likelihood function,

$$L(\alpha \mid \boldsymbol{x}) = \prod_{i=1}^n f_{X_i}(x_i \mid \alpha) = \prod_{i=1}^n \frac{4\alpha^{-3/2}}{\sqrt{\pi}} x_i^2 e^{-x_i^2/\alpha}$$

and the log-likelihood function

$$\ell(\alpha \mid \boldsymbol{x}) = \sum_{i=1}^n \left(-\frac{3}{2}\log(\alpha) + 2\log(x_i) - \frac{x_i^2}{\alpha} + \log\left(\frac{4}{\sqrt{\pi}}\right)\right),$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$. To find the maximizer of $\ell(\alpha \mid \boldsymbol{x})$, we take the first derivative, set it equal to zero, and solve,

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^n \left(-\frac{3}{2\alpha} + \frac{x_i^2}{\alpha^2}\right) = -\frac{3n}{2\alpha} + \frac{\sum_{i=1}^n x_i^2}{\alpha^2} = 0,$$

to obtain the candidate estimator $\hat{\alpha}_{\mathrm{mle}} = \frac{2}{3n}\sum_{i=1}^n X_i^2$. Using the second-derivative test to check if $\hat{\alpha}_{\mathrm{mle}} = \frac{2}{3n}\sum_{i=1}^n X_i^2$ maximises the log-likelihood,

$$\left.\frac{\partial^2 \ell}{\partial \alpha^2}\right|_{\alpha=\hat{\alpha}_{\mathrm{mle}}} = \left[\frac{3n}{2\alpha^2} - \frac{2\sum_{i=1}^n x_i^2}{\alpha^3}\right]_{\alpha=\hat{\alpha}_{\mathrm{mle}}} = -\frac{3n}{2\hat{\alpha}_{\mathrm{mle}}^2} = -\frac{27n^3}{8(\sum_{i=1}^n x_i^2)^2} < 0.$$

Hence $\hat{\alpha}_{\mathrm{mle}}$ is a maximum likelihood estimator for α.

(d) *[Unseen]* We start by deriving the expectation and variance of $\hat{\alpha}_{\mathrm{mle}}$,

$$\begin{aligned}
\mathrm{E}(\hat{\alpha}_{\mathrm{mle}}) &= \frac{2}{3n}\sum_{i=1}^n \mathrm{E}(X_i^2) = \frac{2}{3n}\frac{3\alpha n}{2} = \alpha \\
\mathrm{Var}(\hat{\alpha}_{\mathrm{mle}}) &= \frac{4}{9n^2}\sum_{i=1}^n \mathrm{Var}(X_i^2) = \frac{4}{9n^2}\sum_{i=1}^n \Big(\mathrm{E}(X_i^4) - (\mathrm{E}(X_i^2))^2\Big) = \frac{4n}{9n^2}\left(\frac{15\alpha^2}{4} - \frac{9\alpha^2}{4}\right) = \frac{2\alpha^2}{3n}.
\end{aligned}$$

Because $\hat{\alpha}_{\mathrm{mle}}$ is unbiased, $\mathrm{mse}(\hat{\alpha}_{\mathrm{mle}}) = \mathrm{Var}(\hat{\alpha}_{\mathrm{mle}}) = 2\alpha^2/3n$.

(e) *[Unseen]* To obtain the bias of $\hat{\alpha}_{\text{mom}}$,

$$\begin{aligned}
\text{Bias}(\hat{\alpha}_{\text{mom}}) &= \mathrm{E}(\hat{\alpha}_{\text{mom}}) - \alpha = \mathrm{E}\left(\frac{\pi}{4}\bar{X}^2\right) - \alpha = \frac{\pi}{4}\left(\mathrm{Var}(\bar{X}) + (\mathrm{E}(\bar{X}))^2\right) - \alpha \\
&= \frac{\pi}{4}\left(\frac{\mathrm{Var}(X_1)}{n} + (\mathrm{E}(X_1))^2\right) - \alpha = \frac{\pi}{4}\left(\frac{\alpha}{n}\left(\frac{3}{2} - \frac{4}{\pi}\right) + \frac{4\alpha}{\pi}\right) - \alpha = \frac{\alpha}{n}\left(\frac{3\pi}{8} - 1\right).
\end{aligned}$$

Thus, $\hat{\alpha}_{\text{mom}}$ is not unbiased, but it is asymptotically unbiased.

Marks:

1. (a) 2 marks. One mark for method and one mark for numerical answers.
   (b) 2 marks. One mark for method and one mark for numerical answer.
   (c) 4 marks. One mark for using Bayes Theorem; one for using it correctly; one for plugging in normal probability density functions; and one for final simplified answer.
   (d) 5 marks. Two marks for identifying the number of selected adults taller than 175cm as a Binomial random variable; one mark for exact binomial probability; one mark for normal approximation; and one mark for continuity correction.
   (e) 5 marks. One mark each for trying to apply law of iterated expectations and law of total variance; one mark for writing conditional expectation and variance as linear functions of Y_1 and $1 - Y_1$; and one mark each for correct numerical answer for mean and variance. (Full marks should be given if the correct answers are derived without writing the conditional expectation/variance as a linear function.)
   (f) 2 marks. One mark for mean and one mark for variance.

2. (a) (i) 1 mark. One half mark for "normal" and one half mark for moments.
   (ii) 1 mark. One half mark for "chi square" or "gamma" and one half mark for degrees of freedom or parameters of gamma distribution.
   (iii) 1 mark. One half mark for "gamma" and one half mark for parameters.
   (iv) 3 marks. One mark each for "independent" and for each of the marginal distributions.
   (b) (i) 4 marks. One mark for "multivariate normal", one mark for mean vector, and two marks for variance. Give at least one mark of partial credit for correctly deriving $\boldsymbol{M}$.
   (ii) 2 marks. One mark for formula for correlation and one mark for final answer. (Give full marks if answer is correct based on the answer given in part i.)
   (c) (i) 4 marks. Two marks for arguing independence by factoring the probability density function; one mark for the joint probability density function, and one mark for factoring the probability density function.
   (ii) 4 marks. One mark for correct starting point (conditional density is equal to the given ratio or is proportional to the joint density); two marks for correct algebra; and one mark for arguing/noticing that the derived probability density function is that of the desired conditional density.

3. (a) 5 points. One mark each for (i) infinite or countable sequence of iid random variables (either in words or notation), (ii) finite mean and variance, (iii) standardizing $\bar{X}_n$, (iv) careful statement of conclusion (either in terms of the limiting cumulative distribution function of standardized mean or in terms of convergence in distribution), and (v) noting that our proof requires existance of the moment generating function.

   (b) (i) 6 points. Two marks each for each derivation and one marks for each of the final answers. Full marks should be given for correctly applying general formulae for the cumulative distribution function of an order statistic.

   (ii) 3 points. Two marks for correctly carrying out the limit and one mark for identifying the limiting exponential distribution.

   (c) 6 points. One mark each for (i) infinite or countable sequence of iid random variables (either in words or notation), (ii) two finite moments, (iii) correct statement of conclusion in terms of probability statement, (iv) trying to apply Chebychev's inequality in proof, (v) correctly applying Chebychev's inequality, and (vi) applying limiting argument to obtain the result.

4. (a) 2 marks. One mark for each definition.

   (b) 5 marks. One mark for correct integral form of $E(X^n)$; two marks for evaluating integral and final answer; one mark for variance formula in terms of first two moments; and one mark for final answer for variance.

   (c) 5 marks. One mark for setting up formula for method of moments; one mark for correct method of moments estimator; one mark for correct (log) likelihood function (constants may be omitted or included); one mark for correctly deriving the MLE; and one mark for second derivative test.

   (d) 4 marks. One mark for correct formula for mean square error (either its definition or in terms of variance and bias); one mark for expectation of MLE and/or arguing that the variance equals the mean square error; and two marks for deriving the variance / mean square error. Of the final two marks, one should be awarded for applying basic properties of variances (e.g., variance of the sum is the sum of the variances, squaring constants, and variance equals the expectation of the square minus the square of the expectation) and one mark should be awarded for substituting in the the identify from part (a) to arrive at the correct answer. (In principle the mean square error can be derived directly from its definition, without deriving the expectation and variance of the MLE. If this is done correctly, full marks should be given.)

   (e) 4 marks. One mark for the correctly using the definition of bias; one mark or basic manipulations of means and variance (as in part d); one mark for substituting in the the identify from part (a) to arrive at the correct form of the bias; one mark for noting that the estimator is biased, but asymptotically unbiased.