Imperial College London

MATH95011

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2020

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

Probability and Statistics 2

Date: 20th May 2020

Time: 09.00am – 11.00am (BST)

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 (ONE FOR EACH QUESTION) WITH COMPLETED COVERSHEETS WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. (a) If the independent random variables $Z_1, \ldots, Z_n \sim N(0,1)$, i.e.

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right), \qquad \mathbf{z} \in \mathbf{R}^n,$$

state without proof the distribution of the vector $\mathbf{U} = M\mathbf{Z}$, where M is an invertible linear transformation. (3 marks)

(b) Suppose $\mathbf{Y} = L\mathbf{Z}$, where the first row of the matrix L is $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ and $L^T L = LL^T = I_n$.

(i) Show that

$$\sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2 = \sum_{i=2}^{n} Y_i^2,$$

where $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

(ii) Determine the joint distribution of $\bar{Z}$ and $\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)^2$, stating clearly any results you use.

(5 marks)

Suppose $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ is a random sample from a Normal distribution.

(c) State with brief justification the joint distribution of

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

(4 marks)

(d) Explain how to construct a $(1-\alpha)$ central confidence interval for the parameter μ when the parameter σ^2 is unknown, stating any results needed. (5 marks)

(e) Comment briefly on the differences between intervals as constructed in (d) and $(1-\alpha)$ central confidence intervals for μ in the setting where σ^2 is known. (3 marks)

(Total: 20 marks)

2. Let $(\mathcal{S}, \mathcal{F}, \Pr)$ be a probability space. Let $X\colon \mathcal{S} \to \mathbf{R}$ be a random variable, and let $\mathcal{B}$ be the Borel sigma algebra on $\mathbf{R}$.

(a) Define what it means for $\mathcal{F}$ to be a sigma algebra on $\mathcal{S}$. (3 marks)

(b) Define what it means for the function $X\colon \mathcal{S} \to \mathbf{R}$ to be a random variable. (2 marks)

(c) Show that $\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}\}$ is a sigma algebra on $\mathcal{S}$. (5 marks)

(d) Explain why $\mathcal{F}_X$ is the smallest sigma algebra with respect to which X is a random variable. (2 marks)

(e) Suppose a coin is flipped twice, so that $\mathcal{S} = \{HH, HT, TH, TT\}$. Find $\mathcal{F}_Y$, where $Y : \mathcal{S} \to \mathbf{R}$ counts the number of heads. (5 marks)

(f) Determine $\mathcal{F}_Z$ for $Z = |Y - 1|$. (3 marks)

(Total: 20 marks)

3. The continuous random variable X has probability density function

$$f_X(x) = \begin{cases} \frac{\lambda}{(1+\lambda x)^2} & x > 0, \\ 0 & \text{otherwise}, \end{cases}$$

where $\lambda > 0$.

(a) Find the cumulative distribution function F_X . (2 marks)

(b) Explain in detail how to use a random sample $U_1, \ldots, U_n \sim \textsc{Uniform}(0,1)$ to produce a random sample $X_1, \ldots, X_n$ from the distribution of X. (4 marks)

(c) Find the cumulative distribution function F_{Y_n} of $Y_n = \min\{X_1, \ldots X_n\}$. (4 marks)

(d) Define what it means for a sequence of random variables to converge in distribution. (2 marks)

(e) Determine, for each y, the value of

$$F(y) = \lim_{n\to\infty} F_{Y_n}(y),$$

and state, with reasons, whether or not F is the cumulative distribution function of a random variable. (3 marks)

(f) State, giving brief reasons, whether Y_n converges i) in distribution, ii) in probability. (2 marks)

(g) Find a deterministic (i.e. non-random) sequence α_n and a non-degenerate random variable Y such that $\alpha_n Y_n \to Y$ as $n \to \infty$. (3 marks)

(Total: 20 marks)

4. Let $X_1, X_2, \dots X_n \sim IG(\mu, \lambda)$ be a random sample from the inverse Gaussian distribution with probability density function

$$f_X(x|\lambda,\mu) = \begin{cases} \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) & x > 0. \\ 0 & \text{otherwise,} \end{cases}$$

where λ and μ are positive parameters.

(a) Write down the likelihood function $L(\mu, \lambda)$. (2 marks)

(b) Show that the maximum likelihood estimator of μ is $\widehat{\mu} = \bar{X}$ and determine the maximum likelihood estimator $\widehat{\lambda}$ of λ. (8 marks)

(c) Show that if $Y = \frac{\lambda}{\mu^2}X$, $Y \sim IG(\mu_0, \mu_0^2)$, where the form of μ_0 is to be determined. (2 marks)

(d) Given that Y has moment generating function

$$M_Y(t) = \exp\left(\mu_0\left(1 - \sqrt{1-2t}\right)\right),$$

determine the mean square error of the estimator $\widehat{\mu}$ found in (b) in terms of μ and λ. (6 marks)

(e) Give an approximate sampling distribution for $\widehat{\mu}$ valid for large n, stating clearly any results that you use. (2 marks)

(Total: 20 marks)

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)	$\alpha<\beta\in\mathbb{R}$	$\frac{1}{\beta-\alpha}$	$\frac{x-\alpha}{\beta-\alpha}$	$\frac{(\alpha+\beta)}{2}$	$\frac{(\beta-\alpha)^2}{12}$	$\frac{e^{\beta t}-e^{\alpha t}}{t\,(\beta-\alpha)}$
$Exponential(\lambda)$ (stand. model $\lambda=1$)	$\mathbb{R}^+$	$\lambda\in\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)$
$Gamma(\alpha,\beta)$ (stand. model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$		$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(\frac{\beta}{\beta-t}\right)^\alpha$
$Weibull(\alpha,\beta)$ (stand. model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$	$1-e^{-\beta x^\alpha}$	$\frac{\Gamma\left(1+1/\alpha\right)}{\beta^{1/\alpha}}$	$\frac{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2}{\beta^{2/\alpha}}$	
$Normal(\mu,\sigma^2)$ (stand. model $\mu=0,\sigma=1$)	$\mathbb{R}$	$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\frac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu>1$)	$\frac{\nu}{\nu-2}$ (if $\nu>2$)	
$Pareto(\theta,\alpha)$	$\mathbb{R}^+$	$\theta,\alpha\in\mathbb{R}^+$	$\frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$	$1-\left(\frac{\theta}{\theta+x}\right)^\alpha$	$\frac{\theta}{\alpha-1}$ (if $\alpha>1$)	$\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$ (if $\alpha>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta\in\mathbb{R}^+$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

1. (a) [3 marks] *[Seen]* The distribution of $\mathbf{U}$ is multivariate normal with mean $M\mathbf{0} = 0$ and variance-covariance matrix MM^T.

(b) *[Seen]*

(i)
$$\sum_{i=1}^{n}(Z_i - \bar{Z})^2 = \sum_{i=1}^{n}\left(Z_i^2 - 2Z_i\bar{Z} + \bar{Z}^2\right) = \sum_{i=1}^{n} Z_i^2 - 2n\bar{Z}^2 + n\bar{Z}^2 = \sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2.$$

Now since $L^T L = I_n$, we have that

$$\sum_{i=1}^{n} Y_i^2 = Y^T Y = Z^T L^T L Z = Z^T Z = \sum_{i=1}^{n} Z_i^2,$$

and since the first row of L is $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$,

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \sqrt{n}\bar{Z},$$

hence

$$\sum_{i=1}^{n}(Z_i - \bar{Z})^2 = \sum_{i=2}^{n} Y_i^2.$$

(ii) The variance-covariance matrix of $Y_1, \ldots Y_n$ is $L^T L = I_n$, so that the Y_i variables are independent. Now, since $\bar{Z}$ is a function of Y_1 alone and $\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ is a function of $Y_2, \ldots, Y_n$ alone, these two variables are independent. Since Y_1 is normal with mean zero, $\bar{Z}$ is normal with mean zero and its variance is given by $\frac{1}{n}\mathrm{Var}(Y_1) = \frac{1}{n}$.
Each of $Y_2, \ldots Y_n$ is an independent standard normal variable, and so $Y_i^2 \sim \chi^2(1)$. Since the variables Y_i are independent, and independent chi-square variables add, we see that

$$\sum_{i=1}^{n}(Z_i - \bar{Z})^2 \sim \chi^2(n-1).$$

(c) *[Seen]* By defining $Z_i = \frac{X_i - \mu}{\sigma}$, we can use the answers from the previous part.
As $\bar{Z} \sim N(0, \frac{1}{n})$, because normal random variables respect linearity, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
As $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$, which is a gamma variable with shape parameter $\frac{n-1}{2}$ and scale parameter 2, it follows that S^2 is gamma distributed with shape parameter $\frac{n-1}{2}$ and scale parameter $\frac{2\sigma^2}{n-1}$. The two random variables $\bar{X}$ and S^2 are independent because the corresponding standardized variables are independent.

(d) *[Seen Method]* A random variable is said to follow the t-distribution with k degrees of freedom if it can be written as the ratio

$$T = \frac{Z}{\sqrt{Q/k}},$$

where $Z \sim N(0,1)$ and $Q \sim \chi^2(k)$ are independent. It follows from part (c) that

$$\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{S} \sim t_{n-1}.$$

Hence if t is such that

$$\int_t^\infty f_T(z)dz = \frac{\alpha}{2},$$

where f_T is the t_{n-1} density,

$$\Pr\left(-t \le \frac{\sqrt{n}\,(\bar{X}-\mu)}{S} \le t\right) = 1-\alpha.$$

Rearranging gives the central $1-\alpha$ confidence interval for μ:

$$\Pr\left(\bar{X} - \frac{tS}{\sqrt{n}} \le \mu \le \bar{X} + \frac{tS}{\sqrt{n}}\right) = 1-\alpha.$$

(e) *[Unseen]* Where σ^2 is known, use σ (rather than S) to standardize the sample mean, getting

$$\Pr\left(-z \le \frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \le z\right) = 1-\alpha,$$

where z is an appropriate quantile of the standard normal distribution. In this case, all intervals have length $\frac{2\sigma}{\sqrt{n}}$, whereas for the t-interval the length of the interval is a random quantity, differing from sample to sample. In the case where σ^2 is unknown, the average interval length should be longer than in the known σ^2 case.

2. (a) [3 marks] *[Seen]* $\mathcal{F}$ is a sigma algebra if

   * $\emptyset \in \mathcal{F}$
   * $A^c \in \mathcal{F}$ whenever $A \in \mathcal{F}$
   * $\cup_{i=1}^\infty A_i \in \mathcal{F}$ whenever $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$.

(b) [2 marks]*[Seen]* $X\colon \mathcal{S} \to \mathbf{R}$ is said to be a random variable if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$.

(c) [5 marks] *[Seen Method]* There are three properties to show.

   i. First, since $X^{-1}(\emptyset) = \emptyset$, clearly $\emptyset \in \mathcal{F}_X$.

   ii. Now suppose $A \in \mathcal{F}_X$, say $A = X^{-1}(B)$, for $B \in \mathcal{B}$. Since $\mathcal{B}$ is a sigma algebra, $\mathbf{R}\backslash B \in \mathcal{B}$, and $X^{-1}(\mathbf{R}\backslash B) = A^c$, so that $A^c \in \mathcal{F}_X$.

   iii. Suppose now that $A_1, A_2, \ldots \in \mathcal{F}_X$, say $A_i = X^{-1}(B_i)$, for $B_i \in \mathcal{B}$. Then $\bigcup_{i=1}^\infty B_i \in \mathcal{B}$, since $\mathcal{B}$ is a sigma algebra, so then $\bigcup_{i=1}^\infty A_i = X^{-1}(\bigcup_{i=1}^\infty B_i) \in \mathcal{F}_X$.

(d) [2 marks]*[Unseen]* For X to be a random variable with respect to a sigma algebra $\mathcal{G}$, we need at least that $X^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}$, so $\mathcal{F}_X \subseteq \mathcal{G}$ whenever X is a random variable, and hence $\mathcal{F}_X$ is the smallest sigma algebra making X a random variable.

(e) [5 marks] *[Seen Similar]* The three possible values of the function Y are 0, 1 and 2 so the pre-image of a Borel set B is determined by the combination of these elements that it contains.

$$Y^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B, 2 \notin B \\ \{HH\} & 0 \notin B, 1 \notin B, 2 \in B \\ \{HT, TH\} & 0 \notin B, 1 \in B, 2 \notin B \\ \{TT\} & 0 \in B, 1 \notin B, 2 \notin B \\ \{HH, HT, TH\} & 0 \notin B, 1 \in B, 2 \in B \\ \{HH, TT\} & 0 \in B, 1 \notin B, 2 \in B \\ \{HT, TH, TT\} & 0 \in B, 1 \in B, 2 \notin B \\ \mathcal{S} & 0 \in B, 1 \in B, 2 \in B. \end{cases}$$

The required sigma algebra is then given by the collection of eight sets on the right.

(f) [3 marks] *[Unseen]* The two possible values of Z are 0 and 1, so we obtain the smaller sigma algebra

$$\{\emptyset, \{HT, TH\}, \{HH, TT\}, \mathcal{S}\}.$$

3. (a) [2 marks] *[Seen Method]*

$$F_X(x) = \int_0^x f_X(t)\ dt = \int_0^x \frac{\lambda}{(1+\lambda t)^2} dt = 1 - \frac{1}{1+\lambda x} = \frac{\lambda x}{1+\lambda x}$$

(b) [4 marks]*[Seen Similar]* Let $Y = F_X^{-1}(U)$.
Then since F_X is an increasing, and therefore 1-1, function, whose range is $[0,1)$,

$$\Pr(Y \leq y) = \Pr(F_X^{-1}(U) \leq y) = \Pr(U \leq F_X(y)) = F_X(y).$$

Hence Y has the same cumulative distribution function as X.
Explicitly, if
$U = 1 - \frac{1}{1+\lambda X}$, then
$1 + \lambda X = \frac{1}{1-U}$, so

$$X = \frac{U}{\lambda(1-U)}$$

gives a suitable transformation.

(c) *[Seen Method]* [4 marks]

$$\Pr(Y_n > y) = \Pr(X_1 > y \cap \ldots \cap X_n > y) = \frac{1}{(1+\lambda y)^n},$$

so

$$F_Y(y) = 1 - \frac{1}{(1+\lambda y)^n}, \qquad y > 0.$$

(d) [2 marks] *[Seen]* A sequence $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points of continuity of F_X.

(e) [2 marks] *[Unseen]*

$$F_{Y_n}(y) = 1 - \frac{1}{(1+\lambda y)^n} \to \begin{cases} 0 & y \leq 0 \\ 1 & y > 0 \end{cases} \quad .$$

Since the limit function is not right continuous at zero, it is not the cdf of a random variable.

(f) [2 marks] *[Seen Similar]* Y_n converges in distribution to the constant random variable $Y = 0$, since the sequence of cdfs converges for all values apart from $y = 0$, and the cdf of the constant random variable is not continuous at this exceptional point.

Since Y_n converges in distribution to a constant, by a result from lectures, it also converges in probability to this constant.

(g) [3 marks] *[Seen Similar]* Define $\alpha_n = n$ and $V_n = \alpha_n Y_n$, then

$$\Pr(V_n \leq v) = \Pr\left(Y_n \leq \frac{v}{n}\right) = 1 - \frac{1}{\left(1 + \frac{\lambda v}{n}\right)^n} \to 1 - \exp(-\lambda v).$$

Hence $nY_n \to Y$, where Y has an exponential distribution with rate λ.

4. (a) [2 marks] *[Seen Method]*

$$f_X(x|\lambda,\mu) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right).$$

So for the random sample $x_1, \ldots x_n$,

$$\begin{aligned} L(\lambda,\mu) &= \prod_{i=1}^{n} \left(\frac{\lambda}{2\pi x_i^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x_i-\mu)^2}{2\mu^2 x_i}\right) \\ &= \left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\prod_{i=1}^{n} x_i}\right)^{\frac{3}{2}} \exp\left(-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}\right). \end{aligned}$$

(b) [8 marks] *[Seen Method]*

Compute the log-likelihood

$$l(\lambda,\mu) = -\frac{n}{2}\log 2\pi + \frac{n}{2}\log\lambda - \frac{3}{2}\sum_{i=1}^{n}\log x_i - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}.$$

Determine gradient of log-likelihood

$$\frac{\partial l}{\partial \lambda} = \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}.$$

$$\frac{\partial l}{\partial \mu} = -\frac{\lambda}{2}\frac{\partial}{\partial \mu}\sum_{i=1}^{n}\frac{1}{x_i}\left(\frac{x_i}{\mu}-1\right)^2 = \frac{\lambda}{2}\sum_{i=1}^{n}2\left(\frac{x_i}{\mu}-1\right)\frac{1}{\mu^2} = \frac{\lambda}{\mu^3}\sum_{i=1}^{n}(x_i-\mu).$$

From the second equation, we immediately see that (as $\lambda \neq 0$) if $\frac{\partial l}{\partial \mu} = 0$ we must have $\sum_{i=1}^{n} x_i = n\mu$, i.e. $\widehat{\mu} = \bar{X}$ as required.

From the first equation, if $\frac{\partial l}{\partial \lambda} = 0 = \frac{\partial l}{\partial \mu}$, we have

$$\frac{n}{\lambda} = \frac{1}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i},$$

so that

$$\lambda = \frac{n\mu^2}{\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}} = \frac{n\mu^2}{\sum_{i=1}^{n}\frac{x_i^2-2\mu x_i+x_i^2}{x_i}} = \frac{n\mu^2}{\sum_{i=1}^{n} x_i - 2\mu + \frac{\mu^2}{x_i^2}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i} - \frac{1}{\bar{x}}},$$

where the last equality follows on substituting $\bar{x}$ for μ.

Hence the candidate maximum likelihood estimators are

$\widehat{\mu} = \bar{X}, \qquad \widehat{\lambda} = \frac{n}{\sum_{i=1}^{n}\frac{1}{X_i} - \frac{1}{\bar{X}}}$.

To see that the stationary points we have found are indeed maxima, consider the second partial derivatives:

$$\frac{\partial^2 l}{\partial \mu^2} = \lambda\frac{\partial}{\partial \mu}\sum_{i=1}^{n}\left(\frac{x_i}{\mu^3} - \frac{1}{\mu^2}\right) = \lambda\sum_{i=1}^{n}\left(-\frac{3x_i}{\mu^4} + \frac{2}{\mu^3}\right) = -\frac{\lambda}{\mu^4}\sum_{i=1}^{n}(3x_i - 2\mu)$$

Evaluating at $\mu = \bar{x}$ gives

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{\lambda}{\bar{x}^4}(3n\bar{x} - 2n\bar{x}) = -\frac{n\lambda}{\bar{x}^3} < 0.$$

For λ,

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{n}{2\lambda^2} < 0$$

and the mixed partial derivative is

$$\frac{\partial^2 l}{\partial \mu \partial \lambda} = \frac{1}{\mu^3} \sum_{i=1}^{n} (x_i - \mu),$$

which evaluates to zero when $\mu = \bar{x}$.

When the the hessian matrix is evaluated at the candidate MLE, it has negative trace and positive determinant, and so must be negative definite. Hence our candidates MLEs are indeed maxima.

(c) [2 marks] *[Unseen]*

$$f_Y(y) = f_X\left(\frac{\mu^2 y}{\lambda}\right)\frac{\mu^2}{\lambda} = \left(\frac{\lambda}{2\pi\left(\frac{\mu^2 y}{\lambda}\right)^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda\left(\left(\frac{\mu^2 y}{\lambda}\right) - \mu\right)^2}{2\mu^2\left(\frac{\mu^2 y}{\lambda}\right)}\right)\frac{\mu^2}{\lambda}$$
$$= \frac{\lambda}{\mu}\frac{1}{\sqrt{2\pi y^3}} \exp\left(-\frac{\left(y - \frac{\lambda}{\mu}\right)^2}{2y}\right),$$

which is of the required form with $\mu_0 = \frac{\lambda}{\mu}$.

(d) [6 marks] *[Unseen]*

We are given

$$M_Y(t) = \exp\left(\mu_0\left(1 - \sqrt{1-2t}\right)\right).$$

First we expand the the term inside the exponential to second order in t as

$$1 - \sqrt{1-2t} = 1 - \left(1 - t + \frac{(4t^2)}{2} \times \frac{1}{2} \times \frac{-1}{2} + o(t^2)\right) = t + \frac{t^2}{2} + o(t^2).$$

(Could read off mean and variance directly from this cumulant generating function, but this was not covered in M2S1.)

Keeping terms to second order in t we have

$$M_Y(t) = \exp\left(\mu_0\left(1 - \sqrt{1-2t}\right)\right) = \exp\left(\mu_0 t + \frac{\mu_0 t^2}{2} + o(t^2)\right)$$
$$= \exp(\mu_0 t)\exp\left(\frac{\mu_0 t^2}{2} + o(t^2)\right) = \left(1 + \mu_0 t + \frac{\mu_0^2 t^2}{2} + o(t^2)\right)\left(1 + \frac{\mu_0^2 t^2}{2} + o(t^2)\right)$$
$$= 1 + \mu_0 t + (\mu_0 + \mu_0^2)\frac{t^2}{2} + o(t^2).$$

So then $\mathrm{E}(Y) = \mu_0$ and $\mathrm{E}(Y^2) = \mu_0 + \mu_0^2$, giving $\mathrm{Var}(Y) = \mu_0$.
Hence

$$\mathrm{E}(\bar{Y}) = \mu_0, \qquad \mathrm{Var}(\bar{Y}) = \frac{\mu_0}{n}.$$

This then gives

$$\mathrm{E}(\bar{X}) = \frac{\mu^2}{\lambda}\mu_0 = \frac{\mu^2}{\lambda}\frac{\lambda}{\mu} = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\mu^4}{\lambda^2}\frac{\mu_0}{n} = \frac{\mu^4}{\lambda^2}\frac{\lambda}{n\mu} = \frac{\mu^3}{n\lambda}.$$

Hence we see $\widehat{\mu} = \bar{X}$ is unbiased for μ, and

$$MSE(\mu) = \mathsf{bias}(\mu)^2 + \mathrm{Var}(\mu) = \frac{\mu^3}{n\lambda}.$$

(e) [2 marks] *[Seen Similar]* By the central limit theorem, for large samples $\bar{X} \sim N(\mu, \frac{\mu^3}{\lambda n})$, approximately.