Imperial College London

MATH70043

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

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

Statistical Theory

Date: 24 May 2023

Time: 14:00 – 16:30 (BST)

Time Allowed: 2.5hrs

This paper has 5 Questions.

Please Answer All Questions in 1 Answer Booklet

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

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

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

1. (a) State the Rao-Blackwell theorem. (4 marks)

Let $X_1, \dots, X_n$ be independent and identically distributed random variables with distribution

$$P(X_1 = k) = p^{k-1}(1-p), \qquad k = 1, 2, 3, \dots,$$

where $p \in (0,1)$ is an unknown parameter to be estimated.

[Throughout this question, you may use without proof that $\sum_{i=1}^n X_i$ has negative binomial distribution with parameters n and p, that is

$$P\left(\sum_{i=1}^n X_i = r\right) = \binom{r-1}{n-1}(1-p)^n p^{r-n}, \qquad r = n, n+1, n+2, \dots .]$$

(b) Determine a one-dimensional sufficient statistic $T = T(X_1, \dots, X_n)$ for p. Justify your answer. (3 marks)

(c) Show that

$$\tilde{p}_n = 1\{X_1 \neq 1\}$$

is an unbiased estimator of p. Hence or otherwise, find an unbiased estimator $\hat{p}_n$ for p that is a function of your sufficient statistic T. (8 marks)

(d) Is your estimator $\hat{p}_n$ in (c) the minimum variance unbiased estimator of p?

Without computing the Cramer-Rao lower bound for p, state whether your estimator $\hat{p}_n$ achieves the Cramer-Rao bound or not.

Justify all your answers.

[You may assume that if a power series $h(t) = \sum_{k=0}^\infty a_k t^k$ is identically zero on a non-empty open set in $\mathbb{R}$, then $a_k = 0$ for all $k \geq 0$.] (5 marks)

(Total: 20 marks)

2. A researcher wants to estimate the proportion $\theta \in (0,1)$ of presence of a gene in a population of animals. Every animal receives one chromosome from each of its two parents, each of which carries gene A with probability $1-\theta$ and gene B with probability θ, independently. The researcher has a sample of n animals, and observes whether each animal has two copies of gene A (denoted AA), two copies of gene B (denoted BB) or one of each (denoted AB). Let n_{AA}, n_{BB} and n_{AB} denote the number of each observation among the n animals.

(a) Give the probability of each observation as a function of θ, which you may denote by $f_\theta(x)$, for all three values $x = AA$, BB or AB. (6 marks)

(b) For a vector $v = (v_{AA}, v_{BB}, v_{AB})$, define the estimator

$$\bar{\theta}_v = v_{AA}\frac{n_{AA}}{n} + v_{BB}\frac{n_{BB}}{n} + v_{AB}\frac{n_{AB}}{n}.$$

Find the unique vector v^* such that $\bar{\theta}_{v^*}$ is an unbiased estimator for θ.

Show that $\bar{\theta}_{v^*}$ is consistent.

[You may find it helpful to write $n_{AA} = \sum_{i=1}^n 1\{X_i = AA\}$, where X_i denotes observation i, with similar expressions for n_{BB} and n_{AB}.] (7 marks)

(c) Compute the maximum likelihood estimator $\hat{\theta}_n$ for θ in this model. Find the limit distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ as $n \to \infty$.

[You may use any results from the course as long as these are clearly mentioned.]

(7 marks)

(Total: 20 marks)

3. Let $X \sim \text{Bin}(n, \theta)$ be a binomial random variable, where $n \geq 1$ is a known integer and $\theta \in (0, 1)$ is a parameter to be estimated.

Recall that for positive integers a, b, the Beta(a, b) distribution on $(0, 1)$ has density function

$$f_{a,b}(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}.$$

(a) Compute the maximum likelihood estimator for θ. Consider now a Bayesian model where we assign $\theta \sim U(0, 1)$ a uniform prior on $(0, 1)$. Show that the posterior for θ is a Beta(a, b) distribution for parameters a, b you should specify. (6 marks)

(b) For a loss function L, define the *risk function* of an estimator $\hat{\theta}$ of θ. For a prior π for θ, define the *Bayes risk* and *posterior risk* based on an observation X. (3 marks)

(c) For the loss function

$$L(\hat{\theta}, \theta) = \frac{(\hat{\theta} - \theta)^2}{\theta(1-\theta)},$$

find an optimal Bayes estimator (also called Bayes decision rule) for the uniform $U(0, 1)$ prior. Evaluate its risk function as a function of θ. (8 marks)

(d) Give a minimax optimal estimator for the loss function L. Justify your answer. (3 marks)

(Total: 20 marks)

4. (a) State the weak law of large numbers and the central limit theorem. (2 marks)

Suppose that $X_1, \ldots, X_n \sim^{iid}$ Poi(λ) are independent and identically distribution Poisson random variables with parameter $\lambda > 0$.

(b) Show that the maximum likelihood estimator $\hat{\lambda}_n$ of λ equals $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Show that $\hat{\lambda}_n$ is consistent and derive the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda)$ as $n \to \infty$.

Construct an asymptotic 95%-confidence interval for λ. Briefly justify your construction.

[You may use without proof that $P(Z > 1.96) = 0.025$ for $Z \sim N(0,1)$.] (7 marks)

(c) It is now required to test

$$H_0 : \lambda = 1 \qquad \text{against} \qquad H_1 : \lambda \neq 1.$$

Show that the construction of the likelihood ratio test leads to the statistic

$$2 \log \Lambda_n(x) = 2n \left[\hat{\lambda}_n \log \hat{\lambda}_n + 1 - \hat{\lambda}_n \right].$$

Clearly mentioning any results which you use, for large n, what approximately is the distribution of $2 \log \Lambda_n$ under H_0? (5 marks)

(d) Writing $\hat{\lambda}_n = 1 + Z_n$ and assuming that Z_n is small, show that

$$2 \log \Lambda_n(x) \approx n Z_n^2.$$

Using this result and the central limit theorem, justify the approximate distribution of $2 \log \Lambda_n$ given in (c). What could you say if the assumption that Z_n is small failed to hold?

[You may use without proof that $\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \ldots$ for any $|z| < 1$.] (6 marks)

(Total: 20 marks)

5. Let $X_1, \dots, X_n$ be independent and identically distributed observations and suppose that $T_n = T_n(X_1, \dots, X_n)$ is an estimator of a parameter θ. Recall that the jackknife bias corrected estimator is $\hat{T}_J = T_n - b_J$, where

$$b_J = (n-1)\left(\frac{1}{n}\sum_{i=1}^{n} T_{(-i)} - T_n\right)$$

and $T_{(-i)}$ denotes the estimator based on $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, i.e. with the i^{th} observation removed.

(a) Suppose that the bias function $b_\theta(T_n) = E_\theta[T_n] - \theta$ of T_n satisfies the expansion

$$b_\theta(T_n) = \frac{a}{n} + \frac{b}{n^2} + O\left(\frac{1}{n^3}\right)$$

as $n \to \infty$ for some $a, b \in \mathbb{R}$. Show that

$$b_\theta(\hat{T}_J) = E_\theta[\hat{T}_J] - \theta = O\left(\frac{1}{n^2}\right)$$

as $n \to \infty$.

(4 marks)

Suppose now that $X_1, \dots, X_n \sim^{iid} N(\mu, 1)$, $\mu \in \mathbb{R}$, and consider the parameter $\theta = \mu^2$.

(b) Consider the estimator $T_n = (\bar{X}_n)^2$ for θ, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean. Compute the exact bias of both T_n and $\hat{T}_J$. (5 marks)

(c) What is the asymptotic distribution of $\sqrt{n}(T_n - \mu^2)$ as $n \to \infty$? (3 marks)

(d) Show that in this specific case,

$$b_J = \frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

[Hint: let $\bar{X}_{(-i)} = \frac{1}{n-1}\sum_{j \neq i} X_j$ denote the sample mean of the observations with X_i removed. Note that

$$b_J = \frac{n-1}{n}\sum_{i=1}^{n}[(\bar{X}_{(-i)})^2 - (\bar{X}_n)^2]$$

is the difference of two squares and use the identities

$$\sum_{i=1}^{n}(\bar{X}_{(-i)} - \bar{X}_n) = 0 \qquad \textit{and} \qquad \bar{X}_{(-i)} - \bar{X}_n = \frac{1}{n-1}(\bar{X}_n - X_i) \qquad .]$$

(5 marks)

(e) Deduce that $\sqrt{n}(\hat{T}_J - \mu^2)$ has the same asymptotic distribution as $\sqrt{n}(T_n - \mu^2)$.
[Recall that $\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \chi^2_{n-1}$, and that the χ^2_k distribution has mean k and variance $2k$, respectively.] (3 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2023

This paper is also taken for the relevant examination for the Associateship.

MATH60043/MATH70043

Statistical Theory (Solutions)

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

1. (a) seen ⇓

Theorem 1. *Let $T = T(X)$ be a sufficient statistic for θ and let $\tilde{\theta}(X)$ be an estimator for θ with $E_\theta(\tilde{\theta}^2) < \infty$ for all $\theta \in \Theta$. Let $\hat{\theta}(X) = E[\tilde{\theta}(X)|T(X)]$. Then for all $\theta \in \Theta$,*

$$Bias_\theta(\hat{\theta}) = Bias_\theta(\tilde{\theta}) \qquad and \qquad Var_\theta(\hat{\theta}) \leq Var_\theta(\tilde{\theta}),$$

with strict inequality for some $\theta \in \Theta$ unless $\tilde{\theta}$ is a function of T. 4, A

(b) sim. seen ⇓

$$f_p(x) = \prod_{i=1}^n p^{x_i - 1}(1-p) = p^{\sum_{i=1}^n x_i} p^{-n}(1-p)^n,$$

so by the factorization criterion, $T(X) = \sum_{i=1}^n X_i$ is sufficient for p. 3, A

(c) sim. seen ⇓

$$E_p \tilde{p}_n = P_p(X_1 \neq 1) = 1 - P_p(X_1 = 1) = 1 - (1-p) = p,$$

so $\tilde{p}_n$ is unbiased for p.

2, A

We use the Rao-Blackwell estimator based on $\tilde{p}_n$. For $t \geq n$, meth seen ⇓

$$\begin{aligned}
\hat{p}_n = E_p[\tilde{p}_n|T = t] &= P(X_1 \neq 1|T = t) \\
&= 1 - P(X_1 = 1|T = t) \\
&= 1 - \frac{P(X_1 = 1, T = t)}{P(T = t)} \\
&= 1 - \frac{P(X_1 = 1, \sum_{i=2}^n X_i = t - 1)}{P(T = t)} \\
&= 1 - \frac{P(X_1 = 1)P(\sum_{i=2}^n X_i = t - 1)}{P(T = t)}.
\end{aligned}$$

Substituting in the formula for the pmf

$$\begin{aligned}
\hat{p}_n &= 1 - \frac{(1-p) \times \binom{t-2}{n-2}(1-p)^{n-1}p^{t-1-(n-1)}}{\binom{t-1}{n-1}(1-p)^n p^{t-n}} \\
&= 1 - \frac{\binom{t-2}{n-2}}{\binom{t-1}{n-1}} \\
&= 1 - \frac{(t-2)!}{(t-n)!(n-2)!}\frac{(t-n)!(n-1)!}{(t-1)!} \\
&= 1 - \frac{n-1}{t-1}.
\end{aligned}$$

We thus have $\hat{p}_n = 1 - \frac{n-1}{T(X)-1}$, which is unbiased by the Rao-Blackwell theorem.

6, B

(d) Yes, $\hat{p}_n$ is the uniform minimum variance unbiased estimator (UMVUE) for p. T turns out to be a complete statistic for p: let g be any measurable function such that $E_p g(T) = 0$ for all $p \in (0,1)$. Then

sim. seen ⇓

$$E_p g(T) = \sum_{r=n}^{\infty} g(r)\binom{r-1}{n-1}(1-p)^n p^{r-n} = (1-p)^n p^{-n} \sum_{r=n}^{\infty} g(r)\binom{r-1}{n-1} p^r = 0.$$

But the sum is a power series in p which is zero on a non-empty open set of $\mathbb{R}$ and hence all its coefficients are zero, i.e. $g(r)\binom{r-1}{n-1} = 0$ for $r = n, n+1, \ldots$. Thus $g(r) = 0$ for $r = n, n+1, \ldots$ and thus $P_p(g(T) = 0) = 1$ for all $p \in (0,1)$, i.e. T is complete for p. Hence by the Lehmann-Scheffe theorem, any unbiased estimator that is a function of a complete statistic is UMVUE, which is the case here. 3, C

The joint pmf of $X_1, \ldots, X_n$ can be written in exponential family form as

$$f_p(x) = (1-p)^n p^{t-n} = \exp\{(\log p)t + n\log(1-p) - n\log p\}.$$

We know that an estimator $\hat{p}_n = 1 - \frac{n-1}{t-1}$ attains the Cramer-Rao lower bound if and only if $(X_1, \ldots, X_n)$ comes from an exponential family with natural statistic $\hat{p}_n$. But the above exponential family cannot be written in this way since it has natural statistic t, hence $\hat{p}_n$ does *not* attain the CRLB. 2, D

2. (a) Since each chromosome is inherited independently, unseen ⇓

$$f_\theta(AA) = (1-\theta)^2, \quad f_\theta(BB) = \theta^2, \quad f_\theta(AB) = 2\theta(1-\theta).$$

6, A

(b) We have $E_\theta n_{AA} = n f_\theta(AA) = n(1-\theta)^2$ and similarly for n_{BB} and n_{AB}. Hence we require unseen ⇓

$$\begin{aligned} E_\theta \bar{\theta}_v &= v_{AA}(1-\theta)^2 + v_{BB}\theta^2 + v_{AB}2\theta(1-\theta) \\ &= [v_{BB} - 2v_{AB} + v_{AA}]\theta^2 + [2v_{AB} - 2v_{AA}]\theta + v_{AA} = \theta \end{aligned}$$

for our estimator to be unbiased. Since must hold for all $\theta \in (0,1)$, we can equate coefficients. This implies $v_{AA} = 0$ and hence $2v_{AB} = 1$, i.e. $v_{AB} = 1/2$. Lastly, $v_{BB} = 2v_{AB} = 1$, so that

$$v^* = (0, 1, 1/2).$$

4, B

For consistency, by the weak law of large numbers,

$$\frac{n_x}{n} = \frac{1}{n}\sum_{i=1}^n 1\{X_i = x\} \to^p E_\theta 1\{X_i = x\} = f_\theta(x)$$

for $x \in \{AA, BB, AB\}$. Therefore,

$$\bar{\theta}_{v^*} \to^p f_\theta(BB) + \frac{1}{2} f_\theta(AB) = \theta^2 + \frac{1}{2}2\theta(1-\theta) = \theta,$$

i.e. $\bar{\theta}_{v^*}$ is consistent. 3, B

(c) The likelihood is unseen ⇓

$$\begin{aligned} \prod_{i=1}^n f_\theta(X_i) &= \prod_{i=1}^n [(1-\theta)^2]^{1\{X_i=AA\}} [\theta^2]^{1\{X_i=BB\}} [2\theta(1-\theta)]^{1\{X_i=AB\}} \\ &= (1-\theta)^{2n_{AA}} \theta^{2n_{BB}} [2\theta(1-\theta)]^{n_{AB}} \\ &= 2^{n_{AB}} \theta^{2n_{BB}+n_{AB}} (1-\theta)^{2n_{AA}+n_{AB}}. \end{aligned}$$

Taking logarithms and differentiating,

$$\begin{aligned} \ell_n(\theta) &= n_{AB}\log 2 + (2n_{BB} + n_{AB})\log\theta + (2n_{AA} + n_{AB})\log(1-\theta) \\ \ell_n'(\theta) &= \frac{2n_{BB} + n_{AB}}{\theta} - \frac{2n_{AA} + n_{AB}}{1-\theta} \\ \ell_n''(\theta) &= -\frac{2n_{BB} + n_{AB}}{\theta^2} - \frac{2n_{AA} + n_{AB}}{(1-\theta)^2} < 0. \end{aligned}$$

Hence the MLE is obtained by solving $\ell_n'(\theta) = 0$, which is equivalent to

$$(1-\theta)(2n_{BB} + n_{AB}) = \theta(2n_{AA} + n_{AB}).$$

This is solved by

$$\hat{\theta}_n = \frac{2n_{BB} + n_{AB}}{2[n_{AA} + n_{BB} + n_{AB}]}.$$

4, D

Since the regularity conditions are met for this statistical model, we may use asymptotic normality of the MLE to deduce that $\sqrt{n}(\hat{\theta}_n - \theta) \to^d N(0, I(\theta)^{-1})$, where $I(\theta)$ is the Fisher information. But

$$\begin{aligned} I(\theta) = -E_\theta[\ell_1''(\theta)] &= \frac{2f_\theta(BB) + f_\theta(AB)}{\theta^2} + \frac{2f_\theta(AA) + f_\theta(AB)}{(1-\theta)^2} \\ &= \frac{2\theta^2 + 2\theta(1-\theta)}{\theta^2} + \frac{2(1-\theta)^2 + 2\theta(1-\theta)}{(1-\theta)^2} \\ &= \frac{2}{\theta} + \frac{2}{1-\theta} \\ &= \frac{2}{\theta(1-\theta)}. \end{aligned}$$

Hence

$$\sqrt{n}(\hat{\theta}_n - \theta) \to^d N\left(0, \tfrac{\theta(1-\theta)}{2}\right).$$

$\boxed{3, \text{B}}$

3. (a) The log-likelihood equals $\boxed{\text{seen} \Downarrow}$

$$\ell_n(\theta) = \log\left\{\binom{n}{x}\theta^x(1-\theta)^{n-x}\right\} = \log\binom{n}{x} + x\log\theta + (n-x)\log(1-\theta).$$

Differentiating twice,

$$\begin{aligned}
\ell_n'(\theta) &= \frac{x}{\theta} - \frac{n-x}{1-\theta} \\
\ell_n''(\theta) &= -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} < 0.
\end{aligned}$$

Hence the global maximum is attained at the stationary point where $\ell_n'(\theta) = 0$, which after rearranging gives the MLE $\hat{\theta}_n = x/n$. $\boxed{3, \text{A}}$

The posterior is given by

$$\pi(\theta|x) \propto f_\theta(x)\pi(\theta) \propto \theta^x(1-\theta)^{n-x},$$

which we recognize as the form of a Beta$(x+1, n-x+1)$ distribution. $\boxed{3, \text{A}}$

(b) The *risk function* of an estimator $\hat{\theta}$ is the expected loss L under P_θ as a function of θ: $R(\hat{\theta}, \theta) = E_\theta[L(\hat{\theta}(X), \theta)]$. $\boxed{\text{seen} \Downarrow}$

The *Bayes risk* is $R_\pi(\hat{\theta}) = E_{\theta\sim\pi}[R(\hat{\theta}, \theta)]$, where the expectation is taken over the prior π.

The *posterior risk* given an observation $X = x$ is the average loss under the posterior distribution given X:

$$R_\pi(\hat{\theta}(x)) = E_\pi[L(\hat{\theta}(x), \theta)|x] = \int_\Theta L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta.$$

$\boxed{3, \text{A}}$

(c) We minimize the Bayes risk by finding a minimizer to the posterior risk. For $X = x$, this is $\boxed{\text{meth seen} \Downarrow}$

$$\begin{aligned}
R_\pi(\hat{\theta}(x)) = E_\pi[L(\hat{\theta}(x), \theta)|x] &= \int_0^1 L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta \\
&= \frac{(n+1)!}{x!(n-x)!}\int_0^1 \frac{(\hat{\theta}(x)-\theta)^2}{\theta(1-\theta)}\theta^x(1-\theta)^{n-x}d\theta \\
&= \frac{(n+1)!}{x!(n-x)!}\int_0^1 (\hat{\theta}(x)-\theta)^2\theta^{x-1}(1-\theta)^{n-x-1}d\theta.
\end{aligned}$$

This is a quadratic in $\hat{\theta}(x)$ and so is minimized at its stationary point. Differentiating the last expression with respect to $\hat{\theta}(x)$,

$$\begin{aligned}
\frac{dR_\pi(\hat{\theta}(x))}{d\hat{\theta}(x)} &= \frac{(n+1)!}{x!(n-x)!}\int_0^1 2(\hat{\theta}(x)-\theta)\theta^{x-1}(1-\theta)^{n-x-1}d\theta \\
&= \frac{2(n+1)!}{x!(n-x)!}\left[\hat{\theta}(x)\frac{(x-1)!(n-x-1)!}{(n-1)!} - \frac{x!(n-x-1)!}{n!}\right] \\
&= \frac{2n(n+1)}{x(n-x)}\left[\hat{\theta}(x) - \frac{x}{n}\right] = 0,
\end{aligned}$$

where in the second equality we can evaluate the integral by using the normalizing constant of the Beta distribution given in the question. Solving this gives that $\hat{\theta}(x) = x/n$ minimizes the posterior risk. By a result in the notes, this then minimizes the Bayes risk and is thus an optimal Bayes estimator.

6, D

Its risk function equals

meth seen ⇓

$$R(\hat{\theta}, \theta) = E_\theta \frac{(X/n - \theta)^2}{\theta(1-\theta)} = \frac{\mathsf{Var}_\theta(X)}{n^2\theta(1-\theta)} = \frac{n\theta(1-\theta)}{n^2\theta(1-\theta)} = \frac{1}{n}.$$

using the variance of the binomial distribution.

2, C

(d) A Bayes estimator having constant risk function is a minimax estimator by a result proved in the lectures. Thus the estimator $\hat{\theta}(X) = X/n$ above is minimax optimal.

unseen ⇓

3, D

4. (a) Weak law of large numbers: if $X_1, \dots, X_n$ are i.i.d. random variables with finite mean, then $\frac{1}{n}\sum_{i=1}^n X_i \to^P EX_1$ as $n \to \infty$. $\boxed{\text{seen} \Downarrow}$

Central limit theorem: if $X_1, \dots, X_n$ are i.i.d. random variables with finite mean and variance, then $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n X_i - EX_1) \to^d N(0, \mathrm{Var}(X_1))$ as $n \to \infty$. $\boxed{2, \text{A}}$

(b) The log-likelihood equals $\boxed{\text{seen} \Downarrow}$

$$\ell_n(\lambda) = \log\left(\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right) = \log\left(\frac{\lambda^{\sum x_i}}{\prod_i x_i!}e^{-n\lambda}\right) = (\log\lambda)\sum_{i=1}^n x_i - n\lambda - \log\left(\prod_{i=1}^n x_i!\right).$$

Differentiating twice,

$$\ell_n'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^n x_i - n$$

$$\ell_n''(\lambda) = -\frac{1}{\lambda^2}\sum_{i=1}^n x_i < 0$$

(note the MLE does not exist if $\sum_i x_i = 0$, but it is not necessary to state this). Thus the global maximum is attained when $\ell_n'(\lambda) = 0$, i.e. $\hat\lambda_n = \frac{1}{n}\sum_{i=1}^n X_i = \bar X_n$.

By the weak law of large numbers $\hat\lambda_n = \bar X_n \to^P E_\lambda X_1 = \lambda$, i.e. the MLE is consistent.

By the central limit theorem, $\sqrt{n}(\hat\lambda_n - \lambda) = \sqrt{n}(\bar X_n - E_\lambda X_1) \to^d N(0, \mathrm{Var}_\lambda(X_1)) = N(0, \lambda)$.

[Also acceptable to use consistency and asymptotic normality of MLE provided these are stated].

$\boxed{4, \text{A}}$

The limiting distribution implies $\boxed{\text{seen/sim.seen} \Downarrow}$

$$P_\lambda\left(-1.96\sqrt{\lambda} \le \sqrt{n}(\hat\lambda_n - \lambda) \le 1.96\sqrt{\lambda}\right) \to 0.95$$

as $n \to \infty$. We estimate the standard deviation by the plug-in estimate $\sqrt{\hat\lambda_n}$, since the estimation error for this is of smaller order than the main term, i.e. $\sqrt{\hat\lambda_n} = \sqrt{\lambda} + O(n^{-1/2})$ [any reasonable justification for this is sufficient]. Inverting the inequalities in the probability yields

$$P_\lambda\left(\hat\lambda_n - \frac{1.96}{\sqrt{n}}\sqrt{\hat\lambda_n} \le \lambda \le \hat\lambda_n + \frac{1.96}{\sqrt{n}}\sqrt{\hat\lambda_n}\right) \to 0.95,$$

giving our asymptotic confidence interval $[\hat\lambda_n - \frac{1.96}{\sqrt{n}}\sqrt{\hat\lambda_n}, \hat\lambda_n + \frac{1.96}{\sqrt{n}}\sqrt{\hat\lambda_n}]$. $\boxed{3, \text{B}}$

(c) The likelihood ratio test statistic is $\boxed{\text{unseen} \Downarrow}$

$$\Lambda_n(x) = \frac{\sup_{\lambda>0} f_\lambda(x)}{f_1(x)} = \frac{f_{\hat{\lambda}_n}(x)}{f_1(x)}.$$

Using the form of the log-likelihood derived in (b),

$$\begin{aligned} 2\log\Lambda_n(x) = 2\ell_n(\hat{\lambda}_n) - 2\ell_n(1) &= 2(\log\hat{\lambda}_n - \log 1)\sum_{i=1}^n X_i - 2n(\hat{\lambda}_n - 1) \\ &= 2n\hat{\lambda}_n\log\hat{\lambda}_n + 2n(1-\hat{\lambda}_n) \end{aligned}$$

since $\hat{\lambda}_n = \bar{X}_n$. Using Wilk's theorem, we have $2\log\Lambda_n(X) \to^d \chi^2_1$ as $n \to \infty$ under H_0. $\boxed{3, \text{B}}$

(d) Using the Taylor expansion provided in the question $\boxed{2, \text{C}}$ $\boxed{\text{unseen} \Downarrow}$

$$\begin{aligned} 2\log\Lambda_n(x) &= 2n[(1+Z_n)\log(1+Z_n) + 1 - 1 - Z_n] \\ &= 2n\left[(1+Z_n)\left(Z_n - \frac{Z_n^2}{2} + O(Z_n^3)\right) - Z_n\right] \\ &= 2n\left[Z_n + Z_n^2 - \frac{Z_n^2}{2} + O(Z_n^3) - Z_n\right] \\ &= nZ_n^2(1 + O(Z_n)). \end{aligned}$$

By the central limit theorem, under H_0, $Z_n = \hat{\lambda}_n - 1 = \bar{X}_n - 1 \approx^d N(0, 1/n)$. Thus $\sqrt{n}Z_n \approx^d N(0,1)$ and hence its square $nZ_n^2 \approx \chi^2_1$. $\boxed{5, \text{C}}$

If Z_n were not small, this would imply that the MLE $\hat{\lambda}_n$ is far from the truth, i.e. it is inconsistent. $\boxed{1, \text{D}}$

5. (a) Note that we can rewrite

seen $\Downarrow$

$$\hat{T}_J = T_n - (n-1)\left(\frac{1}{n}\sum_{i=1}^n T_{(-i)} - T_n\right) = nT_n - \frac{n-1}{n}\sum_{i=1}^n T_{(-i)}.$$

Since $E_\theta T_{(-i)} = \theta + b_\theta(T_{n-1}) = \theta + \frac{a}{n-1} + \frac{b}{(n-1)^2} + O(1/n^3)$, we get

$$\begin{aligned}
E\hat{T}_J &= nE_\theta T_n - (n-1)E_\theta T_{n-1} \\
&= \theta + a + \frac{b}{n} + O(1/n^3) - a - \frac{b}{n-1} + O(1/n^3) \\
&= \theta - \frac{b}{n(n-1)} + O(1/n^3) \\
&= \theta + O(1/n^2).
\end{aligned}$$

4, M

(b) For T_n,

unseen $\Downarrow$

$$ET_n = E[(\bar{X}_n)^2] = \frac{1}{n^2}E\left(\sum_{i=1}^n X_i\right)^2 = \frac{1}{n^2}\left\{nEX_1^2 + n(n-1)E[X_1X_2]\right\}.$$

But $EX_1^2 = \mathsf{Var}(X_1) + (EX_1)^2 = 1 + \mu^2$. Writing $X_i =^d \mu + Z_i$ for $Z_1, \ldots, Z_n \sim^{iid} N(0,1)$,

$$E[X_1X_2] = E[(\mu + Z_1)(\mu + Z_2)] = \mu^2.$$

Thus

$$ET_n = \frac{1}{n}(1 + \mu^2 + (n-1)\mu^2) = \mu^2 + \frac{1}{n},$$

so that $b_\theta(T_n) = 1/n$.

For $\hat{T}_J$, we can directly use the computation from (a) to get

$$E\hat{T}_J = nE_\theta T_n - (n-1)E_\theta T_{n-1} = n(\mu^2 + 1/n) - (n-1)(\mu^2 + 1/(n-1)) = \mu^2,$$

so that $b_\theta(\hat{T}_J) = 0$.

5, M

(c) By the central limit theorem, $\sqrt{n}(\bar{X}_n - \mu) \to^d N(0,1)$. Note that $T_n = g(\bar{X}_n)$ for $g(x) = x^2$. Hence applying the delta method with $g'(x) = 2x$ gives

sim. seen $\Downarrow$

$$\sqrt{n}(T_n - \mu^2) \to^d N(0, g'(\mu)^2) = N(0, 4\mu^2).$$

3, M

(d) Writing b_J as the difference of two squares and using both formulas from the hint in the second line, $\boxed{\text{unseen} \Downarrow}$

$$\begin{aligned}
b_J &= \frac{n-1}{n}\sum_{i=1}^{n}(\bar{X}_{(-i)} - \bar{X}_n)(\bar{X}_{(-i)} + \bar{X}_n) \\
&= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{\bar{X}_n - X_i}{n-1}\right)\left(\bar{X}_n + \frac{1}{n-1}(\bar{X}_n - X_i) + \bar{X}_n\right) \\
&= \frac{2\bar{X}_n}{n}\sum_{i=1}^{n}(\bar{X}_n - X_i) + \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{X}_n - X_i)^2 \\
&= 0 + \frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2
\end{aligned}$$

since $\sum_{i=1}^{n}(\bar{X}_n - X_i) = 0$. $\boxed{5, \text{M}}$

(e) $\boxed{\text{meth seen} \Downarrow}$

$$\sqrt{n}(\hat{T}_J - \mu^2) = \sqrt{n}(T_n - \mu^2) - \sqrt{n}b_J,$$

so by Slutsky's theorem, it suffices to show the last term converges to zero in probability. But by the form of b_J derived in (d): $n(n-1)b_J \sim \chi^2_{n-1}$. Therefore for any $\epsilon > 0$, by Markov's inequality,

$$P(|\sqrt{n}b_J| > \epsilon) \leq \frac{\sqrt{n}Eb_J}{\epsilon} = \frac{\sqrt{n}E\chi^2_{n-1}}{\epsilon n(n-1)} = \frac{1}{\epsilon\sqrt{n}} \to 0,$$

i.e. $\sqrt{n}b_J \to^P 0$ as $n \to \infty$. $\boxed{3, \text{M}}$

Review of mark distribution:

Total A marks: 30 of 32 marks

Total B marks: 22 of 20 marks

Total C marks: 12 of 12 marks

Total D marks: 16 of 16 marks

Total marks: 100 of 80 marks

Total Mastery marks: 20 of 20 marks

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.		
ExamModuleCode	**QuestionNumber**	**Comments for Students**
MATH60043/70043	1	No Comments Received
MATH60043/70043	2	Some students were unable to derive the likelihood in part (c), using instead a binomial or products of binomials in place of the required multinomial. Many students found the asymptotic results (consistency in part (b) and limiting distribution in part (c) difficult.
MATH60043/70043	3	Only a small number of students completed the mastery question and they generally found it challenging.
MATH60043/70043	4	No Comments Received
MATH70043	5	Only a small number of students completed the mastery question and they generally found it challenging.