Imperial College London

MATH97073

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

Statistical Theory 1

Date: Wednesday, 12 May 2021

Time: 09:00 to 11:30

Time Allowed: 2.5 hours

Upload Time Allowed: 30 minutes

This paper has 5 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. Consider observing $X_1, \ldots, X_n \sim^{iid} f_\theta$ in the model $\{f_\theta : \theta \in (0, \infty)\}$ with probability density function

$$f_\theta(x) = (1-\alpha)(x-\theta)^{-\alpha} 1_{[\theta, \theta+1]}(x),$$

where $\alpha \in (0,1)$ is **known** and 1_A denotes the indicator function of the set A. The goal is to estimate the parameter θ.

(a) Compute the mean and variance of X_1. Construct an unbiased estimator of θ of the form $\tilde{\theta}_n = \bar{X}_n + c(\alpha)$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, and you should specify $c(\alpha)$. (6 marks)

(b) Show that $\tilde{\theta}_n$ is consistent and find the limit in distribution of $\sqrt{n}(\tilde{\theta}_n - \theta)$ as $n \to \infty$. Clearly mention any theorems that you use. (3 marks)

(c) Find the maximum likelihood estimator $\hat{\theta}_n$ of θ. Compute $P_\theta(\hat{\theta}_n - \theta > t)$ for all $t \in \mathbb{R}$. Is $\hat{\theta}_n$ unbiased? (7 marks)

(d) For $t > 0$, show that $P_\theta(n^\beta(\hat{\theta}_n - \theta) > t)$ has a limit in $(0,1)$ for some $\beta > 0$ as $n \to \infty$. Give explicitly the value of β and the limit. Should one prefer to use $\hat{\theta}_n$ or $\tilde{\theta}_n$? Justify your answer. (4 marks)

(Total: 20 marks)

2. Let $X_1, \ldots, X_n \sim^{iid}$ Exp(θ) have exponential distribution with probability density function $f_\theta(x) = \theta e^{-\theta x}$ for $x > 0$ and $\theta \in (0, \infty)$. Recall that the Gamma(α, β) distribution, with parameters $\alpha, \beta > 0$, has probability density function

$$\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \qquad y > 0,$$

where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the Gamma function.

Note: *throughout this question, you may use without proof that* $\sum_{i=1}^n X_i \sim$ *Gamma*(n, θ).

(a) Show that $T_n = \sum_{i=1}^n X_i$ is *minimal sufficient* and *complete* for θ.

(5 marks)

You may assume the uniqueness of Laplace transforms: if $\mathcal{L}h(t) = \int_0^\infty h(x) e^{-tx} dx = 0$ *for all* $t > 0$, *then* $h(x) = 0$ *for all* $x > 0$.

(b) Suppose we assign the prior $\theta \sim$ Gamma(α, β), $\alpha, \beta > 0$. What is the posterior distribution of θ given $X_1, \ldots, X_n$? (3 marks)

(c) (i) Show that the expectation of $\phi = \theta^{-1}$ under the *prior* distribution in (b) is $\frac{\beta}{\alpha-1}$ if $\alpha > 1$. What is the prior variance of ϕ? (4 marks)

(ii) Deduce the posterior expectation and variance of ϕ given $X_1, \ldots, X_n$. (2 marks)

(d) (i) Let $\tilde{\phi}_n = \tilde{\phi}_{n,\alpha}$ denote the limiting form of the posterior mean for fixed α and as $\beta \to 0$. Consider $\tilde{\phi}_n$ as an estimator in the frequentist framework where $X_1, \ldots, X_n \sim$ Exp(θ_0) for some true $\theta_0 > 0$. Show that for some choice of α, which you should specify, $\tilde{\phi}_n$ is the minimum variance unbiased estimator of ϕ. (4 marks)

(ii) Without computing the Cramer-Rao lower bound for ϕ, state whether or not the variance of $\tilde{\phi}_n$ for the choice of α in (d)(i) achieves the Cramer-Rao lower bound, justifying your answer. (2 marks)

(Total: 20 marks)

3. (a) In the context of decision theory, explain the meaning of the following terms: *loss function*, *decision rule*, the *risk function* of a decision rule and a *Bayes rule* with respect to a prior π. Explain how a Bayes rule with respect to a prior π can be constructed. (6 marks)

Suppose that $X_1 \ldots, X_n \sim^{iid} N(0, \theta)$, where $\theta > 0$ is the **variance**, and assign to θ a prior with probability density function π on $(0, \infty)$. Consider the problem of estimating θ under squared error loss $L(a, \theta) = (a - \theta)^2$. Write $X = (X_1, \ldots, X_n)$ and $\|X\|^2 = \sum_{i=1}^n X_i^2$ for the usual Euclidean norm of X.

You may use without proof that if $Z \sim N(0,1)$, then $EZ^4 = 3$.

(b) (i) Consider a decision rule (estimator) of the form $\hat{\theta} = \alpha\|X\|^2$, $\alpha \in \mathbb{R}$, and an arbitrary prior with probability density function π on $(0, \infty)$. Show that the corresponding Bayes risk can be written as

$$c_n(\alpha) \int_0^\infty \theta^2 \pi(\theta) d\theta,$$

where $c_n(\alpha)$ is a function you should specify. (4 marks)

(ii) By considering decision rules of the form $\hat{\theta} = \alpha\|X\|^2$, prove that if $\alpha \neq \frac{1}{n+2}$ then the estimator $\hat{\theta} = \alpha\|X\|^2$ is not a Bayes rule with respect to any prior as in (b)(i). (2 marks)

(c) By considering decision rules of the form $\hat{\theta}(X) = \alpha\|X\|^2 + \beta$, prove that if $\alpha \neq \frac{1}{n}$ then the estimator $\hat{\theta} = \alpha\|X\|^2$ is not a Bayes rule with respect to *any* prior π as in (b)(i). (8 marks)

(Total: 20 marks)

4. (a) (i) In the context of hypothesis testing, define the following terms: *power function*, *type I error*, *type II error*, *uniformly most powerful test*. (4 marks)

(ii) Let $X \sim P_\theta$, where $\{P_\theta : \theta \in \Theta\}$ is a statistical model with $\Theta \subseteq \mathbb{R}$, and X takes values in a sample space $\mathcal{X}$. Suppose for every $\theta_0 \in \Theta$ there exists a (non-randomized) level α test $\phi_{\theta_0} : \mathcal{X} \to \{0, 1\}$ of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Explain how one can use the tests $\{\phi_{\theta_0} : \theta_0 \in \Theta\}$ to construct a $100(1-\alpha)\%$ confidence set for θ. (2 marks)

Let X be a single observation from a probability density function f. We wish to test the hypotheses

$$H_0 : f = f_0, \qquad \text{against} \qquad H_1 : f = f_1,$$

where $f_0(x) = \frac{1}{2}|x|e^{-x^2/2}$ and $f_1(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ for $x \in \mathbb{R}$. Let $\Phi(x)$ denote the cumulative distribution function of the standard normal $N(0,1)$ distribution.

(b) (i) Find the test of size α, $0 < \alpha < 1$, which maximizes the power. (6 marks)

(ii) Compute the power function of the test in (b)(i) in terms of α and Φ. (3 marks)

(c) Find the size of the test in (b) that minimizes the sum of the error probabilities. (5 marks)

(Total: 20 marks)

5. (a) Given independent and identically distributed observations $X_1, \ldots, X_n$ with finite mean $EX_i = \mu$ and variance $\text{Var}(X_i) = \sigma^2$, explain the notion of a *bootstrap sample* $X_1^*, \ldots, X_n^*$ and explain how you can use it to construct a confidence interval C_n for μ. (5 marks)

(b) Show that, given a sample of n *distinct* observations, the probability that the maximum of the bootstrap sample equals the maximum of the original sample is $1 - (1 - \frac{1}{n})^n$. (2 marks)

Now assume that $X_1, \ldots, X_n \sim^{iid} U[0, \theta]$ are independent and identically distributed uniform random variables. Our goal is to approximate the distribution of the quantity

$$Q_n(X, \theta) = \frac{n(\theta - X_{(n)})}{\theta},$$

where $\hat{\theta}_n = X_{(n)} = \max\{X_1, \ldots, X_n\}$ is the MLE for θ.

(c) (i) Show that $Q_n(X, \theta)$ converges in distribution to an exponential distribution as $n \to \infty$. (2 marks)

[Recall an exponential distribution has density function $\lambda e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$].

(ii) Use this result to construct an asymptotic $(1 - \alpha)100\%$ confidence interval for θ, where $0 < \alpha < 1$. (3 marks)

(d) (i) Let $X_1^*, \ldots, X_n^*$ be a bootstrap sample based on $X_1, \ldots, X_n$. Explain why

$$Q_n(X^*, \hat{\theta}_n) = \frac{n(X_{(n)} - X_{(n)}^*)}{X_{(n)}}$$

is a bootstrap realization of $Q_n(X, \theta)$. (2 marks)

(ii) Show that as $n \to \infty$,

$$\mathbb{P}(Q_n(X^*, \hat{\theta}_n) = 0) \to 1 - e^{-1},$$

where $\mathbb{P}$ denotes the joint distribution of $X_1, \ldots, X_n, X_1^*, \ldots, X_n^*$. (4 marks)
Hint: recall that for any random variables Y, Z, $E[Y] = E[E[Y|Z]]$.

(iii) Based on this last result, is the bootstrap a suitable method to construct a confidence interval for θ in this model for large n? Justify your answer. (2 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2021

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

MATH96046/MATH97073

Statistical Theory (Solutions)

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

1. (a) Using the change of variable $x - \theta = y$ and that $f_0(y) = (1-\alpha)y^{-\alpha}1_{[0,1]}(y)$ is a pdf, the mean is computed as unseen $\Downarrow$

$$
\begin{aligned}
E_\theta X_1 = \int_\theta^{\theta+1} x(1-\alpha)(x-\theta)^{-\alpha}dx = (1-\alpha)\int_0^1 (\theta+y)y^{-\alpha}dy \\
&= \theta + (1-\alpha)\left[\frac{y^{2-\alpha}}{2-\alpha}\right]_0^1 \\
&= \theta + \frac{1-\alpha}{2-\alpha}.
\end{aligned}
$$

For the variance, using a similar computation and the formula derived above, 2, A

$$
\begin{aligned}
E_\theta X_1^2 = \int_\theta^{\theta+1} x^2(1-\alpha)(x-\theta)^{-\alpha}dx = (1-\alpha)\int_0^1 (\theta+y)^2 y^{-\alpha}dy \\
&= \theta^2 + 2\frac{1-\alpha}{2-\alpha}\theta + (1-\alpha)\left[\frac{y^{3-\alpha}}{3-\alpha}\right]_0^1 \\
&= \theta^2 + \frac{2(1-\alpha)}{2-\alpha}\theta + \frac{1-\alpha}{3-\alpha}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathsf{Var}_\theta(X_1) = E_\theta X_1^2 - (E_\theta X_1)^2 &= \theta^2 + \frac{2(1-\alpha)}{2-\alpha}\theta + \frac{1-\alpha}{3-\alpha} - \left(\theta + \frac{1-\alpha}{2-\alpha}\right)^2 \\
&= \frac{1-\alpha}{(2-\alpha)^2(3-\alpha)}.
\end{aligned}
$$

This can be solved more easily by first noting that $X_i =^d \theta + Y_i$ for $Y_1, \dots, Y_n \sim^{iid} f_0$, and then using $\mathsf{Var}_\theta(X_1) = \mathsf{Var}(Y_1)$. 3, A

For an estimator $\tilde\theta_n = \bar X_n + c(\alpha)$ we have bias $E_\theta \tilde\theta_n - \theta = E_\theta X_1 + c(\alpha) - \theta = \frac{1-\alpha}{2-\alpha} + c(\alpha)$, so the estimator is unbiased if $c(\alpha) = -\frac{1-\alpha}{2-\alpha}$. 1, A

(b) By the weak law of large numbers, $\tilde\theta_n = \bar X_n + c(\alpha) \to^p E_\theta X_1 + c(\alpha) = \theta$ by (a). sim. seen $\Downarrow$
By the central limit theorem, the definition of $c(\alpha)$ and the expectation in (a), 1, A

$$
\begin{aligned}
\sqrt{n}(\tilde\theta_n - \theta) &= \sqrt{n}(\bar X_n - E_\theta X_1 + E_\theta X_1 + c(\alpha) - \theta) \\
&= \sqrt{n}(\bar X_n - E_\theta X_1) \\
&\to^d N(0, \mathsf{Var}_\theta(X_1)) = N\left(0, \frac{1-\alpha}{(2-\alpha)^2(3-\alpha)}\right).
\end{aligned}
$$

2, A

(c) The likelihood equals meth seen $\Downarrow$

$$
\begin{aligned}
L_n(\theta) &= \prod_{i=1}^n (1-\alpha)(x_i-\theta)^{-\alpha}1_{[\theta,\theta+1]}(x_i) \\
&= (1-\alpha)^n\left(\prod_{i=1}^n (x_i-\theta)^{-\alpha}\right)1\{\min_i x_i \geq \theta\}1\{\max_i x_i \leq \theta+1\},
\end{aligned}
$$

where we have simplified the indicator functions (seen similar). For $\theta \leq \min_i x_i$, the map $\theta \mapsto (x_i-\theta)^{-\alpha}$ is increasing in θ, and in fact tends to infinity as $\theta \to x_i$. Thus the likelihood is maximized by taking the largest value of θ such that the indicator functions are non-zero, i.e. $\hat\theta_n = \min_i x_i$. 3, B

For $t \in (0,1)$, we have

$$\begin{aligned} P_\theta(\hat{\theta}_n - \theta > t) = P_\theta(\min X_i > \theta + t) &= P_\theta(X_1 > \theta + t)^n \\ &= \left(\int_{\theta+t}^{\theta+1} (1-\alpha)(x-\theta)^{-\alpha} dx \right)^n \\ &= \left(\int_t^1 (1-\alpha) y^{-\alpha} dy \right)^n \\ &= \left(\left[y^{1-\alpha} \right]_t^1 \right)^n = \left(1 - t^{1-\alpha}\right)^n. \end{aligned}$$

We trivially have the probability equals 1 for $t \leq 0$ and 0 for $t \geq 1$. $\boxed{2,\ \mathrm{A}}$

It is clearly biased. This follows since with probability 1, $\min X_i \geq \theta$ and $P_\theta(\min X_i = \theta) = 0$, so $E_\theta \min X_i > \theta$. Alternatively, the students can derive this directly using the pdf g_n of $\hat{\theta}_n - \theta$ from differentiating the last expression:

$$g_n(t) = \frac{d}{dt} P_\theta(\hat{\theta}_n - \theta \leq t) = n(1-\alpha) t^{-\alpha} (1 - t^{1-\alpha})^{n-1},$$

from which $E_\theta \hat{\theta}_n = \theta + \int_0^1 t g_n(t) dt > \theta$ since all the terms in the integral are positive. $\boxed{2,\ \mathrm{C}}$

(d) For $t > 0$, using the formula from (c), $\boxed{\text{unseen} \Downarrow}$

$$P_\theta(n^\beta(\hat{\theta} - \theta) > t) = \left(1 - \left(t/n^\beta\right)^{1-\alpha}\right)^n = \left(1 - \frac{t^{1-\alpha}}{n^{\beta(1-\alpha)}}\right)^n.$$

Setting $\beta = \frac{1}{1-\alpha} > 1$, the right-hand side equals $(1 - t^{1-\alpha}/n)^n \to e^{-t^{1-\alpha}} \in (0,1)$ as $n \to \infty$. $\boxed{2,\ \mathrm{D}}$

Even though $\hat{\theta}_n$ is biased whereas $\tilde{\theta}_n$ is unbiased, the second estimator converges to the truth at a much faster rate ($n^{-1/(1-\alpha)}$ compared to $n^{-1/2}$), so should be preferred. $\boxed{2,\ \mathrm{D}}$

2. (a) Consider the ratio of the pdfs $\boxed{\text{meth seen} \Downarrow}$

$$\frac{\prod_i f_\theta(x_i)}{\prod_i f_\theta(x_i')} = \frac{\prod_i \theta e^{-\theta x_i}}{\prod_i \theta e^{-\theta x_i'}} = e^{\theta(\sum_i x_i' - \sum_i x_i)}.$$

This ratio does not depend on θ if and only if $\sum_i x_i = \sum_i x_i'$ and hence T_n is minimal sufficient for θ by a theorem in the notes. $\boxed{2, \text{A}}$

For completeness, since $T_n \sim$ Gamma(n, θ), suppose g is such that

$$E_\theta g(T_n) = \int_0^\infty g(t) \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} dt = 0 \quad \forall \theta > 0.$$

Writing $h(t) = g(t)t^{n-1}$, the above is equivalent to $\mathcal{L}h(\theta) = 0$ for all $\theta > 0$. Since this matches the Laplace transform of the zero function (or using the hint directly), $h(t) = 0$ for all $t > 0$, and so $g(t) = 0$. Thus T_n is complete.

If the answer uses the general completeness result for exponential families without proving the Gamma distribution is an exponential family, maximum 1 mark. $\boxed{3, \text{B}}$

(b) The posterior distribution is proportional to $\boxed{\text{meth seen} \Downarrow}$

$$\begin{aligned}
\pi(\theta | X_1, \ldots, X_n) &\propto \pi(\theta) \prod_{i=1}^n f_\theta(X_i) \\
&\propto \theta^{\alpha-1} e^{-\beta\theta} \theta^n e^{-\theta \sum_i x_i} \\
&\propto \theta^{n+\alpha-1} e^{-(\beta+T_n)\theta},
\end{aligned}$$

which we recognize is the form of a Gamma$(n + \alpha, \beta + T_n)$ distribution. $\boxed{3, \text{A}}$

(c) (i) Rewriting the integral as the pdf of a Gamma$(\alpha - 1, \beta)$ distribution, the prior expectation is $\boxed{\text{seen/sim.seen} \Downarrow}$

$$\begin{aligned}
E^\pi[1/\theta] &= \int_0^\infty \frac{1}{\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} \int_0^\infty \frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)} \theta^{\alpha-2} e^{-\beta\theta} d\theta \\
&= \frac{\beta}{\alpha - 1}.
\end{aligned}$$

The prior variance can be computed similarly, $\boxed{2, \text{A}}$

$$\begin{aligned}
E^\pi[1/\theta^2] &= \int_0^\infty \frac{1}{\theta^2} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-2)}{\beta^{\alpha-2}} = \frac{\beta^2}{(\alpha-1)(\alpha-2)},
\end{aligned}$$

from which we obtain

$$\begin{aligned}
\text{Var}_\pi(1/\theta^2) = E^\pi[1/\theta^2] - (E^\pi[1/\theta])^2 &= \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \left(\frac{\beta}{\alpha-1}\right)^2 \\
&= \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}.
\end{aligned}$$

$\boxed{2, \text{A}}$

(ii) Using the formulas from (c)(i), but updating the parameters using the conjugate formula from (b), the posterior mean and variance are seen/sim.seen ⇓

$$E^\pi[1/\theta|X_1,\dots,X_n] = \frac{\beta+T_n}{n+\alpha-1},$$
$$\mathsf{Var}_\pi(1/\theta|X_1,\dots,X_n) = \frac{(\beta+T_n)^2}{(n+\alpha-1)^2(n+\alpha-2)}.$$

2, B

(d) (i) The posterior mean satisfies $\frac{\beta+T_n}{n+\alpha-1} \to \frac{T_n}{n+\alpha-1} =: \tilde{\phi}_n$ as $\beta \to 0$. Since $T_n \sim \mathsf{Gamma}(n,\theta)$, we have by a similar computation to previously part seen ⇓

$$E_\theta T_n = \int_0^\infty t\frac{\theta^n}{\Gamma(n)}t^{n-1}e^{-\theta t}dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n+1)}{\theta^{n+1}} = \frac{n}{\theta}.$$

Hence the frequentist expectation of the posterior mean is $E_\theta\tilde{\phi}_n = \frac{n}{n+\alpha-1}\frac{1}{\theta} = \frac{n}{n+\alpha-1}\phi$. If $\alpha = 1$, $\tilde{\phi}_n$ is therefore unbiased and a function of the complete and sufficient statistic T_n, and hence it is uniformly minimum variance unbiased by the Lehmann-Scheffe theorem. 4, C

(ii) The likelihood takes the form part seen ⇓

$$\prod_i \theta e^{-\theta x_i} = \phi^{-n}e^{-\phi^{-1}\sum_i x_i} = \exp(-n\log\phi - \phi^{-1}n\tilde{\phi}_n),$$

which is of exponential family form with $\tilde{\phi}_n$ as natural statistic. It thus attains the Cramer-Rao bound. 2, C

3. (a) A *loss function* is a non-negative function $L : \mathcal{A} \times \Theta \to [0, \infty)$ that determines the cost of action $a \in \mathcal{A}$ for a given parameter $\theta \in \Theta$. seen ⇓

A *decision rule* $\delta : \mathcal{X} \to \mathcal{A}$ makes a decision/action $\delta(X)$ upon observing X.

The *risk function* of δ is the expected loss under P_θ as a function of θ: $R(\delta, \theta) = E_\theta[L(\delta(X), \theta)]$.

A *Bayes rule* with respect to a prior π is any decision rule that minimize the Bayes risk $R_\pi(\delta) = E_{\theta\sim\pi}[R(\delta, \theta)]$, where the expectation is taken over the prior π.

A Bayes rule can be obtained by directly minimizing the Bayes risk. However, a more common approach is to minimize the posterior risk $R_\pi(\delta(x)) = E_\pi[L(\delta(x)), \theta)|x]$, which is the expected loss under the posterior. This is because any minimizer of the posterior risk also minimizes the Bayes risk. 6, A

(b) (i) We write $X_i =^d \sqrt{\theta} Z_i$, for $Z_1, \dots, Z_n \sim^{iid} N(0, 1)$. We compute the risk function of $\hat{\theta}(X) = \alpha\|X\|^2$ for $\theta > 0$: meth seen ⇓

$$\begin{aligned}
R(\hat{\theta}, \theta) &= E_\theta(\alpha\|X\|^2 - \theta)^2 \\
&= E_\theta\left(\alpha\theta(Z_1^2 + \dots + \dots + Z_n^2) - \theta\right)^2 \\
&= \theta^2 E_\theta\left\{1 + \alpha^2 \sum_{i=1}^n Z_i^4 - 2\alpha \sum_{i=1}^n Z_i^2 + \alpha^2 \sum_{i=1}^n \sum_{j\neq i} Z_i^2 Z_j^2\right\} \\
&= \theta^2\left\{1 + 3n\alpha^2 - 2n\alpha + n(n-1)\alpha^2\right\} \\
&= \theta^2\left\{(n^2 + 2n)\alpha^2 - 2n\alpha + 1\right\} =: c_n(\alpha)\theta^2.
\end{aligned}$$

Therefore the Bayes risk is

$$R_\pi(\hat{\theta}) = E_{\theta\sim\pi}[R(\hat{\theta}, \theta)] = c_n(\alpha) \int_0^\infty \theta^2 \pi(\theta) d\theta.$$

4, B

(ii) We see from (b)(i) that any decision rule of the form $\alpha\|X\|^2$ will have larger Bayes risk than the one that minimizes the constant $c_n(\alpha)$. This is a quadratic function in α, so it is minimized at its unique stationary point satisfying unseen ⇓

$$\frac{dc_n(\alpha)}{d\alpha} = 2n(n+2)\alpha - 2n\alpha = 0,$$

which gives $\alpha = \frac{1}{n+2}$. 2, C

(c) For decision rules of the form $\hat{\theta} = \alpha\|X\|^2 + \beta$, we have unseen ⇓

$$E_\theta\hat{\theta} = \alpha E_\theta \sum_{i=1}^n \theta Z_i^2 + \beta = \alpha n\theta + \beta$$

$$\mathsf{Var}_\theta(\hat{\theta}) = \mathsf{Var}_\theta\left(\alpha \sum_{i=1}^n \theta Z_i^2 + \beta\right) = \alpha^2\theta^2 n \mathsf{Var}_\theta(Z_i^2) = \alpha^2 n\theta^2(EZ_i^4 - (EZ_i^2)) = 2\alpha^2 n\theta^2.$$

Hence by the bias-variance decomposition, we have risk $R(\hat{\theta}, \theta) = 2\alpha^2 n\theta^2 + (\alpha n\theta + \beta - \theta)^2$. The Bayes risk is therefore

$$R_\pi(\hat{\theta}, \theta) = 2\alpha^2 n \int_0^\infty \theta^2 \pi(\theta) d\theta + \int_0^\infty (\alpha n\theta + \beta - \theta)^2 \pi(\theta) d\theta.$$

We minimize this in β for a given α. Since this is again quadratic in β, it is minimized at its unique stationary point satisfying

$$\frac{dR_\pi(\hat{\theta}, \theta)}{d\beta} = 2\int_0^\infty [\beta - (1 - \alpha n)\theta]\pi(\theta) d\theta = 2\beta - 2(1 - \alpha n) \int_0^\infty \theta\pi(\theta) d\theta = 0.$$

For $\alpha \neq \frac{1}{n}$, this is minimized by $\beta_\alpha = (1-\alpha n)\int_0^\infty \theta\pi(\theta)d\theta \neq 0$, and hence the decision rule $\alpha\|X\|^2$ has larger Bayes risk than $\alpha\|X\|^2 + \beta_\alpha$. Thus $\alpha\|X\|^2$ is not a Bayes rule for any prior. $\boxed{2, \text{C}}$

$\boxed{6, \text{D}}$

4. (a) (i) The *power function* $\pi_\phi : \Theta \to [0,1]$ of a test ϕ is the probability of rejecting the null hypothesis H_0 under P_θ, i.e. $\pi_\phi(\theta) = P_\theta(\text{reject } H_0)$. seen ⇓

A *type I error* is the error of rejecting the null hypothesis H_0 when it is actually true.

A *type II error* is the error of rejecting the alternative hypothesis H_1 when it is actually true.

A test is *uniformly most powerful* of size α for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ if (i) it is a level α test ($\sup_{\theta\in\Theta_0} \pi_\phi(\theta) \le \alpha$) and (ii) any other level α test ϕ^* has smaller power, i.e. $\pi_{\phi^*}(\theta) \le \pi_\phi(\theta)$ for all $\theta \in \Theta_1$.

4, A

(ii) Let $R(\theta)$ be the critical/rejection region of the test ϕ_{θ_0}, where we reject H_0 if $X \in R(\theta_0)$. Let $A(\theta_0) = R(\theta_0)^c$ be its complement (sometimes called *acceptance region*). We can then construct a $100(1-\alpha)\%$ confidence interval for θ by seen ⇓

$$C(X) = \{\theta \in \Theta : X \in A(\theta)\} = \{\theta \in \Theta : X \in R(\theta)^c\}.$$

2, A

(b) (i) By the Neyman-Pearson lemma, we know the likelihood ratio test is uniformly most powerful (UMP) for testing two simple hypotheses. The test statistic is meth seen ⇓

$$\frac{f_1(x)}{f_0(x)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}}{\frac{1}{2}|x|e^{-x^2/2}} = \sqrt{2/\pi}\frac{1}{|x|}.$$

Thus the UMP test takes the form $1\{\sqrt{2/\pi}|x|^{-1} \ge k\}$, i.e. reject if the above statistic is larger than (or equal to) k. We choose k by considering the type I error: 2, B

$$\begin{aligned}
\alpha = P_0(\text{reject } H_0) &= P_0(\sqrt{2/\pi}/|X| \ge k) \\
&= P_0(\sqrt{2/\pi}\tfrac{1}{k} \ge |X|) \\
&= \int_{-\sqrt{2/\pi}\frac{1}{k}}^{\sqrt{2/\pi}\frac{1}{k}} \frac{1}{2}|x|e^{-x^2/2}dx \\
&= \int_0^{\sqrt{2/\pi}\frac{1}{k}} xe^{-x^2/2}dx = \left[-e^{-x^2/2}\right]_0^{\sqrt{2/\pi}\frac{1}{k}} = 1 - e^{-1/(\pi k^2)}.
\end{aligned}$$

Rearranging gives $1/k^2 = \pi \log(\frac{1}{1-\alpha})$. Therefore the UMP test of size α rejects H_0 if and only if $|X| \le \sqrt{\frac{2}{\pi k^2}} = \sqrt{2\log(\frac{1}{1-\alpha})}$.

4, B

(ii) The power function at H_0 is $P_0(\text{reject } H_0) = \alpha$ by construction. For the alternative H_1, unseen ⇓

$$\begin{aligned}
P_1(\text{reject } H_0) &= P_1\left(|X| \le \sqrt{2\log(\tfrac{1}{1-\alpha})}\right) \\
&= \int_{-\sqrt{2\log(\frac{1}{1-\alpha})}}^{\sqrt{2\log(\frac{1}{1-\alpha})}} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = 2\Phi\left(\sqrt{2\log(\tfrac{1}{1-\alpha})}\right) - 1.
\end{aligned}$$

3, B

(c) The sum of the type I and II error probabilities is

unseen $\Downarrow$

$$\varphi(\alpha) = \alpha + 2 - 2\Phi\left(\sqrt{2\log(\tfrac{1}{1-\alpha})}\right).$$

Note that

$$\frac{d}{d\alpha}\sqrt{2\log(\tfrac{1}{1-\alpha})} = \frac{d}{d\alpha}\sqrt{-2\log(1-\alpha)} = \frac{1}{(1-\alpha)\sqrt{2\log(\tfrac{1}{1-\alpha})}}.$$

Differentiating $\varphi(\alpha)$, since $\Phi'(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$,

$$\begin{aligned}\frac{d\varphi}{d\alpha} &= 1 - 2 \times \frac{1}{\sqrt{2\pi}}e^{-\log(\frac{1}{1-\alpha})} \times \frac{1}{(1-\alpha)\sqrt{2\log(\tfrac{1}{1-\alpha})}} \\ &= 1 - \frac{1}{\sqrt{\pi\log(\tfrac{1}{1-\alpha})}} = 0.\end{aligned}$$

This is solved by $\pi\log(\frac{1}{1-\alpha}) = 1$, which rearranges to $\alpha = 1 - e^{-1/\pi}$. We now check that it is a global minimum on the range $\alpha \in (0,1)$.

$$\frac{d^2\varphi}{d\alpha^2} = -\frac{1}{\sqrt{\pi}}\frac{d}{d\alpha}\left(\log(\frac{1}{1-\alpha})\right)^{-1/2} = \frac{1}{2\sqrt{\pi}(1-\alpha)(\log(\frac{1}{1-\alpha}))^{3/2}} > 0$$

for all $\alpha \in (0,1)$ since $1-\alpha > 0$ and $\log(\frac{1}{1-\alpha}) = -\log(1-\alpha) > 0$. Thus $\alpha = 1 - e^{-1/\pi}$ minimizes the sum of the error probabilities.

5, D

5. (a) A bootstrap sample is a sample of n i.i.d. observations drawn from the empirical distribution function $F_n(t) = \frac{1}{n}\sum_{i=1}^n 1\{X_i \le t\}$. Alternatively, we select each X_i^* with probability $P(X_i^* = X_k | X_1, \ldots, X_n) = 1/n$ for $k = 1, \ldots, n$, i.e. we set X_i^* equal to one of the observations $X_1, \ldots, X_n$, each having equal probability $1/n$. seen ⇓

2, M

We draw B i.i.d. bootstrap samples $X_{b,1}^*, \ldots, X_{b,n}^* \sim^{iid} \hat{F}_n$ for $b = 1, \ldots, B$ large, and for each $b = 1, \ldots, B$ we compute the mean of the bootstrap sample $\bar{X}_{n,b}^* = \frac{1}{n}\sum_{i=1}^n X_{b,i}^*$. We then compute the empirical bootstrap distribution of the pivot $\bar{X}_{n,b}^* - \bar{X}_n$ as

$$\hat{H}(t) = \frac{1}{B}\sum_{b=1}^{B} 1\{\bar{X}_{n,b}^* - \bar{X}_n \le t\},$$

and take the $\alpha/2$ and $1-\alpha/2$ quantiles of this distribution, call them $h_{\alpha/2}$, $h_{1-\alpha/2}$, respectively. A bootstrap confidence interval is then

$$C_n = [\bar{X}_n - h_{\alpha/2}, \bar{X}_n + h_{1-\alpha/2}].$$

Note that if instead we use the quantiles $g_{\alpha/2}$ and $g_{1-\alpha/2}$ of the distribution $\hat{G}(t) = \frac{1}{n}\sum_{b=1}^B 1\{\bar{X}_{n,b}^* \le t\}$ (i.e. $g_{\alpha/2} = \bar{X}_n h_{\alpha/2}$), then the confidence interval is equivalently written as

$$C_n = [2\bar{X}_n - g_{\alpha/2}, 2\bar{X}_n + g_{1-\alpha/2}].$$

3, M

(b) The probability that the maximums are not equal is the probability the bootstrap sample does not contain $X_{(n)} = \max\{X_1, \ldots, X_n\}$. This probability equals $(\frac{n-1}{n})^n = (1 - \frac{1}{n})^n$, so that the desired probability is $1 - (1 - \frac{1}{n})^n$. unseen ⇓

2, M

(c) (i) For $t \in [0, n]$, the distribution of $Q_n(X, \theta)$ can be computed as seen/sim.seen ⇓

$$\begin{aligned} P_\theta(Q_n(X,\theta) \ge t) = P_\theta(n(\theta - X_{(n)}) \ge t\theta) &= P_\theta(\theta - \theta t/n \ge X_{(n)}) \\ &= P_\theta(\theta(1 - t/n) \ge X_1)^n \\ &= (1 - t/n)^n \to e^{-t} \end{aligned}$$

as $n \to \infty$. Thus $Q_n(X,\theta) \to^d \text{Exp}(1)$ as $n \to \infty$.

2, M

(ii) For the exponential distribution, a shortest region of given probability takes the form $[0, c]$ (this is not technically reqiured). We note that $Q_n(X,\theta)$ is an (asymptotically) pivotal quantity for θ, so for large n, unseen ⇓

$$P_\theta\left(0 \le \frac{n(\theta - X_{(n)})}{\theta} \le c_\alpha\right) \approx P_\theta\left(0 \le \text{Exp}(1) \le c_\alpha\right) = 1 - e^{-c_\alpha} = 1 - \alpha,$$

giving $c_\alpha = \log(1/\alpha)$. Rearranging for θ,

$$\begin{aligned} 0 \le \frac{n(\theta - X_{(n)})}{\theta} \le c_\alpha &\iff nX_{(n)} \le n\theta \le nX_{(n)} + c_\alpha\theta \\ &\iff X_{(n)} \le \theta \le \frac{nX_{(n)}}{n - c_\alpha} = \frac{X_{(n)}}{1 - \frac{1}{n}\log(1/\alpha)}. \end{aligned}$$

Therefore, an asymptotic $(1-\alpha)100\%$ confidence interval for θ is

$$[X_{(n)}, \frac{X_{(n)}}{1 - \frac{1}{n}\log(1/\alpha)}].$$

3, M

(d) (i) When making a bootstrap sample, we replace the true underlying distribution F by the empirical distribution function F_n based on $X_1, \ldots, X_n$. Thus we replace the maximum value θ of the support of our distribution with the maximum $X_{(n)}$ of the empirical distribution, and the observations X_i with our bootstrap sample X_i^*. seen/sim.seen ⇓

2, M

(ii) Writing $\mathbb{P}(A) = \mathbb{E}1_A = \mathbb{E}[\mathbb{E}[1_A | X_1, \ldots, X_n]]$, unseen ⇓

$$\begin{aligned}\mathbb{P}\left(Q_n(X^*, \hat{\theta}_n) = 0\right) &= \mathbb{E}\left(\mathbb{E}\left[1\{Q_n(X^*, \hat{\theta}_n) = 0\} | X_1, \ldots, X_n\right]\right) \\ &= \mathbb{E}\left(\mathbb{P}\left(Q_n(X^*, \hat{\theta}_n) = 0 \middle| X_1, \ldots, X_n\right)\right)\end{aligned}$$

Since $Q_n(X^*, \hat{\theta}_n) = 0$ if and only if $X^*_{(n)} = X_{(n)}$, the conditional probability equals

$$\mathbb{P}\left(X_n^* = X_n | X_1, \ldots, X_n\right) = 1 - (1 - 1/n)^n$$

by part (b) since the $X_1, \ldots, X_n$ are distinct with probability 1. Therefore,

$$\mathbb{P}\left(Q_n(X^*, \hat{\theta}_n) = 0\right) = \mathbb{E}\left[1 - (1 - 1/n)^n\right] = 1 - (1 - 1/n)^n \to 1 - e^{-1}$$

as $n \to \infty$.

4, M

(iii) No, since the bootstrap distribution is not a good approximation the distribution of $\hat{\theta}_n = X_{(n)}$ for large n. The MLE has asymptotic distribution equal to an exponential random variables, whereas the bootstrap will put a lot of mass $(1 - e^{-1})$ on a single point. unseen ⇓

2, M

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
MATH96046 MATH97073	1	A significant number of candidates struggled with the integration in (a). (b) was generally well answered. For (c) the MLE was generally well tackled, though many candidates minimized rather than maximized the likelihood. (d) was not well-answered. Many candidates struggled with evaluating the limit and few picked up that the second estimator converges more quickly to the truth.
MATH96046 MATH97073	2	Part a) was generally well answered. Some students missed out the proof of completeness. Part b) was also well answered, and most students computed this correctly. Part c i) involved some straightforward computations which was mostly answered correctly. Similarly for part ii), except some students re-worked the whole computation because they missed the fact they could use results from b) and c i). Some students missed the proof that the estimator is unbiased is part d i). In part d ii) is was important to make clear that it is an exponential family in the \phi parameter, and the natural statistic is equal to the estimator.
MATH96046 MATH97073	3	Part a) was bookwork, and generally well answered. Part b i) often contained mistakes in computing the expectation. There were a few ways to compute this, and the least error-prone method was using that fact that the estimator could be written in terms of a chi-squared random variable. Part ii) was relatively straightforward following part i) and was mostly answered well. Part c) was the most challenging, and many students struggled to answer this.

MATH96046 MATH97073	4	Part a) was generally answered well. There were a few cases where students misstated the type I and type II error definitions. A common mistake in part b) was to not calculate the constant k (likelihood ratio test constant) in terms of alpha, or to not state the UMPT test. The power function of part ii) was sometimes defined incorrectly as a sum of the type I and type II errors. Part c) was quite challenging, although many students had the correct method, few students managed to compute the derivative correctly.
MATH96046 MATH97073	5	Parts (a) and (b) were generally well answered. Part (c)(ii) caused a lot more difficulty and many candidates struggled with using a pivotal quantity to compute a confidence set. Part (d) was generally either well answered or not attempted. It's a shame some candidates did not attempt (d)(iii) since this could be answered without the preceeding parts.