Imperial College London

MATH97073 MATH70043

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2022

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

Statistical Theory

Date: 26 May 2022

Time: 09:00 – 11:30 (BST)

Time Allowed: 2:30 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 ONE PDF TO THE RELEVANT DROPBOX ON BLACKBOARD WITH COMPLETED COVERSHEETS WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. (a) Consider observing $X_1, \dots, X_n \sim^{iid}$ Poisson(θ), where $\theta \in (0, \infty)$.

   (i) Compute the mean and variance of X_1 and verify that they are equal.

   (4 marks)

   Hint: for the variance, you may find it useful to consider $E_\theta X_1(X_1 - 1)$.

   (ii) Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ for θ. Find the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ as $n \to \infty$.

   (4 marks)

(b) In many applications, Poisson-modelled data show *overdispersion*, i.e. the sample variance is larger than the sample expectation. For $p \in [0, 1]$ and $\phi \in (0, \infty)$, consider the function $f_{p,\phi} : \{0, 1, 2, \dots\} \to [0, 1]$ given by

$$f(k) = f_{p,\phi}(k) = \begin{cases} p\frac{\phi^k}{k!}e^{-\phi}, & \text{if } k = 1, 2, 3, \dots, \\ (1-p) + pe^{-\phi} & \text{if } k = 0. \end{cases}$$

Show that this defines a valid probability mass function. Does it model *overdispersion*? Justify your answer.

(3 marks)

(c) Let now $Y_1, \dots, Y_n$ be i.i.d. observations from $f_{p,\phi}$ with ϕ **known**. Find the maximum likelihood estimator (MLE) $\hat{p}_n$ for p.

(5 marks)

Hint: you may find it useful to split the likelihood as

$$\prod_{i=1}^{n} f(X_i) = \prod_{i:X_i=0} f(X_i) \prod_{i:X_i\neq 0} f(X_i).$$

(d) Suppose that for any $p \in [0, 1]$, we know that $\sqrt{n}(\hat{p}_n - p) \to^d Z$ converges in distribution to a random variable Z. Suppose we wanted to test the null hypotheses that our data $Y_1, \dots, Y_n$ comes from a Poisson distribution as in (a). State the corresponding null hypotheses in terms of p.

Compute the limiting distribution of $\sqrt{n}(\hat{p}_n - p)$ under H_0, i.e. assuming $Y_1, \dots, Y_n$ are actually Poisson data.

(4 marks)

(Total: 20 marks)

2. Consider observing $X_1, \ldots, X_n \sim^{iid}$ Exp(λ), $\lambda \in (0, \infty)$, that is with density function

$$f_\lambda(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

(a) Show that $EX_1 = 1/\lambda$ and Var$(X_1) = 1/\lambda^2$.

(4 marks)

(b) Find the maximum likelihood estimator (MLE) $\hat{\lambda}_n$ for λ. Show that $\hat{\lambda}_n$ is consistent and derive the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda)$ as $n \to \infty$. Clearly mention any theorems that you use.

Use this result to construct an (asymptotic) 95%-confidence interval for λ. Clearly explain how you estimate any quantities you need (you do not need to mathematically justify the approximation, just state it).

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

(7 marks)

(c) Find an injective function $h : (0, \infty) \to \mathbb{R}$ such that the limit variance of $\sqrt{n}(h(\hat{\lambda}_n) - h(\lambda))$ as $n \to \infty$ does not depend on λ.

Use this result to construct an (asymptotic) 95%-confidence interval for $h(\lambda)$, and hence deduce one for λ.

Show that this confidence interval and the one in (b) have approximately the same size for large n.

(6 marks)

Upon learning new information about the values that λ can take, you decide to restrict the parameter space to $\lambda \in [\delta, \infty)$ for some $\delta > 0$.

(d) Compute the maximum likelihood estimator $\hat{\lambda}_n^\delta$ in this restricted model.

Now suppose that $\lambda = \delta$ so that $X_1, \ldots, X_n \sim^{iid}$ Exp(δ). Briefly comment on whether the general asymptotic theory for MLEs derived in the course applies in this case.

(3 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 the *Bayes risk* of a decision 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 $X \in N(\theta, \sigma^2)$, where $\sigma^2 > 0$ is known. Consider the parameter space $\Theta = \mathbb{R}$ and action space $\mathcal{A} = \mathbb{R}$, with loss function $L : \Theta \times \mathcal{A} \to \mathbb{R}$ given by

$$L(\theta, a) = e^{-\lambda a \theta},$$

where $\lambda > 0$ is a positive constant.

(b) Consider assigning a prior $\theta \sim N(0, \tau^2)$ for some $\tau^2 > 0$. Show that the posterior distribution is

$$\theta | X \sim N\left(\frac{\tau^2 X}{\sigma^2 + \tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}\right).$$

(3 marks)

(c) Show that for the loss function L given above,

$$\delta_0(X) = \frac{X}{\lambda \sigma^2}$$

is a Bayes rule for the $N(0, \tau^2)$ prior. Is it the unique Bayes estimator for this prior?

(4 marks)

You may use that for $Z \sim N(m, \nu^2)$, we have moment generating function

$$M_Z(t) = Ee^{tZ} = e^{mt + \frac{1}{2}t\nu^2 t^2}$$

for any $t \in \mathbb{R}$.

(d) Show that the maximal risk of δ_0 over the parameter space equals

$$\sup_{\theta \in \mathbb{R}} R(\theta, \delta_0) = 1.$$

(3 marks)

(e) By computing the Bayes risk of $\delta_0(X) = X/(\lambda \sigma^2)$ or otherwise, show that δ_0 is minimax for this decision problem.

(4 marks)

(Total: 20 marks)

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

(5 marks)

Let X be a single observation from a probability density function

$$f_\theta(x) = \frac{1}{2}\frac{1}{\Gamma(\theta)}|x|^{\theta-1}e^{-|x|}, \qquad x \in \mathbb{R}, \quad \theta \in [1, \infty),$$

where $\Gamma(\theta) = \int_0^\infty t^{\theta-1}e^{-t}dt$ is the usual Gamma function.

Recall that $\Gamma(z+1) = z\Gamma(z)$ *for any* $z \geq 1$*, and in particular* $\Gamma(m) = (m-1)!$ *for* $m = 1, 2, 3, \ldots$.

(b) Consider the hypotheses

$$H_0 : \theta = 1, \qquad\qquad H_1 : \theta = 2.$$

Find the most powerful test of size α, $0 < \alpha < 1$, for testing H_0 against H_1.

(6 marks)

(c) Compute the power of your test in (b) in terms of α.

What happens to the probability of making a type II error in the two cases $\alpha \to 0$ and $\alpha \to 1$. Is this what you expected?

(5 marks)

(d) Consider now the hypotheses

$$H_0 : \theta = 1, \qquad\qquad H_1' : \theta > 1.$$

Is your test in (b) uniformly most powerful for testing H_0 against H_1'. Carefully explain your answer.

(4 marks)

(Total: 20 marks)

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. observations from the following probability density function:

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x \geq \theta, \\ 0 & \text{if } x < \theta, \end{cases}$$

where $\theta \in \mathbb{R}$.

(a) Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of θ. (3 marks)

(b) Show that the MLE has distribution function $P_\theta(\hat{\theta}_n \leq t) = 1 - e^{-n(t-\theta)}$ for $t \geq \theta$. Hence or otherwise, compute the bias of the MLE of θ. (4 marks)

(c) Obtain the jackknife bias estimate of $\hat{\theta}_n$. Derive the jackknifed estimator of θ based on $\hat{\theta}_n$. Which estimator has smaller bias: the MLE of θ or the jackknifed estimator of θ? Briefly justify your answer.

(8 marks)

Consider now $X_1, \ldots, X_n \sim^{iid}$ Exp(1), i.e. f_θ above with $\theta = 0$. Suppose that $n = 2m$ is even and we are interested in estimating the median γ of the distribution of the $X_i's$. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics and $\gamma_n = \frac{X_{(m)}+X_{(m+1)}}{2}$ be the sample median. Recall that the *jackknife estimator of the variance of* γ_n is

$$v_{jack} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\tilde{\gamma}_i - \frac{1}{n} \sum_{j=1}^{n} \tilde{\gamma}_i \right)^2,$$

where

$$\tilde{\gamma}_i = n\gamma_n - (n-1)\gamma_{(-i)}$$

with $\gamma_{(-i)}$ the sample median based on the observations with the i^{th} observation removed.

(d) Show that

$$v_{jack} = \frac{n-1}{4}(X_{(m+1)} - X_{(m)})^2.$$

(5 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2022

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

MATH60043/MATH70043/MATH97073

Statistical Theory (Solutions)

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

1. (a) (i) seen ⇓

$$EX_1 = \sum_{k=0}^{\infty} k \frac{\theta^k}{k!} e^{-\theta} = \theta e^{-\theta} \sum_{k=1}^{\infty} \frac{\theta^{k-1}}{(k-1)!} = \theta e^{-\theta} e^{\theta} = \theta,$$

$$EX_1(X_1 - 1) = \sum_{k=0}^{\infty} k(k-1) \frac{\theta^k}{k!} e^{-\theta} = \theta^2 e^{-\theta} \sum_{k=2}^{\infty} \frac{\theta^{k-2}}{(k-2)!} = \theta^2.$$

Therefore $\text{Var}(X_1) = EX_1(X_1 - 1) + EX_1 - (EX_1)^2 = \theta^2 + \theta - \theta^2 = \theta$, and they are equal.

4, A

(ii) The log-likelihood satisfies

seen ⇓

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

$$\ell_n'(\theta) = \frac{1}{\theta} \sum x_i - n,$$

$$\ell_n''(\theta) = -\frac{1}{\theta^2} \sum x_i.$$

We see that $\ell_n''(\theta) < 0$ for all $\theta > 0$ as long as $\sum x_i > 0$, in which case the MLE is given by solving $\ell_n'(\theta) = 0$, i.e. $\hat{\theta}_n = \frac{1}{n} \sum x_i = \bar{x}_n$. There is no MLE if $\sum x_i = 0$.

For any $\theta > 0$, $P_\theta(\sum_{i=1}^n X_i = 0) = (1 - e^{-\theta})^n \to 0$ as $n \to \infty$, so the MLE exists and equals $\bar{X}_n$ with probability tending to 1. By the central limit theorem,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(\bar{X}_n - E_\theta X_1) \to^d N(0, \text{Var}_\theta(X_1)) = N(0, \theta)$$

as $n \to \infty$.

Note: do not penalize the students here if they forget the case $\sum X_i = 0$.

4, A

(b) Clearly $f(k) \in [0, 1]$ for all $k = 0, 1, 2, \ldots$. Then

meth seen ⇓

$$\sum_{k=0}^{\infty} f(k) = (1 - p) + p \sum_{k=0}^{\infty} \frac{\phi^k}{k!} e^{-k} = 1 - p + p = 1.$$

Using the same arguments as above, $Y \sim f_{p,\phi}$ satisfies $EY = p\phi$ and $EY(Y-1) = p\phi^2$, and hence

$$\text{Var}(Y) = EY(Y-1) + EY - (EY)^2 = p\phi^2 + p\phi - p^2\phi_i^2 = p\phi(1 + (1-p)\phi).$$

Comparing the mean to the variance, we have $\text{Var}(Y) > EY$ if and only if

$$\begin{aligned} \text{Var}(Y) = p\phi(1 + (1-p)\phi) &> p\phi = EY \\ \iff 1 + (1-p)\phi &> 1 \\ \iff 1 - p &> 0, \end{aligned}$$

since $\phi > 0$, i.e. we have overdispersion as soon as $p < 1$.

3, B

(c) Letting $N = \#\{i : X_i \neq 0\}$, the log-likelihood equals

unseen ⇓

$$\begin{aligned} \ell_n(p) &= \log \left(\prod_{i:X_i=0} [1 - p(1 - e^{-\phi})] \prod_{i:X_i \neq 0} p \frac{\phi^{x_i}}{x_i!} e^{-\phi} \right) \\ &= (n - N) \log(1 - p(1 - e^{-\phi})) + N \log p + C \end{aligned}$$

for some C not depending on p. Differentiating,

$$\begin{aligned}\ell_n'(p) &= -(n-N)\frac{1-e^{-\phi}}{1-p(1-e^{-\phi})} + \frac{N}{p} \\ &= \frac{1}{p(1-p(1-e^{-\phi}))}\left[N(1-p(1-e^{-\phi})) - (n-N)p(1-e^{-\phi})\right] \\ &= \frac{1}{p(1-p(1-e^{-\phi}))}\left[N - np(1-e^{-\phi})\right].\end{aligned}$$

Since the denominator is positive for all $p \in [0,1]$, we see that $\boxed{2, \mathrm{B}}$

$$\begin{aligned} &\ell_n'(p) \geq 0 \\ \iff &N - np(1-e^{-\phi}) \geq 0 \\ \iff &\frac{N}{n(1-e^{-\phi})} \geq p,\end{aligned}$$

with equality at $\hat{p}_n = \frac{N}{n(1-e^{-\phi})}$. Thus $\hat{p}_n$ is the MLE. $\boxed{3, \mathrm{C}}$

(d) $H_0 : p = 1$. We can rewrite the MLE as $\boxed{\text{unseen} \Downarrow}$

$$\hat{p}_n = \frac{1}{n(1-e^{-\phi})}\sum_{i=1}^{n} W_i$$ $\boxed{1, \mathrm{A}}$

for $W_i = 1\{X_i \neq 0\} \sim^{iid} \mathsf{Bin}(1, P_{H_0}(X_i \neq 0) = 1 - e^{-\phi})$ under H_0. We have $EW_i = 1 - e^{-\phi}$ and $\mathsf{Var}(W_i) = e^{-\phi}(1-e^{-\phi})$, so by the central limit theorem,

$$\sqrt{n}(\hat{p}_n - 1) \to^d N\left(0, \frac{e^{-\phi}(1-e^{-\phi})}{(1-e^{-\phi})^2}\right) = N\left(0, \frac{e^{-\phi}}{1-e^{-\phi}}\right).$$

$\boxed{3, \mathrm{D}}$

2. (a) meth seen ⇓

$$EX_1 = \int_0^\infty \lambda x e^{-\lambda x} dx = [-xe^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$$
$$EX_1^2 = \int_0^\infty \lambda x^2 e^{-\lambda x} dx = [-x^2 e^{-\lambda x}]_0^\infty + 2\int_0^\infty x e^{-\lambda x} dx = 2/\lambda^2,$$

using the first integral in the last step. Thus $\mathrm{Var}(X_1) = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$. 4, A

(b) The log-likelihood satisfies seen ⇓

$$\ell_n(\lambda) = \log\left(\lambda^n e^{-\lambda \sum x_i}\right) = n\log\lambda - \lambda \sum_i x_i$$
$$\ell_n'(\lambda) = \frac{n}{\lambda} - \sum_i x_i$$
$$\ell_n''(\lambda) = -\frac{n}{\lambda^2} < 0,$$

so that the global maximum is found by solving $\ell_n'(\lambda) = 0$. Thus the MLE is $\hat{\lambda}_n = n/\sum x_i = 1/\bar{x}_n$. 2, A

We the weak law of large numbers, $\bar{X}_n \to^p E_\lambda X_1 = 1/\lambda$ and thus by the continuous mapping theorem, $1/\bar{X}_n \to^p \lambda$, i.e. $\hat{\lambda}_n$ is consistent. By the central limit theorem, $\sqrt{n}(\bar{X}_n - 1/\lambda) \to^d N(0, 1/\lambda^2)$ as $n \to \infty$. Setting $g(t) = 1/t$, so that $g'(t) = -1/t^2$, the delta method yields

$$\sqrt{n}(g(\bar{X}_n) - g(1/\lambda)) = \sqrt{n}(\hat{\lambda}_n - \lambda) \to^d N(0, g'(1/\lambda)^2\lambda^{-2}) = N(0, \lambda^2).$$

2, B

The limiting distribution implies sim. seen ⇓

$$P_\lambda\left(-1.96\lambda \leq \sqrt{n}(\hat{\lambda}_n - \lambda) \leq 1.96\lambda\right) \to 0.95.$$

We estimate the standard deviation by $\hat{\lambda}_n$ since the error is of smaller order after rescaling by $1/\sqrt{n}$ (it is enough just to state the estimation by $\hat{\lambda}_n$). Inverting the inequalities in the probability yield

$$P_\lambda\left(\hat{\lambda}_n - \frac{1.96}{\sqrt{n}}\hat{\lambda}_n \leq \lambda \leq \hat{\lambda}_n + \frac{1.96}{\sqrt{n}}\hat{\lambda}_n\right),$$

which gives our 95% confidence interval. 3, B

(c) Again by the delta method, unseen ⇓

$$\sqrt{n}(h(\hat{\lambda}_n) - h(\lambda)) \to^d N(0, h'(\lambda)^2\lambda^2)$$

as $n \to \infty$. Therefore we want h such that $h'(\lambda)\lambda = c > 0$, i.e. $h'(\lambda) = c\lambda^{-1}$. This gives $h(\lambda) = c\log\lambda + d$ for any $c > 0$ and $d \in \mathbb{R}$ (any choice is fine, e.g. $h(\lambda) = \log\lambda$).

Setting $c = 1$ and $d = 0$, $\sqrt{n}(\log\hat{\lambda}_n - \log\lambda) \to^d N(0,1)$. Therefore

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

giving our confidence interval for $h(\lambda) = \log\lambda$. Since $t \mapsto \log t$ is strictly increasing, we can invert this to get

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

and hence our confidence interval is

$$\left[\hat{\lambda}_n e^{-1.96n^{-1/2}}, \hat{\lambda}_n e^{1.96n^{-1/2}}\right].$$

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

For n large, the width of this confidence interval is by Taylor expansion:

$$\begin{aligned}\hat{\lambda}_n(e^{1.96/\sqrt{n}} - e^{-1.96/\sqrt{n}}) &= \hat{\lambda}_n(1 + 1.96n^{-1/2} - 1 + 1.96n^{-1/2} + O(n^{-1})) \\ &= \hat{\lambda}_n(2 \times 1.96n^{-1/2} + O(n^{-1}),\end{aligned}$$

which matches the length of that derived in (b) up to the smaller order terms. $\boxed{2, \text{D}}$

(d) Returning to our previous expression, we see that $\ell_n'(\lambda) \ge 0$ if and only if $\lambda \le 1/\bar{X}_n$. $\boxed{\text{meth seen} \Downarrow}$
Thus the likelihood is decreasing after $1/\bar{X}_n$. So if $1/\bar{X}_n < \delta$, we select the smallest value of λ in our restricted parameter space, i.e. δ. Therefore,

$$\hat{\lambda}_n^\delta = \max(1/\bar{X}_n, \delta) = \begin{cases} 1/\bar{X}_n & \text{if } 1/\bar{X}_n \ge \delta, \\ \delta & \text{otherwise.}\end{cases}$$

Since δ lies on the boundary of our parameter space, we cannot necessarily expect the MLE to be normally distributed. $\boxed{3, \text{D}}$

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) seen/sim.seen ⇓

$$\begin{aligned}
\pi(\theta|x) \propto f_\theta(x)\pi(\theta) &\propto \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)\exp\left(-\frac{\theta^2}{2\tau^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2 + \frac{x}{\sigma^2}\theta\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)\left(\theta - \frac{x/\sigma^2}{1/\sigma^2+1/\tau^2}\right)^2\right),
\end{aligned}$$

which is the form of a

$$N\left(\frac{\tau^2 x}{\sigma^2+\tau^2}, \frac{\tau^2\sigma^2}{\sigma^2+\tau^2}\right)$$

distribution, as desired. 3, A

(c) We obtain the Bayes rule by minimizing the posterior risk. Writing $\theta|X \sim N(m, \nu^2)$ for the quantities m, ν^2 derived in (b), for fixed observation $x \in \mathbb{R}$, meth seen ⇓

$$\begin{aligned}
E_\pi[L(\delta(x), \theta)|x] &= E_\pi[e^{-\lambda\delta(x)\theta}|x] \\
&= e^{-\lambda\delta(x)m + \frac{1}{2}\lambda^2\delta(x)^2\nu^2}.
\end{aligned}$$

The exponent is a quadratic in $\delta(x)$, so it can be minimized by finding the stationary point (do not penalize if candidates simply differentiate at set to 0). Differentiating this expression with respect to $\delta(x)$, gives

$$\left[-\lambda m + \lambda^2\delta(x)\nu^2\right] e^{-\lambda\delta(x)m + \frac{1}{2}\lambda^2\delta(x)^2\nu^2} = 0$$

which implies

$$\delta(x) = \frac{m}{\lambda\nu^2} = \frac{1}{\lambda}\frac{\tau^2 x}{\sigma^2+\tau^2}\frac{\sigma^2+\tau^2}{\sigma^2\tau^2} = \frac{x}{\lambda\sigma^2}.$$

This is the unique minimizer of the posterior risk and hence also the unique Bayes rule. 4, C

(d) The risk equals unseen ⇓

$$\begin{aligned}
R(\delta_0(x), \theta) &= E_\theta e^{-\lambda\delta_0(X)\theta} \\
&= E_\theta e^{-\frac{\theta}{\sigma^2}X} \\
&= e^{-\theta^2/\sigma^2 + \theta^2\sigma^2/(2\sigma^4)} = e^{-\frac{\theta^2}{2\sigma^2}}.
\end{aligned}$$

This is maximized at $\theta = 0$, yielding maximal risk $\sup_{\theta\in\mathbb{R}} R(\delta_0, \theta) = \exp(\mu^2\sigma^2/2)$.

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

(e) Using the formula for the risk derived above and considering the prior $\theta \sim N(0, \tau^2)$, the Bayes risk equals:

$\boxed{\text{unseen} \Downarrow}$

$$
\begin{aligned}
R_\pi(\delta_0) = E_\pi[R(\delta_0, \theta)] &= E_\pi[e^{-\frac{1}{2\sigma^2}\theta^2}] \\
&= \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}\theta^2} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{\theta^2}{2\tau^2}} \, d\theta \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2}(1/\tau^2 + 1/\sigma^2)\theta^2} \, d\theta.
\end{aligned}
$$

Recognizing the integrand has the form of a $N(0, (1/\tau^2 + 1/\sigma^2)^{-1})$ density function up to the normalizing constant, we get

$$
R_\pi(\delta_0) = \sqrt{\frac{(1/\tau^2 + 1/\sigma^2)^{-1}}{\tau^2}} = \sqrt{\frac{\sigma^2}{\sigma^2 + \tau^2}}.
$$

Letting $\tau \to 0$, we get

$$
R_{N(0,\tau^2)}(\delta_0) \to \sqrt{\frac{\sigma^2}{\sigma^2}} = 1 = \sup_{\theta\in\mathbb{R}} R(\delta_0, \theta),
$$

which corresponds to a point mass prior $P(\theta = 0) = 1$. Note since the Bayes rule does not depend on τ here, δ_0 will again be Bayes for this point mass prior. From a result in the notes, if the Bayes risk of a Bayes decision rule equals the maximal risk over the parameter space, then the decision rule is minimax. Hence δ_0 is minimax.

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

4. (a) 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. $\pi_\phi(\theta) = P_\theta(\text{reject } H_0)$. $\boxed{\text{seen} \Downarrow}$

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.

The *critical region* of a test ϕ is the region R where we reject the null hypothesis if the data (or statistic) falls in R, i.e. $X \in R$.

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 ($\sup_{\theta\in\Theta_0} \pi_\phi(\theta) \leq \alpha$) and (ii) any other level α test ϕ^* has smaller power, i.e. $\pi_{\phi^*}(\theta) \leq \pi_\phi(\theta)$ for all $\theta \in \Theta_1$. $\boxed{5, \text{A}}$

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

$$\frac{f_2(x)}{f_1(x)} = \frac{\Gamma(1)}{\Gamma(2)}|x| = |x|.$$

Thus the UMP test takes the form reject H_0 if and only if $|x| \geq k$ where k is determined by the probability of making a type I error. In particular, by symmetry of the distributions about 0, $\boxed{2, \text{A}}$

$$\begin{aligned} P_{\theta=1}(\text{reject } H_0) &= P_1(|x| \geq k) \\ &= 2\int_k^\infty \frac{1}{2}\frac{1}{\Gamma(1)} e^{-|x|}dx \\ &= \left[-e^{-x}\right]_k^\infty = e^{-k} = \alpha, \end{aligned}$$

yielding $k = -\log\alpha = \log(1/\alpha)$. Thus the UMP test is to reject H_0 if and only if $|X| \geq \log(1/\alpha)$. $\boxed{4, \text{B}}$

(c) The power of the test is the probability of rejecting the null hypothesis under the alternative, namely $\boxed{\text{seen/sim.seen} \Downarrow}$

$$\begin{aligned} P_{H_1}(\text{reject } H_0) = P_2(|X| \geq k) &= 2\int_k^\infty \frac{1}{2\Gamma(2)}|x|e^{-|x|}dx \\ &= \int_k^\infty xe^{-x}dx \\ &= \left[-xe^{-x}\right]_k^\infty + \int_k^\infty e^{-x}dx \\ &= ke^{-k} + e^{-k} = \alpha + \alpha\log(1/\alpha). \end{aligned}$$

Your probability of making a type II error is 1 minus the power, i.e.

$$P_{H_1}(\text{do not reject } H_0) = 1 - \alpha - \alpha\log(1/\alpha)).$$

As $\alpha \to 0$, this tends to 1, while as $\alpha \to 1$, this tends to 0. This is what is expected, since when $\alpha \to 0$, we have little tolerance for falsely rejecting H_0 and thus make it hard to reject, even when H_1 is true. Conversely when $\alpha \to 1$, we are happy to falsely reject H_0, also when H_1 is true. $\boxed{5, \text{B}}$

(d) For $\theta_2 > \theta_1$, we have

meth seen $\Downarrow$

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{\Gamma(\theta_1)}{\Gamma(\theta_2)}|x|^{\theta_2-\theta_1},$$

which is an increasing function of $|x|$. Thus the family of distributions being considered has *monotone likelihood ratio*. By the Karlin-Rubin theorem, the likelihood ratio test (which we just derived above) is uniformly most powerful for testing H_0 against H_1'.

4, D

5. (a) The likelihood equals $\boxed{\text{seen} \Downarrow}$

$$L_n(\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} 1\{x_i \geq \theta\} = e^{n\theta - \sum x_i} 1\{\min x_i \geq \theta\}.$$

Since the function $\theta \mapsto e^{n\theta}$ is increasing, the likelihood is maximized by the largest value of θ such that the indicator is 1, i.e. $\hat{\theta}_n = \min x_i$. $\boxed{3, \mathrm{M}}$

(b) For $t \geq \theta$, $\boxed{\text{seen} \Downarrow}$

$$P_\theta(\hat{\theta}_n \geq t) = P_\theta(\min X_i \geq t) = P_\theta(X_1 \geq t)^n = e^{-n(t-\theta)}.$$

Differentiating yields the density function $ne^{-n(t-\theta)} 1_{[\theta,\infty)}(t)$ for $\hat{\theta}_n = \min X_i$, and hence integrating by parts

$$\begin{aligned} E_\theta \hat{\theta}_n &= \int_\theta^\infty tne^{n(\theta-t)} dt \\ &= \Big[-te^{n(\theta-t)}\Big]_\theta^\infty + \int_\theta^\infty e^{n(\theta-t)} dt \\ &= \theta + \Big[-\tfrac{1}{n} e^{n(\theta-t)}\Big]_\theta^\infty = \theta + \tfrac{1}{n}. \end{aligned}$$

Hence the bias is $b(\theta) = E_\theta \hat{\theta}_n - \theta = 1/n > 0$. $\boxed{4, \mathrm{M}}$

(c) By removing the i^{th} observation, we have $\hat{\theta}_{(-i)} = \min_{j \neq i} X_j$, which takes the value $X_{(1)} = \min X_i$ $n-1$ times and $X_{(2)}$ once (when $X_i = \min_j X_j$). [Note that ties occur with probability zero]. Thus our averaged version is $\boxed{\text{unseen} \Downarrow}$

$$\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(-i)} = \frac{1}{n} \left((n-1) X_{(1)} + X_{(2)} \right) = \frac{n-1}{n} X_{(1)} + \frac{1}{n} X_{(2)}.$$

The jackknife bias estimate is then

$$b_{jack} = (n-1)(\bar{\theta}_n - \hat{\theta}_n) = (n-1) \left[\frac{n-1}{n} X_{(1)} + \frac{1}{n} X_{(2)} - X_{(1)} \right] = \frac{n-1}{n} \left[X_{(2)} - X_{(1)} \right]$$

$\boxed{4, \mathrm{M}}$

The bias-corrected estimator is $\boxed{\text{unseen} \Downarrow}$

$$\theta_{jack} = \hat{\theta}_n - b_{jack} = X_{(1)} - \frac{n-1}{n} \left[X_{(2)} - X_{(1)} \right].$$

$\boxed{2, \mathrm{M}}$

The jackknifed estimator has smaller bias since it is a bias-corrected version of the MLE $\hat{\theta}_n$. We saw in (b) that the bias of the MLE equals $1/n$. We know from the textbook that if the bias of $\hat{\theta}_n$ takes the form $\frac{a}{n} + \frac{b}{n^2} + O(1/n^3)$, then

$$E b_{jack} = \text{bias}(\hat{\theta}_n) + O(1/n^2).$$

Hence the jackknifed estimator will have bias

$$E_\theta \hat{\theta}_n - E_\theta b_{jack} - \theta = O(1/n^2),$$

which is smaller than that of $\hat{\theta}_n$. (Any brief justification is sufficient). $\boxed{2, \mathrm{M}}$

(d) $\boxed{\text{unseen} \Downarrow}$

$$
\begin{aligned}
v_{jack} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\tilde{\gamma}_i - \frac{1}{n} \sum_{j=1}^{n} \tilde{\gamma}_i \right)^2 \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(n\gamma_n - (n-1)\gamma_{(-i)} - \frac{1}{n} \sum_{j=1}^{n} [n\gamma_n - (n-1)\gamma_{(-j)}] \right)^2 \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left((n-1)\gamma_{(-i)} + \frac{1}{n} \sum_{j=1}^{n} (n-1)\gamma_{(-j)} \right)^2 .
\end{aligned}
$$

With probability one, there are no ties between observations. Since $n = 2m$ is even, the medians $\gamma_{(-j)}$ are equal to $X_{(m)}$ or $X_{(m+1)}$, each occurring $n/2 = m$ times. Thus

$$
\sum_{j=1}^{n} \gamma_{(-j)} = \frac{n(X_{(m)} + X_{(m+1)})}{2} = m(X_{(m)} + X_{(m+1)}).
$$

Therefore,

$$
v_{jack} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\gamma_{(-i)} + \frac{X_{(m)} + X_{(m+1)}}{2} \right)^2 .
$$

Again using the values of $\gamma_{(-i)}$,

$$
\begin{aligned}
v_{jack} &= \frac{n-1}{n} \left\{ \frac{n}{2} \left(X_{(m)} - \frac{X_{(m)} + X_{(m+1)}}{2} \right)^2 + \frac{n}{2} \left(X_{(m+1)} - \frac{X_{(m)} + X_{(m+1)}}{2} \right)^2 \right\} \\
&= (n-1) \left(\frac{X_{(m+1)} - X_{(m)}}{2} \right)^2
\end{aligned}
$$

as required. $\boxed{5, \mathrm{M}}$

Review of mark distribution:
Total A marks: 31 of 32 marks
Total B marks: 19 of 20 marks
Total C marks: 14 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.
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 separate pdf file with your email.

ExamModuleCode	QuestionNumber	Comments for Students
Statistical Theory_MATH60043 MATH97073 MATH70043	1	This question was generally well-answered. Several candidates forget to justify why their proposed MLE was actually a maximum.
Statistical Theory_MATH60043 MATH97073 MATH70043	2	The first parts were generally well-answered. Few candidates thought to Taylor expand the limits of the confidence intervals in (c) .
Statistical Theory_MATH60043 MATH97073 MATH70043	3	This question caused the most difficulty. In (b) several candidates just wrote the answer without showing any working. Generally candidates who were clear on the basic definitions involved did best.
Statistical Theory_MATH60043 MATH97073 MATH70043	4	Several candidates got confused between the notions of power and probability of type II error. In (d), many candidates did not provide a clear explanation of why the model has monotone likelihood ratio.
Statistical Theory_MATH60043 MATH97073 MATH70043	5	Parts (a)-(b) were generally well answered. For (c)-(d), several candidates did not write clear solutions, which then seemed to confuse them later on. Candidates who clearly wrote down at the start what their estimator with one observation removed equaled tended to make significant progress with these parts.