IMPERIAL

MATH70043

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
Summer 2025

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

Statistical Theory

Date: Tuesday, May 27, 2025

Time: Start time 10:00 – End time 12:30 (BST)

Time Allowed: 2.5 hours

This paper has 5 Questions.

Please Answer Each Question in a Separate Answer Booklet

This is a closed book examination.

Candidates should start their solutions 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.

Credit will be given for all questions attempted.

Allow margins for marking.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO DO SO

1. (a) Let $X_1, \ldots, X_n \sim^{iid} f_\theta$ denote independent and identically distributed (i.i.d.) random variables with f_θ a probability density function coming from a statistical model $\{f_\theta : \theta \in \Theta\}$ satisfying the usual regularity conditions.

If $\hat{\theta}_n$ is the maximum likelihood estimator (MLE) and $\theta_0 \in \Theta$ is the true parameter, write down the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ as $n \to \infty$. You should define any quantities involved in the limiting distribution. (4 marks)

Suppose now that $X_1, \ldots, X_n \sim^{iid} U[-\theta, \theta]$ are i.i.d. uniform random variables on $[-\theta, \theta]$ with $\theta > 0$.

(b) Compute the maximum likelihood estimator (MLE) $\hat{\theta}_n$ for θ. (4 marks)

(c) Let θ_0 denote the true parameter value. Compute the cumulative distribution function of $\hat{\theta}_n$. Hence or otherwise, show that $n(\theta_0 - \hat{\theta}_n) \to^d Y$ for a random variable Y whose distribution you should identify. (6 marks)

(d) Does it occur that $\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d Z$ for a random variable Z with positive variance? Justify your answer. (2 marks)

(e) What is the limiting distribution of $n(\hat{\theta}_n^2 - \theta_0^2)$? Justify your answer. (4 marks)

(Total: 20 marks)

2. Let $X_1, \dots, X_n$ be i.i.d. random variables having exponential distribution with density function $f_\lambda(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$. Recall that $T_n = \sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$, where the Gamma distribution has density

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda x}, \qquad x > 0.$$

(a) Show that $T_n = \sum_{i=1}^n X_i$ is *minimal sufficient* and *complete* for λ. (6 marks)

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

(b) It is desired to estimate the quantity $\theta = P_\lambda(X_1 > 1) = e^{-\lambda}$. Compute the Fisher information $I(\theta)$ of X_1 with respect to the parameter θ. (4 marks)

(c) State the Lehmann-Scheffé theorem. (2 marks)

(d) By considering the estimator $\tilde{\theta}_n = 1\{X_1 > 1\}$ of θ or otherwise, show that

$$\hat{\theta}_n = \begin{cases} 0, & \text{if } T_n < 1, \\ \left(1 - \frac{1}{T_n}\right)^{n-1} & \text{if } T_n \geq 1 \end{cases}$$

is the minimum variance unbiased estimator of θ based on $X_1, \dots, X_n$. (8 marks)

You may find it useful to find the conditional density of X_1 given T_n, which can be found via Bayes formula:

$$f_{X_1|T_n}(x|t) = \begin{cases} \frac{f_{T_n|X_1}(t|x) f_{X_1}(x)}{f_{T_n}(t)} & \textit{if } 0 \leq x \leq t, \\ 0 & \textit{otherwise.} \end{cases}$$

(Total: 20 marks)

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

(5 marks)

(b) Show that if $X \sim U(0,1)$, then $-\log X \sim$ Exp(1).
Recall that the Exp(λ) *distribution has density* $\lambda e^{-\lambda x}$ *for* $x > 0$. (2 marks)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with density $f_\beta(x) = (\beta+1)x^\beta 1_{(0,1)}(x)$ with $\beta \geq 0$.

Recall that if $V_1, \ldots, V_n \sim^{iid}$ Exp(λ), then $\sum_{i=1}^n V_i \sim \Gamma(n, \lambda)$, where the Gamma distribution has density $\frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y}$ for $y > 0$. For $0 < \alpha < 1$, define the corresponding quantile $\gamma_{n,\lambda}(\alpha)$ of the Gamma distribution by:

$$P\left(\Gamma(n,\lambda) \leq \gamma_{n,\lambda}(\alpha)\right) = \alpha.$$

(c) Find the most powerful test of size $\alpha \in (0,1)$ for testing $H_0 : \beta = 0$ against $H_1 : \beta = 1$. The critical region should be defined in terms of the quantile of a gamma distribution which you should specify.

(6 marks)

(d) Show that the power of your test in (c) can be written in the form

$$P\left(\Gamma(n,\lambda) \leq \gamma_{n,\lambda'}(\alpha)\right)$$

for suitable λ and λ' that you should specify. (3 marks)

(e) Construct a uniformly most powerful test of size α for $H_0 : \beta = 0$ against $H_1' : \beta > 0$.

(4 marks)

(Total: 20 marks)

4. (a) In the context of decision theory, explain the meaning of the following terms: *loss function*, 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.

(5 marks)

Suppose $X_1, \dots, X_n \sim^{iid} f_\theta$, where $\theta > 0$ is a positive parameter. Consider estimation of θ with the loss function

$$L(a, \theta) = \frac{a}{\theta} - 1 - \log \frac{a}{\theta}.$$

(b) Assign to θ an arbitrary prior with probability density function π on $(0, \infty)$. For the loss L above, show that the corresponding Bayes rule is given by

$$\tilde{\theta}_n = \frac{1}{E^\Pi[1/\theta | X_1, \dots, X_n]},$$

where $E^\Pi[1/\theta | X_1, \dots, X_n] = \int \frac{1}{\theta} \pi(\theta | X_1, \dots, X_n) d\theta$. (5 marks)

Suppose now $X_1, \dots, X_n \sim^{iid}$ Ber(θ), $0 < \theta < 1$, are Bernoulli random variables and assign θ a Beta(a, b) prior, where a, b are positive integers.

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

$$\frac{(a + b - 1)!}{(a - 1)!(b - 1)!} y^{a-1} (1 - y)^{b-1}, \qquad 0 < y < 1,$$

mean $\frac{a}{a+b}$ and variance $\frac{ab}{(a+b)^2(a+b+1)}$.

(c) Compute the resulting posterior distribution for θ based on a Beta(a, b) prior. (3 marks)

(d) Compute the Bayes rule $\tilde{\theta}_n$ for loss L for a Beta(a, b) prior. (3 marks)

(e) Assume now $X_1, \dots, X_n \sim^{iid}$ Ber(θ_0) for some true $\theta_0 \in (0, 1)$. Derive the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ as $n \to \infty$.

Hint: you may find it useful to consider the difference $\sqrt{n}(\tilde{\theta}_n - \bar{X}_n)$, where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

(4 marks)

(Total: 20 marks)

5. (a) Given independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_n$, explain the notion of a *bootstrap sample* $X_1^*, \ldots, X_n^*$. (3 marks)

Consider now the statistical model $X_1, \ldots, X_n \sim^{iid} N(\mu, 1)$, but with **restricted** parameter space $\mu \in [0, \infty)$, and let μ_0 denote the true parameter.

(b) Show that the MLE equals $\hat{\mu}_n = \max(\bar{X}_n, 0)$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Compute the asymptotic distribution of $\sqrt{n}(\hat{\mu}_n - \mu_0)$ when $\mu_0 > 0$.
When $\mu_0 = 0$, show that $\sqrt{n}(\hat{\mu}_n - \mu_0) \to^d W_0$ for a random variable W_0, whose distribution you should specify. (7 marks)

Our goal is to approximate the distribution of the quantity

$$R_n(X, \mu) = \sqrt{n}(\hat{\mu}_n - \mu_0)$$

using the bootstrap.

(c) Let $X_1^*, \ldots, X_n^*$ be a bootstrap sample based on $X_1, \ldots, X_n$ and $\hat{\mu}_n^* = \max(\bar{X}_n^*, 0)$, where $\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*$. Explain why

$$R_n(X^*, \hat{\mu}_n) = \sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n)$$

is a bootstrap realization of $R_n(X, \mu)$. (2 marks)

(d) Let $-t < x < 0$ for some $t > 0$. For $\mathbb{P}$ the joint distribution of $X_1, \ldots, X_n, X_1^*, \ldots, X_n^*$, show that

$$\mathbb{P}\left(\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) \leq x | \sqrt{n}\bar{X}_n > t\right) \geq \mathbb{P}\left(\max\{\sqrt{n}(\bar{X}_n^* - \bar{X}_n), -t\} \leq x | \sqrt{n}\bar{X}_n > t\right). \quad (1)$$

By taking the limit $n \to \infty$ on the right-hand side of (1), deduce that if $\mu_0 = 0$, then for n large,

$$\mathbb{P}\left(\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) \leq x | \sqrt{n}\bar{X}_n > t\right) > \mathbb{P}(W_0 \leq x),$$

where W_0 is the random variable from (b).
[You may assume without proof that $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)|X_1, \ldots, X_n \to^d N(0, 1)$ as $n \to \infty$ for all sequences $X_1, X_2, \ldots$ in a set of $\mathbb{P}$-probability one.] (6 marks)

(e) Based on this last result, is the bootstrap quantity $R_n(X^*, \hat{\mu}_n)$ a good distributional approximation for $R_n(X, \mu) = \sqrt{n}(\hat{\mu}_n - \mu_0)$? Justify your answer. (2 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2025

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) The limiting distribution is $\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N(0, I(\theta_0)^{-1})$, where seen $\Downarrow$

$$I(\theta)_{ij} = E_\theta\left[\frac{\partial}{\partial\theta_i}\log f_\theta(X_1)\frac{\partial}{\partial\theta_j}\log f_\theta(X_1)\right]$$

is the Fisher information [full marks for giving the formula correctly in one-dimension]. 4, A

(b) The likelihood equals sim. seen $\Downarrow$

$$\prod_{i=1}^n \frac{1}{2\theta}1_{[-\theta,\theta]}(x_i) = \frac{1}{(2\theta)^n}\prod_{i=1}^n 1\{|x_i| \leq \theta\} = \frac{1}{(2\theta)^2}1\{\max_i |x_i| \leq \theta\}.$$

This is a decreasing function of θ while the indicator function is equal to 1, hence the MLE is the smallest value of θ such that the indicator remains 1, i.e. $\hat{\theta}_n = \max_i |X_i|$. 4, A

(c) For $t \in [0, \theta]$, the CDF is sim. seen $\Downarrow$

$$P_{\theta_0}\left(\hat{\theta}_n \leq t\right) = P_{\theta_0}\left(\max_i |X_i| \leq t\right) = P_{\theta_0}(|X_i| \leq t)^n$$
$$\left(\frac{2t}{2\theta_0}\right)^n = (t/\theta_0)^n,$$

using $X_1, \ldots, X_n$ are iid. Therefore, for $t > 0$, 2, B

$$\begin{aligned}
P_{\theta_0}(n(\theta_0 - \hat{\theta}_n) \leq t) &= P_{\theta_0}\left(\theta_0 - t/n \leq \hat{\theta}_n\right) \\
&= 1 - P_{\theta_0}\left(\hat{\theta}_n \leq \theta_0 - t/n\right) \\
&= 1 - \left(\frac{\theta_0 - t/n}{\theta_0}\right)^n \\
&= 1 - \left(1 - \frac{t}{n\theta_0}\right)^n \to 1 - e^{-t/\theta_0}
\end{aligned}$$

as $n \to \infty$. We see that the limiting distribution is an exponential distribution with mean θ_0, i.e. with pdf $\frac{1}{\theta_0}e^{-t/\theta_0}$. 4, B

(d) No, this cannot be the case. By (c), the asymptotic variance of $\hat{\theta}_n$ is $1/n^2$ (since Y is non-degenerate), whereas if Z had positive variance, this would mean the asymptotic variance is like $1/n$, which is a contradiction [any reasonable justification is fine]. sim. seen $\Downarrow$ 2, B

(e) Expanding out the difference of squares, unseen $\Downarrow$

$$n(\hat{\theta}_n^2 - \theta_0^2) = n(\hat{\theta}_n - \theta_0)(\hat{\theta}_n + \theta_0).$$

The first term converges in distribution to Y by (c). Using the CDF in (c), for any $0 < \epsilon < \theta_0$,

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| \geq \epsilon) = P_{\theta_0}(\hat{\theta}_n \leq \theta_0 - \epsilon) = (1 - \epsilon/\theta_0)^n \to 0,$$

i.e. $\hat{\theta}_n \to^p \theta_0$ (any reasonable justification is fine). Thus $\hat{\theta}_n + \theta_0 \to^p 2\theta_0$ and hence by Slutsky's theorem,

$$n(\hat{\theta}_n^2 - \theta_0^2) \to^d 2\theta_0 Y.$$

4, C

2. (a) Consider the ratio of pdfs

meth seen ⇓

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

This ratio does not depend on λ if and only if $\sum_i x_i = \sum_i x_i'$ and hence T_n is minimal sufficient for λ by a theorem in the notes.

3, A

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

$$E_\theta g(T) = \int_0^\infty g(t) \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt = 0, \qquad \forall \theta > 0.$$

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), $h(t) = 0$ for all $t > 0$, and hence also $g(t) = 0$. Thus T_n is complete.

3, A

An alternative approach is to argue using the general completeness result for exponential families.

sim. seen ⇓

(b) Since $\theta = e^{-\lambda}$, we have $\lambda = -\log\theta$. Thus the log-likelihood equals

$$\ell(\theta) = \log f_\lambda(x) = \log\lambda - \lambda x = \log(-\log\theta) + x\log\theta.$$

Differentiating twice with respect to θ,

$$\begin{aligned}
\ell'(\theta) &= \frac{1}{\theta\log\theta} + \frac{x}{\theta}, \\
\ell''(\theta) &= -\frac{1}{\theta^2\log\theta} - \frac{1}{\theta^2(\log\theta)^2} - \frac{x}{\theta^2} \\
&= \frac{-1}{\theta^2}\left[\frac{1}{\log\theta} + \frac{1}{(\log\theta)^2} + y\right].
\end{aligned}$$

Since $EY = 1/\lambda = -1/(\log\theta)$, we have

$$\begin{aligned}
I(\theta) = -E_\theta \ell''(\theta; Y) &= \frac{1}{\theta^2}\left[\frac{1}{\log\theta} + \frac{1}{(\log\theta)^2} - \frac{1}{\log\theta}\right] \\
&= \frac{1}{\theta^2(\log\theta)^2}.
\end{aligned}$$

4, B

(c)

seen ⇓

Theorem 1 (Lehmann-Scheffé). *Let T be a sufficient and complete statistic for θ and $\tilde{g}$ be an unbiased estimator of $g(\theta)$ with $var_\theta(\tilde{g}) < \infty$ for all $\theta \in \Theta$. If $\hat{g}(T(X)) = E[\tilde{g}(X)|T(X)]$, then $\hat{g}$ is the unique uniformly minimum variance unbiased estimator (UMVUE) of $g(\theta)$.*

[Do not penalize if conditions such as finite variance are missing].

2, A

(d) We have $E_\theta 1\{X_1 > 1\} = P_\theta(X_1 > 1) = \theta$ and hence $\tilde{\theta}_n$ is unbiased for θ. The Lehmann-Scheffé theorem tells us that $\hat{\theta}_n = E[\tilde{\theta}_n|T_n] = P(X_1 > 1|T_n)$ will be the minimum variance unbiased estimator for θ. To work this out, consider the conditional distribution of $X_1|T_n$. Using the hint, if $x \le t$,

unseen ⇓

$$\begin{aligned}
f_{X_1|T_n}(x|t) = \frac{f_{T_n|X_1}(t|x) f_{X_1}(x)}{f_{T_n}(t)} &= \frac{f_{T_{n-1}}(t-x) f_{X_1}(x)}{f_{T_n}(t)} \\
&= \frac{\frac{\lambda^{n-1}}{(n-2)!}(t-x)^{n-2} e^{-\lambda(t-x)} \lambda e^{-\lambda x}}{\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}} \\
&= (n-1)\frac{1}{t}(1 - x/t)^{n-2}.
\end{aligned}$$

If $T_n < 1$, our estimator $\hat{\theta}_n = P(X_1 > 1|T_n) = 0$ trivially since $X_1 \leq T_n$. If $T_n = t > 1$, then

$$\begin{aligned}\hat{\theta}_n = P(X_1 > 1|T_n = t) &= \int_1^\infty f_{X_1|T_n}(x|t)dx \\ &= \frac{n-1}{t}\int_1^t \left(1 - \tfrac{x}{t}\right)^{n-2} dx \\ &= \left[-\left(1 - \tfrac{x}{t}\right)^{n-1}\right]_1^t = \left(1 - \tfrac{1}{t}\right)^{n-1}\end{aligned}$$

as required. $\boxed{8, \mathrm{D}}$

3. (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. $\pi_\phi(\theta) = P_\theta(\text{reject } H_0)$. 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.

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

5, A

(b) For $t > 0$, sim. seen ⇓

$$P(-\log X \le t) = P(e^{-t} \le X) = 1 - e^{-t},$$

which is the CDF of an Exp(1) random variable (alternatively, differentiating this gives the pdf). 2, A

(c) By the Neyman-Pearson lemma, the uniformly most powerful (UMP) test is given by the likelihood ratio test. We have sim. seen ⇓

$$\frac{f_1(x)}{f_0(x)} = \frac{\prod_{i=1}^n 2x_i}{\prod_{i=1}^n 1} = 2^n \prod_{i=1}^n x_i.$$

The likelihood ratio test rejects H_0 if this ratio exceeds some threshold, which is equivalent to the log-ratio $\log(f_1(x)/f_0(x)) = n\log 2 + \sum_{i=1}^n \log x_i$ exceeding a threshold. Thus we reject H_0 if and only if $\sum \log x_i \ge k$ for a suitable k. But

$$P_{H_0}\left(\sum \log X_i \ge k\right) = P_0\left(-k \ge -\sum \log X_i\right).$$

Since $X_i \sim U(0,1)$ under H_0, we have from (b) and the hint that $-\sum \log X_i \sim \Gamma(n,1)$. Thus the above probability equals

$$P\left(\Gamma(n,1) \le -k\right) = \alpha,$$

which gives $-k = \gamma_{n,1}(\alpha)$. We thus reject H_0 if and only if $\sum_{i=1}^n \log X_i \ge k = -\gamma_{n,1}(\alpha)$. 6, B

(d) For $X \sim f_\beta$, we similarly have sim. seen ⇓

$$P(-\log X \le t) = P(e^{-t} \le X) = \int_{e^{-t}}^1 (1+\beta)x^\beta dx = \left[x^{\beta+1}\right]_{e^{-t}}^1 = 1 - e^{-(\beta+1)t},$$

i.e. $-\log X \sim$ Exp$(1+\beta)$. So under H_1, $-\sum_{i=1}^n \log X_i \sim \Gamma(n,2)$, and so the power is

$$P\left(\Gamma(n,2) \le \gamma_{n,1}(\alpha)\right).$$

3, B

(e) Two possible solutions: meth seen ⇓

1. (Karlin-Rubin theorem) For $\beta_2 > \beta_1$,

$$\frac{f_{\beta_2}(x)}{f_{\beta_1}(x)} = \left(\frac{\beta_2+1}{\beta_1+1}\right)^n \left(\prod x_i\right)^{\beta_2-\beta_1},$$

   which is a monotone increasing function of $\prod x_i$. Hence the Karlin-Rubin theorem tells us the uniformly most powerful test takes the form to reject if $\prod x_i \geq k'$, or equivalently if $\sum \log x_i \geq k$, i.e. exactly the test in (c).

2. (Arguing directly). For any simple hypotheses $H_0 : \beta = 0$ and $H_{\beta'} : \beta = \beta'$ with $\beta' > 0$, the Neyman Pearson lemma is optimal with log-likelihood ratio

$$\log \frac{f_{\beta'}(x)}{f_0(x)} = \log(\beta+1)^n \prod x_i = n\log(1+\beta) + \sum \log x_i.$$

   We see that this exceeds a threshold if and only if $\log x_i \geq k$, where k is calibrated by H_0, i.e. we recover the test in (c). Since this is UMP for all alternatives in H_1', it is UMP for H_1. $\boxed{4, \mathrm{D}}$

4. (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 ⇓

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. 5, A

(b) We find the minimizer of the posterior risk, which by a result in the notes gives the minimizer of the Bayes risk. For $\pi(\theta|X)$ the posterior distribution, the posterior risk equals sim. seen ⇓

$$\begin{aligned} E_\pi[L(\delta(x), \theta)|x] &= \int \left(\frac{\delta}{\theta} - 1 - \log \frac{\delta}{\theta} \right) \pi(\theta|x) d\theta \\ &= \delta \int \frac{1}{\theta} \pi(\theta|x) d\theta - (1 + \log \delta) + \int \log \theta \pi(\theta|x) d\theta. \end{aligned}$$

Differentiating with respect to δ,

$$\int \frac{1}{\theta} \pi(\theta|x) d\theta - \frac{1}{\delta} = 0.$$

Solving for δ gives the desired answer

$$\tilde{\theta}_n = \frac{1}{E_\pi[1/\theta|X]}.$$

(One can check that the second derivative equals $-1/\delta^2 < 0$, and hence solving the above gives a global maximum, but this is not necessary). 5, D

(c) seen ⇓

$$\begin{aligned} \pi(\theta|X_1, \dots, X_n) &\propto \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} \theta^{a-1} (1-\theta)^{b-1} \\ &\propto \theta^{a + \sum x_i - 1} (1-\theta)^{b + n - \sum x_i - 1}, \end{aligned}$$

which we recognize as the form of a Beta$(a + \sum x_i, b + n - \sum x_i)$ distribution. 3, A

(d) For $\theta \sim$ Beta(a, b), we have meth seen ⇓

$$\begin{aligned} E[1/\theta] &= \int_0^1 \frac{1}{\theta} \frac{(a+b-1)!}{(a-1)!(b-1)!} \theta^{a-1} (1-\theta)^{b-1} d\theta \\ &= \frac{(a+b-1)!}{(a-1)!(b-1)!} \int_0^1 \theta^{a-2} (1-\theta)^{b-1} d\theta \\ &= \frac{(a+b-1)!}{(a-1)!(b-1)!} \frac{(a-2)!(b-1)!}{(a+b-2)!} = \frac{a+b-1}{a-1}. \end{aligned}$$

Plugging in the parameters for the posterior distribution Beta$(a + \sum x_i, b + n - \sum x_i)$,

$$E[1/\theta|X_1, \dots, X_n] = \frac{a+b+n-1}{a + \sum x_i - 1}.$$

Hence we have

$$\tilde{\theta}_n = \frac{a + \sum x_i - 1}{a+b+n-1}.$$

3, C

(e) Following the hint, sim. seen ⇓

$$\begin{aligned}\sqrt{n}(\tilde{\theta}_n - \bar{X}_n) &= \frac{\sqrt{n}}{n(a+b+n-1)}\left\{an - (a+b)\sum x_i\right\} \\ &= \frac{\sqrt{n}a}{(a+b+n-1)} - \frac{(a+b)\sqrt{n}}{(a+b+n-1)}\bar{X}_n.\end{aligned}$$

By the weak law of large numbers and Slutsky's theorem, this converges in probability to 0. Hence we can decompose,

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \sqrt{n}(\tilde{\theta}_n - \bar{X}_n) + \sqrt{n}(\bar{X}_n - \theta_0).$$

The first term tends to 0 in probability as shown, while the second converges in distribution to a $N(0, \theta_0(1-\theta_0))$ by the central limit theorem. Hence by Slutsky's theorem, $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ also converges to this normal distribution. 4, C

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 $1/n$. $\boxed{\text{seen} \Downarrow}$

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

(b) When there is no restriction on μ, we saw that the MLE is $\bar{X}_n$. However, if $\bar{X}_n < 0$, this is outside of the parameter space $[0, \infty)$. If $\bar{x}_n < 0$, then the log-likelihood $\ell_n : [0, \infty) \to \mathbb{R}$ equals $\boxed{\text{seen} \Downarrow}$

$$\ell_n(\mu) = \log\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i-\mu)^2/2}\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n x_i^2 + n\bar{x}_n\mu - \frac{n\mu^2}{2}.$$

Differentiating, $\ell_n'(\mu) = n(\bar{x}_n - \mu) < 0$ if $\mu > 0$ and $\bar{x}_n < 0$. Thus ℓ_n is a decreasing function of μ for $\mu \ge 0$ if $\bar{x}_n < 0$, and is thus maximized at $\mu = 0$. The MLE is therefore

$$\hat{\mu}_n = \bar{X}_n 1\{\bar{X}_n \ge 0\} = \max(\bar{X}_n, 0).$$

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

For the limit distributions, consider first $\mu_0 > 0$. Then

$$P_\mu(\sqrt{n}(\hat{\mu}_n - \mu_0) \le x) = P_\mu(\hat{\mu}_n \le \mu_0 + x/\sqrt{n}) = P_\mu(\bar{X}_n \le \mu_0 + x/\sqrt{n})$$

if $\mu_0 + x/\sqrt{n} \ge 0$. For $\mu_0 > 0$ this is fulfilled for each x if n is large enough. Consequently,

$$\lim_{n\to\infty} P_\mu(\sqrt{n}(\hat{\mu}_n - \mu_0) \le x) = \lim_{n\to\infty} P_\mu(\sqrt{n}(\bar{X}_n - \mu_0) \le x) = \phi(x)$$

since $\sqrt{n}(\bar{X}_n - \mu_0) \to^d N(0,1)$ (actually the $\to^d$ is $=^d$). Thus for $\mu_0 > 0$, $\sqrt{n}(\hat{\mu}_n - \mu_0) \to^d N(0,1)$.
If $\mu_0 = 0$, then

$$\begin{aligned}\sqrt{n}(\hat{\mu}_n - \mu_0) &= \sqrt{n}\max(\bar{X}_n, 0)\\ &= \sqrt{n}\max(\bar{X}_n - \mu_0, 0)\\ &= \max(\sqrt{n}(\bar{X}_n - \mu_0), 0)\\ &\to^d \max(Z, 0) =: W_0,\end{aligned}$$

for $Z \sim N(0,1)$ by the continuous mapping theorem. $\boxed{4, \mathrm{M}}$

(c) When using a bootstrap sample for a statistic $T(X_1, \ldots, T_n)$, 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 target quantity μ_0 by its estimated value $\hat{\mu}_n$ and the observations X_i with our bootstrap sample X_i^*. $\boxed{\text{seen/sim.seen} \Downarrow}$

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

(d) On the event $\{\sqrt{n}\bar{X}_n > t\}$, we have $\boxed{\text{unseen} \Downarrow}$

$$\begin{aligned}\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) &= \sqrt{n}(\max(\bar{X}_n^*, 0) - \max(\bar{X}_n, 0))\\ &= \sqrt{n}\max(\bar{X}_n^* - \bar{X}_n, -\bar{X}_n)\\ &\le \max(\sqrt{n}(\bar{X}_n^* - \bar{X}_n), -t).\end{aligned}$$

Thus

$$\mathbb{P}\left(\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) \le x|\sqrt{n}\bar{X}_n > t\right) \ge \mathbb{P}\left(\max\{\sqrt{n}(\bar{X}_n^* - \bar{X}_n), -t\} \le x|\sqrt{n}\bar{X}_n > t\right).$$

3, M

Using the hint, for every sequence $X_1, X_2, \ldots$ on a set of probability one, $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)|X_1, \ldots, X_n \to^d N(0,1)$, and hence by the continuous mapping theorem,

$$\mathbb{P}\left(\max\{\sqrt{n}(\bar{X}_n^* - \bar{X}_n), -t\} \leq x | \sqrt{n}\bar{X}_n > t\right) \to \mathbb{P}\left(\max(Z, -t) \leq x\right)$$

as $n \to \infty$, where $Z \sim N(0,1)$. Since $\max(Z, -t) \leq \max(Z, 0) =^d W_0$, this is lower bounded by $\mathbb{P}(W_0 \leq x)$. 3, M

(e) No, the last result shows that the bootstrap distribution of $R_n(X^*, \hat{\mu}_n)$ does not match that of limit of $R_n(X, \mu_0)$, since the probabilities they are in the sets $(-\infty, x]$ are different. unseen $\Downarrow$ 2, M

Review of mark distribution:

Total A marks: 31 of 32 marks
Total B marks: 21 of 20 marks
Total C marks: 11 of 12 marks
Total D marks: 17 of 16 marks
Total marks: 100 of 80 marks
Total Mastery marks: 20 of 20 marks

MATH70043 Statistical Theory Markers Comments

Question 1	1(b) Many students derived the MLE as max X_i instead of max |X_i|. Alternative correct answers such as max(max X_i,-min X_i) were given full marks. 1(c) some students claimed the limiting distribution is normal while in fact it is exponential. 1(e) some students claimed that the limiting the distribution is normal (likely a consequence of earlier mistakes). The result can either be derived as in the solution usinf Slutsky's theorem, or by deriving cdf of the random variable of interest and finding its limit, or by using the delta rule. Solutions by any of these methods were given full marks.
Question 2	Overall well done everyone, especially for parts (a) and (c). Parts (b) and (d) were more challenging, with some mistakes in the derivations, please do review the solutions.
Question 3	No Comment
Question 4	This question was well done. Be careful: the risk is the expectation of the loss, holding theta fixed and taking the expectation over the distribution of X given theta.
Question 5	This question was not done very well. Answers showed good understanding of the ideas, but a lot of the attempts were unconvincing of the technical derivations in (b), (d).