M3/M4
S1

Imperial College
London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2015

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

# Statistical Theory I

Date: Monday, 11 May 2015. Time: 2.00pm – 4.00pm. Time allowed: 2 hours.

This paper has FOUR questions.

Candidates should use ONE main answer book.

Supplementary books may only be used after the main book is full.

Statistical tables are provided on pages 5 & 6.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.
- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.
- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

| Raw mark | up to 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Extra credit | 0 | $\frac{1}{2}$ | 1 | $1\frac{1}{2}$ | 2 | $2\frac{1}{2}$ | 3 | $3\frac{1}{2}$ | 4 |

- Each question carries equal weight.
- Calculators may not be used.

 

1. (a) Let $X$ denote the observed data, whose distribution depends on an unknown parameter $\theta \in \Theta$. Let $T = t(X)$ be some statistic. Write down the definition for each of the following concepts:

   (i) $T$ is sufficient.

   (ii) $T$ is minimal sufficient.

   (iii) $T$ is complete.

   (b) Let $X_1, \ldots, X_n$ be i.i.d. $\text{Exponential}(\theta)$ random variables.

   (i) Show that $\bar{X}$ is the Cramér-Rao Unbiased estimator of some function $\mu(\theta)$ of $\theta$, and write down its variance.

   (ii) Find a variance stabilising transformation $g$ such that

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} \mathcal{N}(0,1).$$

   (iii) Show that for such $\text{Exponential}(\theta)$ observations, the $\text{Gamma}(\alpha, \beta)$ prior is a conjugate Bayesian prior.

   (iv) Let the prior distribution of $\theta$ be $\text{Gamma}(\alpha, \beta)$. Compute the posterior mode.

2. (a) Let $\theta \in \Theta$ be an unknown parameter and let $X$ denote the observed data. Consider the null hypothesis $H_0 : \theta \in \Theta_0$ and alternative hypothesis $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$.

   (i) Give the definitions for the size $\alpha$ and the power function $\beta$ of a hypothesis test with critical region $R$. You may use the notation $P_\theta$ to denote the dependence of the probability measure on $\theta$.

   (ii) Give the definition of an unbiased test.

   (iii) Explain how one can construct a $100(1-\alpha)\%$ confidence interval for $\theta$ by first considering, for various $\theta_0$, size $\alpha$ tests of $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta \neq \theta_0$.

   (b) Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim \mathcal{N}(\theta x_i, \sigma^2)$, where $\sigma^2$ and the $x_i$ are known constants.

   (i) Find the Cramér-Rao unbiased estimator for $\theta$, and write down the corresponding Cramér-Rao lower bound.

   (ii) Find an unbiased estimator of $\theta$ which is a function of $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. What is the efficiency of this estimator?

   (c) Suppose that we observe both $X \sim \text{Geometric}(1-\theta)$ and $Y \sim \text{Poisson}(\theta)$. $X$ and $Y$ are independent. $\theta \in (0,1)$ is an unknown parameter.

   (i) Write down a minimal sufficient statistic $t(x, y)$ for $\theta$.

   (ii) Show that the likelihood satisfies the monotone likelihood ratio criterion.

   (iii) Does a similar uniformly most powerful randomised test of size $\alpha = 0.01$ for testing $H_0 : \theta = 0.5$ against $H_1 : \theta < 0.5$ exist? Write one sentence justifying your answer.

3. Let $X_1, \ldots, X_n$ be i.i.d. samples with PMF $f_X(x) = \theta(1-\theta)^x \mathbb{1}_{x \geq 0}$, and unknown parameter $\theta \in (0,1)$. This distribution is an alternative version of the geometric distribution with range $\{0, 1, 2, \ldots\}$.

(a) Justify without proof why $S = \sum_{i=1}^n X_i$ is a complete sufficient statistic for $\theta$.

(b) Find an unbiased estimator for $\theta^2$ in the case where $n = 1$, by comparing coefficients of $(1-\theta)^k$ in a suitable expansion.

(c) Compute the total score function $U_\bullet(\theta)$ and the total Fisher information $I_\bullet(\theta)$.

(d) Explain why there is no Cramér-Rao unbiased estimator of $\theta$.

(e) $\mathbb{1}_{X_1=0}$ is an unbiased estimator for $\theta$. Assuming $n > 1$, obtain an improved estimator by applying the Rao-Blackwell procedure using the sufficient statistic $S$.
*[Hint: S follows a Negative-Binomial distribution with range $\{0, 1, 2, \ldots\}$ because it is a sum of $n$ Geometric random variables with ranges $\{0, 1, 2, \ldots\}$.]*

(f) Is the improved estimator obtained in (e) the minimum-variance unbiased estimator for $\theta$? Justify your answer.

4. Let $X_1, \ldots, X_n$ be i.i.d. samples from a $\text{Uniform}(0, \theta_X)$ distribution, and let $Y_1, \ldots, Y_n$ be independent i.i.d. samples from a $\text{Uniform}(0, \theta_Y)$ distribution. We are interested in the hypotheses $H_0 : \theta_X = \theta_Y$ and $H_1 : \theta_X \neq \theta_Y$. Let $X_{(n)} = \max(X_1, \ldots, X_n)$, let $Y_{(n)} = \max(Y_1, \ldots, Y_n)$ and let $T = \max(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = \max(X_{(n)}, Y_{(n)})$.

(a) Show that $(X_{(n)}, Y_{(n)})$ is a sufficient statistic for $(\theta_X, \theta_Y)$.

(b) Is the hypothesis $H_0$ simple or composite? Is the hypothesis $H_1$ simple or composite? You are not required to justify your answers.

(c) (i) Let $\Lambda = \Lambda(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = 2\log(\lambda)$, where $\lambda = \frac{\sup_{\theta_X, \theta_Y \in (0,\infty)} L(\theta_X, \theta_Y)}{\sup_{\theta_X = \theta_Y \in (0,\infty)} L(\theta_X, \theta_Y)}$ is the (generalised) likelihood ratio. Assuming $H_0$ is true, show that $\Lambda$ is ancillary for $\theta = \theta_X = \theta_Y$.

(ii) Assuming $H_0$ is true, show that the distribution of $\Lambda$ is $\chi^2_2$. You may use the following facts without proof:

1. If $A \sim \chi^2_m$ is independent of $B \sim \chi^2_k$ then $A + B \sim \chi^2_{m+k}$.
2. If $A$ is independent of $B \sim \chi^2_k$ and $A + B \sim \chi^2_{m+k}$ then $A \sim \chi^2_m$.
3. If $Z \sim \text{Beta}(1, \beta)$ then $-2\beta \log(Z) \sim \chi^2_2$.
4. $\frac{X_{(n)}}{\theta_X} \sim \text{Beta}(1, n)$, independently of $\frac{Y_{(n)}}{\theta_Y} \sim \text{Beta}(1, n)$.
5. If $\theta = \theta_X = \theta_Y$ then $T$ is a complete sufficient statistic for $\theta$ and $\frac{T}{\theta} \sim \text{Beta}(1, 2n)$.

*[Hint: First show that $T$ is independent of $\Lambda$.]*

(iii) Compute the critical region for a likelihood ratio test of size $\alpha$ of $H_0$ v.s. $H_1$. *[Hint: The $\chi^2_2$ distribution, the* $\text{Exponential}\left(\frac{1}{2}\right)$ *distribution and the* $\text{Gamma}\left(1, \frac{1}{2}\right)$ *distribution are all identical.]*

(d) You may assume without loss of generality that $n$ is large. Comment on how the distribution in (c)(ii) relates to Wilks' Theorem. Explain why this test violates the regularity conditions for Wilks' Theorem. *[Hint: The regularity conditions for Wilks' Theorem are the same as those given for the asymptotic normality of maximum likelihood estimators. The regularity condition which is violated here is also one of the conditions needed to prove the Cramér-Rao lower bound.]*

## DISCRETE DISTRIBUTIONS

| | RANGE $\mathbb{X}$ | PARAMETERS | MASS FUNCTION $f_X$ | CDF $F_X$ | $\mathsf{E}_{f_X}[X]$ | $\mathsf{Var}_{f_X}[X]$ | MGF $M_X$ |
|---|---|---|---|---|---|---|---|
| $Bernoulli(\theta)$ | $\{0,1\}$ | $\theta \in (0,1)$ | $\theta^x(1-\theta)^{1-x}$ | | $\theta$ | $\theta(1-\theta)$ | $1-\theta+\theta e^t$ |
| $Binomial(n,\theta)$ | $\{0,1,...,n\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{n}{x}\theta^x(1-\theta)^{n-x}$ | | $n\theta$ | $n\theta(1-\theta)$ | $(1-\theta+\theta e^t)^n$ |
| $Poisson(\lambda)$ | $\{0,1,2,...\}$ | $\lambda \in \mathbb{R}^+$ | $\dfrac{e^{-\lambda}\lambda^x}{x!}$ | | $\lambda$ | $\lambda$ | $\exp\{\lambda(e^t-1)\}$ |
| $Geometric(\theta)$ | $\{1,2,...\}$ | $\theta \in (0,1)$ | $(1-\theta)^{x-1}\theta$ | $1-(1-\theta)^x$ | $\dfrac{1}{\theta}$ | $\dfrac{(1-\theta)}{\theta^2}$ | $\dfrac{\theta e^t}{1-e^t(1-\theta)}$ |
| $NegBinomial(n,\theta)$ | $\{n,n+1,...\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$ | | $\dfrac{n}{\theta}$ | $\dfrac{n(1-\theta)}{\theta^2}$ | $\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$ |
| or | $\{0,1,2,...\}$ | $n \in \mathbb{Z}^+, \theta \in (0,1)$ | $\binom{n+x-1}{x}\theta^n(1-\theta)^x$ | | $\dfrac{n(1-\theta)}{\theta}$ | $\dfrac{n(1-\theta)}{\theta^2}$ | $\left(\dfrac{\theta}{1-e^t(1-\theta)}\right)^n$ |

For **CONTINUOUS** distributions (see over), define the **GAMMA FUNCTION**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

and the LOCATION/SCALE transformation $Y = \mu + \sigma X$ gives

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) \qquad M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathsf{E}_{f_Y}[Y] = \mu+\sigma\mathsf{E}_{f_X}[X] \qquad \mathsf{Var}_{f_Y}[Y] = \sigma^2\mathsf{Var}_{f_X}[X]$$

## CONTINUOUS DISTRIBUTIONS

| | $\mathbb{X}$ | PARAMS. | PDF $f_X$ | CDF $F_X$ | $\mathsf{E}_{f_X}[X]$ | $\mathsf{Var}_{f_X}[X]$ | MGF $M_X$ |
|---|---|---|---|---|---|---|---|
| $Uniform(\alpha,\beta)$ (standard model $\alpha=0,\beta=1$) | $(\alpha,\beta)$ | $\alpha<\beta\in\mathbb{R}$ | $\frac{1}{\beta-\alpha}$ | $\frac{x-\alpha}{\beta-\alpha}$ | $\frac{(\alpha+\beta)}{2}$ | $\frac{(\beta-\alpha)^2}{12}$ | $\frac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)}$ |
| $Exponential(\lambda)$ (standard model $\lambda=1$) | $\mathbb{R}^+$ | $\lambda\in\mathbb{R}^+$ | $\lambda e^{-\lambda x}$ | $1-e^{-\lambda x}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | $\left(\frac{\lambda}{\lambda-t}\right)$ |
| $Gamma(\alpha,\beta)$ (standard model $\beta=1$) | $\mathbb{R}^+$ | $\alpha,\beta\in\mathbb{R}^+$ | $\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ | | $\frac{\alpha}{\beta}$ | $\frac{\alpha}{\beta^2}$ | $\left(\frac{\beta}{\beta-t}\right)^\alpha$ |
| $Weibull(\alpha,\beta)$ (standard model $\beta=1$) | $\mathbb{R}^+$ | $\alpha,\beta\in\mathbb{R}^+$ | $\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$ | $1-e^{-\beta x^\alpha}$ | $\frac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}$ | $\frac{\Gamma(1+2/\alpha)-\Gamma(1+1/\alpha)^2}{\beta^{2/\alpha}}$ | |
| $Normal(\mu,\sigma^2)$ (standard model $\mu=0,\sigma=1$) | $\mathbb{R}$ | $\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$ | $\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ | | $\mu$ | $\sigma^2$ | $e^{\{\mu t+\sigma^2 t^2/2\}}$ |
| $Student(\nu)$ | $\mathbb{R}$ | $\nu\in\mathbb{R}^+$ | $\frac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$ | | $0$ (if $\nu>1$) | $\frac{\nu}{\nu-2}$ (if $\nu>2$) | |
| $Pareto(\theta,\alpha)$ | $\mathbb{R}^+$ | $\theta,\alpha\in\mathbb{R}^+$ | $\frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$ | $1-\left(\frac{\theta}{\theta+x}\right)^\alpha$ | $\frac{\theta}{\alpha-1}$ (if $\alpha>1$) | $\frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}$ (if $\alpha>2$) | |
| $Beta(\alpha,\beta)$ | $(0,1)$ | $\alpha,\beta\in\mathbb{R}^+$ | $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ | | $\frac{\alpha}{\alpha+\beta}$ | $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ | |

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2015

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

# M3S1/M4S1

## Statistical Theory I

Date: Monday, 11th May 2015 Time: 2 pm – 4 pm

# Solutions

seen ⇓

1. (a) (i) The conditional distribution of $X$ given $T$ does not depend on $\theta$. 2

(ii) $T$ is sufficient and for any other sufficient statistic $S$, $T$ is a function of $S$. 2

(iii) The only function $h$ which satisfies $\mathrm{E}[h(T)] = 0 \ \forall\theta$ is $h(T) = 0$ (almost surely). 2

(b) (i) $\ell(\theta) = n\log(\theta) - \theta\sum_{i=1}^{n} x_i$. meth seen ⇓
$U_\bullet(\theta) = \ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} X_i = n(\mu(\theta) - \bar{X})$. By inspection $\bar{X}$ is the CRUE of $\mu(\theta) = \frac{1}{\theta}$. 3
$\mathrm{var}(\bar{X}) = \frac{1}{n\theta^2} = \frac{\mu^2}{n}$ 1

(ii) Take

$$g(\mu) = \int \left(n\mathrm{var}(\bar{X})\right)^{-\frac{1}{2}} \,\mathrm{d}\mu = \int \frac{1}{\mu}\,\mathrm{d}\mu = \log(\mu) \ (+C).$$

3

(iii) Applying Bayes theorem the posterior is

$$\begin{aligned}\pi(\theta|\mathrm{x}) &\propto \pi(\theta)L(\theta)\\ &\propto \theta^{\alpha-1}\mathrm{e}^{-\beta\theta}\theta^{n}\mathrm{e}^{-\theta\sum_{i=1}^{n} x_i}\\ &= \theta^{\alpha+n-1}\mathrm{e}^{-\theta\left(\beta+\sum_{i=1}^{n} x_i\right)}\\ &\propto f_{\mathrm{Gamma}(\alpha+n,\beta+n\bar{x})}(\theta),\end{aligned}$$

which is again a Gamma distribution. 4

(iv) $\frac{\mathrm{d}\log\pi(\theta|\mathrm{x})}{\mathrm{d}\theta} = \frac{\alpha+n-1}{\theta} - (\beta + \sum_{i=1}^{n} x_i)$.
Solving $\frac{\mathrm{d}\log\pi(\theta|\mathrm{x})}{\mathrm{d}\theta} = 0$ gives us the posterior mode $\frac{\alpha+n-1}{\beta+\sum_{i=1}^{n} X_i}$ 3

2. (a) (i) $\alpha = \sup_{\theta \in \Theta_0} \mathrm{P}_\theta(X \in R)$.
$\beta(\theta) = \mathrm{P}_\theta(X \in R)$. seen $\Downarrow$ $\boxed{2}$

(ii) A test is unbiased if $\beta(\theta_0) \leq \beta(\theta_1)$ for all $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$. Alternatively $\beta(\theta_1) \geq \alpha \ \forall \theta_1 \in \Theta_1$ $\boxed{1}$

(iii) Let $R(\theta_0)$ denote the critical region for a test of size $\alpha$ of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Let $\Psi = \{\theta_0 : X \notin R(\theta_0)\}$. Then $\Psi$ is a $100(1-\alpha)\%$ confidence interval since $\mathrm{P}_\theta(\theta \in \Psi) = \mathrm{P}_\theta(X \notin R(\theta)) = 1 - \alpha$. $\boxed{3}$

(b) (i) $\ell(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \theta x_i)^2$.
$U_\bullet(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^n x_i(Y_i - \theta x_i) = \frac{\sum_{i=1}^n x_i^2}{\sigma^2}\left(\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} - \theta\right)$.
Hence by inspection $\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ is the CRUE of $\theta$. $\boxed{3}$
The CRLB is $\frac{\sigma^2}{\sum_{i=1}^n x_i^2}$. $\boxed{1}$

(ii) $\mathrm{E}[\bar{Y}] = \frac{\sum_{i=1}^n x_i}{n}\theta$. Hence an unbiased estimator is $T = \frac{n\bar{Y}}{\sum_{i=1}^n x_i}$. meth seen $\Downarrow$ $\boxed{1}$

$\mathrm{var}\left(\frac{n\bar{Y}}{\sum_{i=1}^n x_i}\right) = \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2}$. Hence
$\mathrm{Efficiency}(T) = \frac{CRLB}{\mathrm{var}(T)} = \frac{\left(\sum_{i=1}^n x_i\right)^2}{n\sum_{i=1}^n x_i^2}$. $\boxed{2}$

(c) (i) $L(\theta) = (1-\theta)\theta^{x-1}\frac{\mathrm{e}^{-\theta}\theta^y}{y!} = \frac{1}{y!}(1-\theta)\mathrm{e}^{-\theta}\theta^{x+y-1}$.
$t = x + y$ is minimal sufficient. $\boxed{2}$

(ii) Let $0 < \theta_0 < \theta_1 < 1$. The likelihood ratio for $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is

$$\lambda = \frac{(1-\theta_1)\mathrm{e}^{-\theta_1}}{(1-\theta_0)\mathrm{e}^{-\theta_0}}\left(\frac{\theta_1}{\theta_0}\right)^{t-1}.$$

$\frac{\theta_1}{\theta_0} > 1$ hence this is an increasing function of the sufficient statistic $t$ and the monotone likelihood ratio criterion is satisfied. $\boxed{4}$

(iii) Yes. The likelihood ratio test is uniformly most powerful because the monotone likelihood ratio criterion is satisfied. $\boxed{1}$

3. (a) $S = \sum_{i=1}^{n} X_i$ is complete and sufficient because it is the natural statistic $\tau$ of the 1-parameter exponential family followed by $X_1, \ldots, X_n$. meth seen $\Downarrow$ $\boxed{2}$

(b) Let $t(x)$ be an unbiased estimate of $\theta^2$. Then

$$\begin{aligned}
&\sum_{x=0}^{\infty} t(x)\theta(1-\theta)^x = \theta^2 \\
\iff &\sum_{x=0}^{\infty} t(x)(1-\theta)^x = \theta \\
\iff &t(0) + t(1)(1-\theta) = \theta \\
\iff &t(0) = 1,\ t(1) = -1,\ \&\, t(x) = 0 \text{ otherwise.}
\end{aligned}$$

Hence an unbiased estimator of $\theta^2$ is given by $T = \mathbb{1}_{X_1=0} - \mathbb{1}_{X_1=1}$. $\boxed{4}$

(c) $\ell(\theta) = n\log(\theta) + S\log(1-\theta)$.
$U_\bullet(\theta) = \ell'(\theta) = \frac{n}{\theta} - \frac{S}{1-\theta} = \frac{n}{\theta-1}\left(\bar{X} + 1 - \frac{1}{\theta}\right)$. $\boxed{2}$
$I_\bullet(\theta) = \mathrm{E}\left[-\frac{\mathrm{d}}{\mathrm{d}\theta}U_\bullet(\theta)\right] = \mathrm{E}\left[\frac{n}{\theta^2} + \frac{S}{(1-\theta)^2}\right] = \frac{n}{\theta^2} + \frac{\frac{n}{\theta}-n}{(1-\theta)^2} = \frac{n}{\theta^2(1-\theta)}$. $\boxed{2}$

(d) $U_\bullet(\theta)$ cannot be written in the form $\frac{1}{c(\theta)}(T-\theta)$ for any statistic $T$. Alternatively, note that $\theta \pm \frac{U_\bullet(\theta)}{I_\bullet(\theta)}$ is not a statistic. $\boxed{2}$

(e) The Rao-Blackwell estimate is given by

$$\begin{aligned}
\mathrm{E}[\mathbb{1}_{X_1=0}|S=s] &= \mathrm{P}(X_1=0|S=s) \\
&= \frac{\mathrm{P}(S=s|X_1=0)\mathrm{P}(X_1=0)}{\mathrm{P}(S=s)} \\
&= \frac{\binom{n-1+s-1}{s}\theta^{n-1}(1-\theta)^s \cdot \theta(1-\theta)^0}{\binom{n+s-1}{s}\theta^n(1-\theta)^s} \\
&= \frac{(n+s-2)!s!(n-1)!}{(n+s-1)!s!(n-2)!} \\
&= \frac{n-1}{s+n-1}.
\end{aligned}$$

$\boxed{6}$

(f) Yes. The improved estimator in (f) is unbiased, and it is a function of the complete sufficient statistic $S$. Hence by the Lehman-Scheffé theorem it is the MVUE. $\boxed{2}$

4. (a) sim. seen ⇓

$$
\begin{aligned}
L(\theta_X, \theta_Y) &= \prod_{i=1}^{n} \left( \frac{1}{\theta_X} \mathbb{1}_{X_i \le \theta_X} \frac{1}{\theta_Y} \mathbb{1}_{Y_i \le \theta_Y} \right) \\
&= \theta_X^{-n} \theta_Y^{-n} \mathbb{1}_{X_{(n)} \le \theta_X} \mathbb{1}_{Y_{(n)} \le \theta_Y} \\
&= g\left((\theta_X, \theta_Y), (X_{(n)}, Y_{(n)})\right).
\end{aligned}
$$

Hence $(X_{(n)}, Y_{(n)})$ is sufficient by the Neyman factorisation theorem. 2

(b) Both $H_0$ and $H_1$ are composite. 1

(c) (i) unseen ⇓

$$
\lambda = \frac{L(X_{(n)}, Y_{(n)})}{L(T, T)} = T^{2n} X_{(n)}^{-n} Y_{(n)}^{-n}.
$$

Hence $\Lambda = 4n \log(T) - 2n \log(X_{(n)}) - 2n \log(Y_{(n)})$. 3

Assuming $H_0$ is true, then $\theta = \theta_X = \theta_Y$ is a scale parameter, and

$$
\begin{aligned}
\Lambda(\theta X_1, \ldots, \theta Y_n) &= 4n \log(\theta T) - 2n \log(\theta X_{(n)}) - 2n \log(\theta Y_{(n)}) \\
&= \Lambda(X_1, \ldots, Y_n) - 4n \log(\theta) - 2n \log(\theta) - 2n \log(\theta) \\
&= \Lambda(X_1, \ldots, Y_n).
\end{aligned}
$$

Hence $\Lambda$ is ancillary for $\theta$. 2

(ii) $T$ is a complete sufficient statistic and $\Lambda$ is ancillary. Hence by Basu's theorem $T$ is independent of $\Lambda$. 2

Using the same approach as in (c)(i), we may write
$\Lambda = 4n \log(\theta^{-1} T) - 2n \log(\theta^{-1} X_{(n)}) - 2n \log(\theta^{-1} Y_{(n)})$,
which can be rearranged as
$-2n \log(\theta^{-1} X_{(n)}) - 2n \log(\theta^{-1} Y_{(n)}) = \Lambda - 4n \log(\theta^{-1} T)$.
(Alternatively note that we may assume $\theta = 1$ without affecting the the distribution of $\Lambda$ because it is ancillary.)
From facts 4 and 3, $-2n \log(\theta^{-1} X_{(n)}) \sim \chi_2^2$ independently of $-2n \log(\theta^{-1} Y_{(n)}) \sim \chi_2^2$. Hence by fact 1 the LHS is $\chi_4^2$.
From facts 5 and 3 $-4n \log(\theta^{-1} T) \sim \chi_2^2$, and we showed that it is independent of $\Lambda$. Hence $\chi_4^2 = \Lambda + \chi_2^2$. Now $\Lambda \sim \chi_2^2$ follows from fact 2. 4

(iii) The critical region is $\Lambda > z$, where $z$ is chosen such that $\mathrm{P}(\Lambda > z | H_0) = 1 - F_{\chi_2^2}(z) = \alpha$. Using the hint this simplifies to $e^{-\frac{1}{2} z} = \alpha$ and the solution is $z = -2 \log(\alpha)$.
Equivalently, the critical region can be written as
$R = \{(X_1, \ldots, X_n, Y_1, \ldots, Y_n) : \Lambda(X_1, \ldots, X_n, Y_1, \ldots, Y_n) > -2 \log(\alpha)\}$. 2

(d) Under $H_1$ the parameter space is 2-dimensional and under $H_0$ it is 1-dimensional. Based on Wilks' Theorem one would expect $\Lambda \xrightarrow{d} \chi_1^2$, where the degrees of freedom is the difference between the two dimensionalities. However this is contraddicted by (c)(ii) which states that $\Lambda \sim \chi_2^2$. Wilks' Theorem does not apply here because one of the regularity conditions requires that the range of the samples does not depend on the parameter(s). 4