Imperial College
London

M4/5S1

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2017

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

Statistical Theory I

Date: Friday 02 June 2017

Time: 14:00 - 16:30

Time Allowed: 2.5 Hours

This paper has 5 Questions.

Candidates should use ONE main answer book.

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

All required additional material will be provided.

- 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	½	1	1 ½	2	2 ½	3	3 ½	4

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

1. (a) State and prove Basu's Theorem.

(b) Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with the probability function

$$P_\theta(X = x) = \theta^x (1-\theta)^{1-x}, \qquad x = 0, 1, \;\; 0 < \theta < 1.$$

(i) Find a one-dimensional sufficient statistic for θ.

(ii) Show whether the sufficient statistic in part (i) is complete or not.

(iii) Obtain the UMVUE of θ^r, where r is a positive integer and $r \leq n$.

(iv) Determine the value(s) of r for which the UMVUE achieves the Cramer-Rao lower bound. Justify your answer.

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables from a Weibull distribution with the probability density function

$$f(x) = \alpha\beta x^{\alpha-1} e^{-\beta x^\alpha}, \qquad x > 0, \;\; \alpha, \beta > 0,$$

where α is known.

(a) Obtain a method of moments (MM) estimator of β.

(b) For the case when $\alpha = \frac{1}{2}$, derive the asymptotic distribution of the MM estimator of β in part (a).

(c) Find the maximum likelihood estimator (MLE) of β.

(d) Derive the asymptotic distribution of the MLE of β.

(e) For the case when $\alpha = \frac{1}{2}$, write down the asymptotic relative efficiency of the MLE of β to the MM estimator of β and find out which estimator is more efficient.

3. Let $X_1, X_2, \ldots$ be i.i.d. random variables each with the exponential distribution *Exp*(1). Suppose that we observe $Y = X_1 + \ldots + X_{\theta+1}$, where θ is an unknown integer $\theta \geq 1$. Assume θ has as prior distribution a geometric distribution given by

$$\pi(\theta) = p(1-p)^{\theta-1}, \qquad \theta = 1, 2, \ldots$$

where p is a known constant.

(a) Obtain the posterior distribution of θ.
[Hint: You may use the fact that $\sum_{k=0}^{\infty} \frac{A^k}{k!} = e^A$.*]*

(b) Is the prior a conjugate prior? Justify your answer.

(c) Find the Bayesian point estimator of θ under the squared error loss function.

(d) State whether or not the Bayes estimator obtained in part (c) is admissible, and explain why.

4. (a) Consider the following probability density function

$$f_\theta(x) = \frac{e^{(x-\theta)}}{\left(1+e^{(x-\theta)}\right)^2}, \qquad -\infty < x < \infty, \ -\infty < \theta < \infty.$$

Based on one observation X, find the most powerful size α test of $H_0 : \theta = 0$ versus $H_1 : \theta = 1$.
Note that the cumulative distribution function of X is $F_\theta(x) = e^{(x-\theta)}/(1+e^{(x-\theta)})$.

(b) Consider the following probability density function

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}, \qquad -\infty < \mu < x < \infty, \ 0 < \sigma < \infty,$$

where μ and σ are unknown parameters.

(i) Based on n i.i.d. observations $X_1, \ldots, X_n$, derive the likelihood ratio test statistic for testing $H_0 : \sigma = 1$ versus $H_1 : \sigma \neq 1$.

(ii) For a given significance level α, perform the likelihood ratio test in part (i) using the asymptotic distribution of the likelihood ratio test statistic under H_0.

(iii) Assume $\mu = 0$. Based on n i.i.d. observations $X_1, \ldots, X_n$, construct an approximate two-sided confidence interval for σ with the confidence coefficient $1 - \alpha$.

Mastery Question:

5. Let $X_1, \ldots, X_n$ be a random sample (i.i.d.) from a population distribution. Suppose that the true population distribution is Uniform distribution on the interval $(0, b)$, but we erroneously assume that the X_i's have Gamma distribution

$$f_{\alpha,\lambda}(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \qquad x > 0,\ \alpha, \lambda > 0.$$

(a) Write down the log-likelihood equations.

(b) If $\hat{\alpha}_{MLE}$ and $\hat{\lambda}_{MLE}$ are the solutions to the log-likelihood equations in part (a), then show that $\hat{\alpha}_{MLE} \xrightarrow{P} \alpha_0$ and $\hat{\lambda}_{MLE} \xrightarrow{P} \lambda_0$ where α_0 and λ_0 satisfy

$$\begin{aligned} &\log(\lambda_0) - \tfrac{\Gamma'(\alpha_0)}{\Gamma(\alpha_0)} + \tfrac{1}{b} \textstyle\int_0^b \log(x)\, dx = 0, \\ &\tfrac{\alpha_0}{\lambda_0} - \tfrac{1}{b} \textstyle\int_0^b x\, dx = 0. \end{aligned}$$

(c) Suppose that we are interested to estimate the population mean. Obtain the maximum likelihood estimate of the population mean using the assumed Gamma distribution, and show that it will converge to the true population mean. What do you conclude?

DISCRETE DISTRIBUTIONS

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

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

CONTINUOUS DISTRIBUTIONS

	$\mathbb{X}$	PARAMS.	PDF f_X	CDF F_X	$\mathrm{E}_{f_X}[X]$	$\mathrm{Var}_{f_X}[X]$	MGF M_X
$Uniform(\alpha,\beta)$ (standard model $\alpha=0,\beta=1$)	(α,β)	$\alpha<\beta\in\mathbb{R}$	$\frac{1}{\beta-\alpha}$	$\frac{x-\alpha}{\beta-\alpha}$	$\frac{(\alpha+\beta)}{2}$	$\frac{(\beta-\alpha)^2}{12}$	$\frac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)}$
$Exponential(\lambda)$ (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 $\mu=0,\sigma=1$)	$\mathbb{R}$	$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\frac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		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 2017

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

M3S1/M4S1

Statistical Theory I

Date: Friday, 2 June 2017 Time: 14.00 – 16.00

Solutions

1. (a) **Basu's Theorem:** seen $\Downarrow$

If T is a complete and sufficient statistic for θ, then any ancillary statistic V is independent of T.

Proof:

From the lecture notes we know that a statistic V is ancillary for parameter θ if the distribution of V does not depend on θ. To prove the theorem, it is enough to show that

$$P_\theta(V \in A|T) = P_\theta(V \in A), \qquad \forall A.$$

For this, first note that

$$E_\theta\left(P_\theta(V \in A|T)\right) = E_\theta\left(E_\theta(I(V \in A|T))\right) = E_\theta(I(V \in A)) = P_\theta(V \in A), \quad \forall\theta, \forall A.$$

But, since V is ancillary, the probability $P_\theta(V \in A)$ does not depend on θ, and we then get

$$E_\theta\left[P_\theta(V \in A|T) - P_\theta(V \in A)\right] = 0, \quad \forall\theta, \forall A.$$

Hence, because T is complete we conclude that

$$P_\theta(V \in A|T) - P_\theta(V \in A) = 0, \quad \forall A,$$

and this completes the proof. $\boxed{4}$

(b) (i) According to Neyman Factorization Criterion, because seen $\Downarrow$

$$f_\theta(x) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i} = g\Big(\sum_{i=1}^{n} x_i, \theta\Big).h(x),$$

where $h(x) = 1$, we conclude that $T = \sum\limits_{i=1}^{n} X_i$ is a sufficient statistic for θ, which is also one-dimensional. $\boxed{4}$

(ii) Yes, $T = \sum\limits_{i=1}^{n} X_i$ is complete because Bernoulli distribution is a member of "full rank" exponential family of distributions since seen $\Downarrow$

$$f_\theta(x) = \exp\left(\ln\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^{n} x_i + n\ln(1-\theta)\right).$$

Another way to show T is complete is to follow the definition of completeness.

$\boxed{4}$

(iii) An unbiased estimator of θ^r is $I(X_1 = 1, \ldots, X_r = 1)$. Also, from parts (i) and (ii) above, $T = \sum\limits_{i=1}^{n} X_i$ is a complete and sufficient statistic. Therefore, unseen $\Downarrow$

using the Lehmann-Scheffe Theorem, the UMVUE of θ^r is as follows

$$\begin{aligned}
E(I(X_1=1,\dots,X_r=1)|T=t) &= P(X_1=1,\dots,X_r=1|\sum_{i=1}^{n} X_i=t) \\
&= \frac{P(X_1=1,\dots,X_r=1,\sum_{i=1}^{n} X_i=t)}{P(\sum_{i=1}^{n} X_i=t)} \\
&= \frac{P(X_1=1,\dots,X_r=1,\sum_{i=r+1}^{n} X_i=t-r)}{P(\sum_{i=1}^{n} X_i=t)} \\
&= \frac{\theta^r \binom{n-r}{t-r} \theta^{t-r}(1-\theta)^{n-r-t+r}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\binom{n-r}{t-r}}{\binom{n}{t}} = \frac{t(t-1)...(t-r+1)}{n(n-1)...(n-r+1)},
\end{aligned}$$

and, in fact, the UMVUE of θ^r is

$$\frac{T(T-1)...(T-r+1)}{n(n-1)...(n-r+1)},$$

where $T=\sum_{i=1}^{n} X_i$. $\boxed{4}$

There is another way to obtain the above UMVUE which is as follow.
Again using the Lemmann-Scheffe Theorem, the UMVUE of θ^r is a function of the complete and sufficient statistic $T=\sum_{i=1}^{n} X_i$ like $h(T)$ that is unbiased for θ^r. For this, we must have

$$\sum_{t=0}^{n} h(t)\binom{n}{t}\theta^t(1-\theta)^{n-t}=\theta^r \quad \Rightarrow \sum_{t=0}^{n} h(t)\binom{n}{t}\theta^{t-r}(1-\theta)^{n-t}=1.$$

To obtain $h(t)$, we assume $h(t)=0$ when $t<r$, then we get

$$\sum_{t=r}^{n} h(t)\binom{n}{t}\theta^t(1-\theta)^{n-t}=\theta^r \quad \Rightarrow \sum_{t=0}^{n-r} h(t+r)\binom{n}{t+r}\theta^t(1-\theta)^{n-r-t}=1$$

$$\Rightarrow \sum_{t=0}^{n-r} \frac{h(t+r)\binom{n}{t+r}}{\binom{n-r}{t}}\binom{n-r}{t}\theta^t(1-\theta)^{n-r-t}=1$$

$$\Rightarrow \sum_{t=0}^{n-r}\left(\frac{h(t+r)\binom{n}{t+r}}{\binom{n-r}{t}}-1\right)\binom{n-r}{t}\theta^t(1-\theta)^{n-r-t}=0$$

and because T is complete we must have

$$\frac{h(t+r)\binom{n}{t+r}}{\binom{n-r}{t}}-1=0 \quad \Rightarrow \quad h(t+r)=\frac{\binom{n-r}{t}}{\binom{n}{t+r}}, \quad t=0,\ldots,n-r.$$

or

$$h(t)=\frac{\binom{n-r}{t-r}}{\binom{n}{t}}=\frac{t(t-1)...(t-r+1)}{n(n-1)...(n-r+1)}, \quad t=r,r+1,\ldots,n.$$

and since $h(t)=0$ for $t<r$, the UMVUE is indeed the same as above.

meth seen ⇓

(iv) Considering the exponential form in part (ii) and using a theorem in the lecture notes, only estimators of the form $\left\{a\sum_{i=1}^{n}X_i+b\right\}$ achieve the Cramer-Rao lower bound. In other words, only linear function of $\sum_{i=1}^{n}X_i$ can attain the lower bound. Because the UMVUE of θ^r is a linear function of $\sum_{i=1}^{n}X_i$ only when $r=1$, therefore only for $r=1$ the UMVUE of θ^r achieves the Cramer-Rao lower bound. 4

sim. seen ⇓

2. (a) From the formulae sheets we know $E(X)=\frac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}$, therefore a method of moments estimator of β is as follows

$$\frac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}=\bar{X} \quad \Rightarrow \quad \widehat{\beta}_{MM}=\left(\frac{\Gamma(1+1/\alpha)}{\bar{X}}\right)^{\alpha}.$$

4

(b) First, when $\alpha = 1/2$, the MM estimator becomes:

meth seen ⇓

$$\widehat{\beta}_{MM} = \sqrt{\frac{2}{\bar{X}}}.$$

To obtain its asymptotic distribution, using the central limit theorem and considering the expectation and variance of Weibull distribution in the formulae sheets, we get

$$\sqrt{n}(\bar{X} - \frac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}) \xrightarrow{D} N(0, \frac{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)}{\beta^{2/\alpha}}),$$

which for $\alpha = 1/2$ becomes:

$$\sqrt{n}(\bar{X} - \frac{2}{\beta^2}) \xrightarrow{D} N(0, \frac{20}{\beta^4}).$$

Now, using the Delta method, we obtain the asymptotic distribution of the MM estimator of β, which is as follows

$$\sqrt{n}(\widehat{\beta}_{MM} - \beta) \xrightarrow{D} N(0, \frac{5\beta^2}{4}),$$

since $(g'(\theta))^2 = \frac{1}{2\theta^3}$ with $\theta = \frac{2}{\beta^2}$. 4

(c) First, the likelihood function is

sim. seen ⇓

$$L(\beta) = \alpha^n \beta^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\beta \sum\limits_{i=1}^{n} x_i^\alpha}.$$

Then, the log-likelihood is as follows

$$l(\beta) = n\log(\alpha) + n\log(\beta) + (\alpha-1)\sum_{i=1}^{n}\log x_i - \beta\sum_{i=1}^{n} x_i^\alpha.$$

By taking the derivative w.r.t β, we get the log-likelihood equation as follows

$$\frac{d}{d\beta}l(\beta) = \frac{n}{\beta} - \sum_{i=1}^{n} x_i^\alpha = 0,$$

which gives

$$\widehat{\beta}_{MLE} = \frac{n}{\sum\limits_{i=1}^{n} x_i^\alpha}.$$

Note that $\widehat{\beta}_{MLE}$ is indeed the maximiser of the log-likelihood function since

$$\frac{d^2}{d\beta^2}l(\beta) = \frac{-n}{\beta^2} < 0.$$

4

(d) From a theorem in the lecture notes we know that the MLEs are asymptotically normally distributed, and in fact for this question we have

meth seen ⇓

$$\sqrt{n}(\widehat{\beta}_{MLE} - \beta) \xrightarrow{D} N(0, \frac{1}{I(\beta)}),$$

where $I(\beta)$ is the Fisher information for one observation, which from part (b) is calculated as follows:

$$I(\beta) = -E(\frac{d^2}{d\beta^2}l(\beta)) = -E(\frac{-1}{\beta^2}) = \frac{1}{\beta^2}.$$

4

(e) When $\alpha = 1/2$, from Parts (b) and (d), the asymptotic relative efficiency of $\widehat{\beta}_{MLE}$ to $\widehat{\beta}_{MM}$ would be

meth seen ⇓

$$ARE(\widehat{\beta}_{MLE}, \widehat{\beta}_{MM}) = \frac{5\beta^2/4}{\beta^2} = \frac{5}{4} > 1,$$

and hence the maximum likelihood estimator $\widehat{\beta}_{MLE}$ is better.

4

unseen ⇓

3. (a) Since we observe Y, we need to obtain the posterior of θ given Y. For this, first note that $Y \sim \Gamma(\theta + 1, 1)$. Then, applying the Bayes theorem, we obtain the posterior distribution of θ given Y as follows

$$\pi(\theta|y) = \frac{p(1-p)^{\theta-1} y^\theta e^{-y}/\theta!}{\sum_{\theta=1}^{\infty} p(1-p)^{\theta-1} y^\theta e^{-y}/\theta!} = \frac{(1-p)^{\theta-1} y^\theta/\theta!}{\sum_{\theta=1}^{\infty} (1-p)^{\theta-1} y^\theta/\theta!},$$

and because, using the hint, the denominator is

$$\sum_{\theta=1}^{\infty} (1-p)^{\theta-1} y^\theta/\theta! = \frac{1}{(1-p)} \left(\sum_{\theta=0}^{\infty} \frac{(1-p)^\theta y^\theta}{\theta!} - 1 \right) = \frac{1}{(1-p)} \left(e^{(1-p)y} - 1 \right),$$

we get

$$\pi(\theta|y) = \frac{(1-p)^{\theta-1} y^\theta/\theta!}{\frac{1}{(1-p)} \left(e^{(1-p)y} - 1 \right)} = \frac{(1-p)^\theta y^\theta/\theta!}{e^{(1-p)y} - 1}.$$

8

(b) No, because the posterior distribution is not a geometric distribution anymore. 3

(c) Under the squared error loss, we know that the Bayes estimator is the posterior mean, which can be directly obtained as follows:

meth seen ⇓

$$\begin{aligned}
\widehat{\theta}_{\text{Bayes}} &= E(\theta|y) = \sum_{\theta=1}^{\infty} \theta \frac{(1-p)^\theta y^\theta/\theta!}{e^{(1-p)y} - 1} = \frac{1}{e^{(1-p)y} - 1} \sum_{\theta=1}^{\infty} \frac{(1-p)^\theta y^\theta}{(\theta-1)!} \\
&= \frac{(1-p)y}{e^{(1-p)y} - 1} \left(\sum_{\theta=1}^{\infty} \frac{(1-p)^{\theta-1} y^{\theta-1}}{(\theta-1)!} \right) = \frac{(1-p)y}{e^{(1-p)y} - 1} e^{(1-p)y}.
\end{aligned}$$

6

(d) Because the Bayes estimator obtained in part (c) is unique, therefore it is admissible using a theorem in the lecture notes.

meth seen ⇓

3

meth seen $\Downarrow$

4. (a) First note that

$$\frac{f_{H_1}(x)}{f_{H_0}(x)} = \frac{e^{(x-1)}}{e^{(x-0)}} \frac{\left(1+e^{(x-0)}\right)^2}{\left(1+e^{(x-1)}\right)^2} = e^{-1}\left(\frac{1+e^x}{1+e^{x-1}}\right)^2.$$

The above ratio is increasing in x because the derivative of the quantity in brackets is negative since

$$\frac{d}{dx}\left(\frac{1+e^x}{1+e^{x-1}}\right) = \frac{e^x(1+e^{x-1}) - e^{x-1}(1+e^x)}{(1+e^{x-1})^2} = \frac{e^x - e^{x-1}}{(1+e^{x-1})^2} > 0.$$

Now, using the Neyman-Person Lemma, the most powerful size α test rejects H_0 if $X > c$ and c is calculated such that

$$P_{H_0}(X > c) = \alpha \quad \Rightarrow \quad 1-F_{H_0}(x) = \alpha \quad \Rightarrow \quad 1-\frac{e^c}{(1+e^c)} = \alpha \quad \Rightarrow \quad c = \log\left(\frac{1-\alpha}{\alpha}\right).$$

$\boxed{5}$

(b) (i) First note that $X_{(1)}$ is both the restricted MLE ($\widehat{\mu}_{0MLE}$) and the unrestricted MLE ($\widehat{\mu}_{MLE}$) of μ. Also, $\widehat{\sigma}_{0MLE} = 1$ while the MLE of σ under the whole parameter space is

meth seen $\Downarrow$

$$\widehat{\sigma}_{MLE} = \frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{n}.$$

Now, the likelihood ratio statistic is as follows

$$\begin{aligned}
\lambda(x) &= \frac{L\left(\widehat{\mu}_{0MLE}, \widehat{\sigma}_{0MLE}\right)}{L\left(\widehat{\mu}_{MLE}, \widehat{\sigma}_{MLE}\right)} = \frac{e^{-\sum_{i=1}^{n}(x_i - x_{(1)})}}{\left(\sum_{i=1}^{n}(x_i - x_{(1)})/n\right)^n e^{-n}} \\
&= \left(\frac{n}{\sum_{i=1}^{n}(x_i - x_{(1)})}\right)^n e^{-\sum_{i=1}^{n}(x_i - x_{(1)}) + n}
\end{aligned}$$

$\boxed{5}$

(ii) Under H_0 and under regularity conditions, the asymptotic distribution of $-2\log(\lambda(x))$ is $\chi^2(1)$. Hence, the likelihood ratio level α test based on the asymptotic distribution rejects H_0 if $-2\log(\lambda(x)) \geq \chi^2_\alpha(1)$.

$\boxed{5}$

(iii) When $\mu = 0$, the distribution $f(x)$ becomes an exponential distribution with parameter σ. An approximate two-sided confidence interval based on the MLE of σ and its asymptotic normal distribution is as follows

meth seen $\Downarrow$

$$\left(\bar{X} - Z_{\alpha/2}\frac{1}{\sqrt{nI(\widehat{\sigma}_{MLE})}}, \bar{X} + Z_{\alpha/2}\frac{1}{\sqrt{nI(\widehat{\sigma}_{MLE})}}\right),$$

where

$$I(\sigma) = -E\left(\frac{d^2}{d\sigma^2}\log f_\sigma(X)\right) = -E\left(\frac{1}{\sigma^2} - \frac{2X}{\sigma^3}\right) = \frac{1}{\sigma^2}$$

and

$$\widehat{\sigma}_{MLE} = \frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{n}.$$

$\boxed{5}$

5. (a) Since the log-likelihood function is $\boxed{\text{unseen} \Downarrow}$

$$l(\alpha,\lambda) = n\alpha\log(\lambda) - n\log\Big(\Gamma(\alpha)\Big) + (\alpha-1)\sum_{i=1}^{n}\log(x_i) - \lambda\sum_{i=1}^{n}x_i,$$

we get the log-likelihood equations as follows

$$\begin{aligned}\tfrac{\partial}{\partial\alpha}l(\alpha,\lambda) &= n\log(\lambda) - \tfrac{n\Gamma'(\alpha)}{\Gamma(\alpha)} + \textstyle\sum_{i=1}^{n}\log(x_i) = 0,\\ \tfrac{\partial}{\partial\lambda}l(\alpha,\lambda) &= \tfrac{n\alpha}{\lambda} - \textstyle\sum_{i=1}^{n}x_i = 0.\end{aligned}$$

$\boxed{4}$

(b) From a theorem in the mastery notes, under mild regularity conditions, the maximum likelihood estimates $\widehat{\theta}_{MLE}$ of misspecified models converges to $\theta(F)$ (i.e., $\widehat{\theta}_{MLE} \xrightarrow{P} \theta(F)$) where $\theta(F)$ is the solution to the equation $\boxed{\text{unseen} \Downarrow}$

$$\int_{-\infty}^{\infty} l'(\theta;x)dF(x)dx = 0,$$

in which $l'(\theta;x)$ is the derivative of the log-likelihood (the score function) and F is the true distribution. $\boxed{4}$

Now, for this question because the true distribution is a Uniform distribution, we obtain, using the log-likelihood equations in part (a), that

$$\begin{aligned}&\tfrac{1}{b}\textstyle\int_0^b \Big(n\log(\lambda_0) - \tfrac{n\Gamma'(\alpha_0)}{\Gamma(\alpha_0)} + \sum_{i=1}^{n}\log(x_i)\Big)dx = 0\\ &\tfrac{1}{b}\textstyle\int_0^b \Big(\tfrac{n\alpha_0}{\lambda_0} - \sum_{i=1}^{n}x_i\Big)dx = 0\end{aligned}$$

which becomes

$$\begin{aligned}&\log(\lambda_0) - \tfrac{\Gamma'(\alpha_0)}{\Gamma(\alpha_0)} + \tfrac{1}{b}\textstyle\int_0^b \log(x)\,dx = 0,\\ &\tfrac{\alpha_0}{\lambda_0} - \tfrac{1}{b}\textstyle\int_0^b x\,dx = 0.\end{aligned}$$

$\boxed{4}$

(c) Using the invariance property of maximum likelihood estimates, the maximum likelihood estimate of the population mean is $\frac{\widehat{\alpha}_{MLE}}{\widehat{\lambda}_{MLE}}$, which converges to $\frac{\alpha_0}{\lambda_0}$. From the second equation in part (b), we have $\boxed{\text{unseen} \Downarrow}$

$$\frac{\alpha_0}{\lambda_0} = \frac{1}{b}\int_0^b x\,dx = \frac{b}{2},$$

and $\frac{b}{2}$ is indeed the true population mean. $\boxed{4}$

We conclude that correctly specifying the distribution is not a necessary condition for maximum likelihood estimate to converge to the true parameter value. $\boxed{4}$