Imperial College London

M3S1

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2016

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

Statistical Theory 1

Date: **Tuesday 10th May 2016**

Time: **09.30 – 11.30**

Time Allowed: **2 Hours**

This paper has Four Questions.

Candidates should use ONE main answer book.

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

Statistical tables are 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 the Neyman Factorization Criterion and prove it for the case of discrete distributions.

   (b) Suppose that $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$, where $0 < \theta < 1$.

      (i) Show that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ. Is T complete? Why? Also, argue whether or not T is minimal sufficient for θ.

      (ii) Find the UMVUE of $\theta(1-\theta)$.
      [Hint: $I(X_1 = 0, X_2 = 1)$ is an unbiased estimator of $\theta(1-\theta)$.]

      (iii) Compute the Cramér-Rao lower bound for the variance of unbiased estimators of $\theta(1-\theta)$. Does the UMVUE of $\theta(1-\theta)$ obtained in (ii) attain this lower bound? Why or why not?

2. Let $(X_1, Y_1),...,(X_n, Y_n)$ be independent pairs of Normal random variables where X_i and Y_i are independent $N(\mu_i, \sigma^2)$ random variables.

   (a) Find the MLEs of $\mu_1,...,\mu_n$ and σ^2.

   (b) Now, suppose we observe only $Z_1,...,Z_n$ where $Z_i = X_i - Y_i$.

      (i) Find the MLE of σ^2 based on $Z_1,...,Z_n$ and discuss whether or not it is consistent.

      (ii) Obtain a method of moments (MM) estimator of σ^2 based on $Z_1,...,Z_n$.

      (iii) Consider testing $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$. Find the UMP test at level α based on $Z_1,...,Z_n$.

      (iv) Is the UMP level α test obtained in (iii) unbiased? Justify your answer.

3. Let $X_1,\ldots,X_n$ be i.i.d. random variables from the delayed exponential distribution having the probability density function

$$f_\theta(x) = \theta e^{-\theta(x-2)}, \qquad x > 2,$$

where θ is unknown. Suppose that the prior distribution for θ is $\mathrm{Exponential}(\lambda)$ where λ is a known positive constant.

(a) Obtain the posterior distribution of θ.

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

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

(d) Verify whether or not the Bayes estimator obtained in (c) is admissible.

4. Suppose that $X_1,\ldots,X_m \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\theta_1)$ and $Y_1,\ldots,Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\theta_2)$, and assume the X_i and the Y_i are independent. Consider testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$.

(a) Show that the likelihood ratio test statistic is as follows

$$\lambda(x,y) = \left(\frac{m}{m+n} + \frac{n}{m+n}\frac{\bar{Y}}{\bar{X}}\right)^{-m}\left(\frac{n}{m+n} + \frac{m}{m+n}\frac{\bar{X}}{\bar{Y}}\right)^{-n}.$$

(b) Obtain a test at level α using the test statistic $\frac{\bar{X}}{\bar{Y}}$.
[Hint: Use the fact that if χ_1^2 and χ_2^2 are two independent chi-squared random variables with degrees of freedom v_1 and v_2 respectively, then $\frac{\chi_1^2/v_1}{\chi_2^2/v_2} \sim F(v_1,v_2)$.]

(c) Obtain the likelihood ratio test using the asymptotic distribution of $-2\log(\lambda(x,y))$ under H_0.

(d) Construct a confidence interval for $\frac{\theta_1}{\theta_2}$ with confidence coefficient $1-\alpha$.

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	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
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

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

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)	$\alpha<\beta\in\mathbb{R}$	$\dfrac{1}{\beta-\alpha}$	$\dfrac{x-\alpha}{\beta-\alpha}$	$\dfrac{(\alpha+\beta)}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	$\dfrac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)}$
$Exponential(\lambda)$ (standard model $\lambda=1$)	$\mathbb{R}^+$	$\lambda\in\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda-t}\right)$
$Gamma(\alpha,\beta)$ (standard model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$		$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta-t}\right)^\alpha$
$Weibull(\alpha,\beta)$ (standard model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$	$1-e^{-\beta x^\alpha}$	$\dfrac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}$	$\dfrac{\Gamma(1+2/\alpha)-\Gamma(1+1/\alpha)^2}{\beta^{2/\alpha}}$	
$Normal(\mu,\sigma^2)$ (standard model $\mu=0,\sigma=1$)	$\mathbb{R}$	$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu>1$)	$\dfrac{\nu}{\nu-2}$ (if $\nu>2$)	
$Pareto(\theta,\alpha)$	$\mathbb{R}^+$	$\theta,\alpha\in\mathbb{R}^+$	$\dfrac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$	$1-\left(\dfrac{\theta}{\theta+x}\right)^\alpha$	$\dfrac{\theta}{\alpha-1}$ (if $\alpha>1$)	$\dfrac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}$ (if $\alpha>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

Imperial College
London

M4S1

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2016

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

Statistical Theory 1

Date: **Tuesday 10th May 2016**

Time: **09.30 – 12.00**

Time Allowed: **2 Hours 30 Mins**

This paper has Five Questions.

Candidates should use ONE main answer book.

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

Statistical tables are 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 the Neyman Factorization Criterion and prove it for the case of discrete distributions.

   (b) Suppose that $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$, where $0 < \theta < 1$.

      (i) Show that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ. Is T complete? Why? Also, argue whether or not T is minimal sufficient for θ.

      (ii) Find the UMVUE of $\theta(1-\theta)$.
      [Hint: $I(X_1 = 0, X_2 = 1)$ is an unbiased estimator of $\theta(1-\theta)$.]

      (iii) Compute the Cramér-Rao lower bound for the variance of unbiased estimators of $\theta(1-\theta)$. Does the UMVUE of $\theta(1-\theta)$ obtained in (ii) attain this lower bound? Why or why not?

2. Let $(X_1, Y_1),...,(X_n, Y_n)$ be independent pairs of Normal random variables where X_i and Y_i are independent $N(\mu_i, \sigma^2)$ random variables.

   (a) Find the MLEs of $\mu_1,...,\mu_n$ and σ^2.

   (b) Now, suppose we observe only $Z_1,...,Z_n$ where $Z_i = X_i - Y_i$.

      (i) Find the MLE of σ^2 based on $Z_1,...,Z_n$ and discuss whether or not it is consistent.

      (ii) Obtain a method of moments (MM) estimator of σ^2 based on $Z_1,...,Z_n$.

      (iii) Consider testing $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$. Find the UMP test at level α based on $Z_1,...,Z_n$.

      (iv) Is the UMP level α test obtained in (iii) unbiased? Justify your answer.

3. Let $X_1,\ldots,X_n$ be i.i.d. random variables from the delayed exponential distribution having the probability density function

$$f_\theta(x) = \theta e^{-\theta(x-2)}, \qquad x > 2,$$

where θ is unknown. Suppose that the prior distribution for θ is $\mathrm{Exponential}(\lambda)$ where λ is a known positive constant.

(a) Obtain the posterior distribution of θ.

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

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

(d) Verify whether or not the Bayes estimator obtained in (c) is admissible.

4. Suppose that $X_1,\ldots,X_m \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\theta_1)$ and $Y_1,\ldots,Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\theta_2)$, and assume the X_i and the Y_i are independent. Consider testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$.

(a) Show that the likelihood ratio test statistic is as follows

$$\lambda(x,y) = \left(\frac{m}{m+n} + \frac{n}{m+n}\frac{\bar{Y}}{\bar{X}}\right)^{-m}\left(\frac{n}{m+n} + \frac{m}{m+n}\frac{\bar{X}}{\bar{Y}}\right)^{-n}.$$

(b) Obtain a test at level α using the test statistic $\frac{\bar{X}}{\bar{Y}}$.
[Hint: Use the fact that if χ_1^2 and χ_2^2 are two independent chi-squared random variables with degrees of freedom v_1 and v_2 respectively, then $\frac{\chi_1^2/v_1}{\chi_2^2/v_2} \sim F(v_1,v_2)$.]

(c) Obtain the likelihood ratio test using the asymptotic distribution of $-2\log(\lambda(x,y))$ under H_0.

(d) Construct a confidence interval for $\frac{\theta_1}{\theta_2}$ with confidence coefficient $1-\alpha$.

Mastery Question:

5. Let $X_1,\ldots,X_n$ be i.i.d. Cauchy random variables with density function

$$f_\theta(x) = \frac{1}{\pi\left(1 + (x-\theta)^2\right)} \qquad x \in R,$$

and suppose outcomes $x_1,\ldots,x_n$ are observed.

(a) Write down the likelihood equation for estimating θ and discuss whether it has a unique solution for the given sample $x_1,\ldots,x_n$.

(b) Given an estimate $\hat{\theta}^{(k)}$ for θ at iteration k, obtain a new estimate $\hat{\theta}^{(k+1)}$ using the Newton-Raphson method.

(c) Show that a new estimate $\hat{\theta}^{(k+1)}$ using the Fisher scoring algorithm is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{4}{n}\sum_{i=1}^{n}\frac{x_i - \hat{\theta}^{(k)}}{1 + \left(x_i - \hat{\theta}^{(k)}\right)^2}.$$

[Hint: $\int_0^\infty \frac{1-t^2}{(1+t^2)^3}dt = \frac{\pi}{8}$.]

(d) Is the sample mean $\bar{x}$ an appropriate initial value for the Newton-Raphson and the Fisher scoring methods here? If not, suggest a good starting point. Briefly explain your thinking.

(e) Which method has a faster convergence: the Newton-Raphson method or the Fisher scoring algorithm? Why?

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	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
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

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

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 2016

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

M3S1/M4S1

Statistical Theory I

Date: Tuesday, 10th May 2016 Time: 9:30 – 12:00

Solutions

1. (a) *Neyman Factorization Criterion:* Suppose that $X = (X_1, ..., X_n)$ has a joint distribution $f_\theta(x)$. Then $T = T(X)$ is a sufficient statistic for θ if and only if $f_\theta(x) = g(T(x), \theta)\, h(x)$. $\boxed{\text{seen} \Downarrow}$ $\boxed{2}$

Proof: First suppose that T is sufficient. Then

$$\begin{aligned} f_\theta(x) &= P_\theta(X = x) = \sum_t P_\theta(X = x, T = t) = P_\theta(X = x, T = T(x)) \\ &= P_\theta(T = T(x)) P_\theta(X = x | T = T(x)) = g(T(x), \theta) h(x) \end{aligned}$$

Now, suppose that $f_\theta(x) = g(T(x), \theta)\, h(x)$. Then, if $T(x) = t$,

$$\begin{aligned} P_\theta(X = x | T = t) &= \frac{P_\theta(X = x)}{P_\theta(T = t)} = \frac{P_\theta(X = x)}{\sum\limits_{T(y)=t} P_\theta(X = y)} \\ &= \frac{g(T(x), \theta) h(x)}{\sum\limits_{T(y)=t} g(T(y), \theta) h(y)} = \frac{h(x)}{\sum\limits_{T(y)=t} h(y)} \end{aligned}$$

which does not depend on θ. If $T(x) \neq t$, then $P_\theta(X = x | T = t) = 0$. In both cases, $P_\theta(X = x | T = t)$ is independent of θ and so T is sufficient. $\boxed{3}$

(b) (i) According to Neyman Factorization Criterion, since

$$f_\theta(x) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i},$$

$T = \sum_{i=1}^n X_i$ is a sufficient statistic for θ. $\boxed{2}$

Yes, T is complete because Bernoulli distribution is a member of "full rank" exponential family of distributions since $f_\theta(x) = \exp\left(\ln\left(\frac{\theta}{1-\theta}\right) \sum_{i=1}^n x_i + n \ln(1-\theta) \right)$. $\boxed{2}$

Also, T is minimal sufficient because it is a complete and sufficient statistic. $\boxed{1}$

(ii) An unbiased estimator of $\theta(1-\theta)$ is $I(X_1 = 0, X_2 = 1)$. By the Lehmann- $\boxed{\text{sim. seen} \Downarrow}$

Scheffe Theorem, the UMVUE of $\theta(1-\theta)$ is

$$
\begin{aligned}
E(I(X_1=0,X_2=1)|T=t) &= P(X_1=0,X_2=1|\sum_{i=1}^{n}X_i=t) \\
&= \frac{P(X_1=0,X_2=1,\sum_{i=1}^{n}X_i=t)}{P(\sum_{i=1}^{n}X_i=t)} \\
&= \frac{P(X_1=0,X_2=1,\sum_{i=3}^{n}X_i=t-1)}{P(\sum_{i=1}^{n}X_i=t)} \\
&= \frac{\theta(1-\theta)\binom{n-2}{t-1}\theta^{t-1}(1-\theta)^{n-2-t+1}}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t}} \\
&= \frac{\binom{n-2}{t-1}}{\binom{n}{t}} = \frac{t(n-t)}{n(n-1)}.
\end{aligned}
$$

Therefore, $\frac{\sum_{i=1}^{n}X_i\left(n-\sum_{i=1}^{n}X_i\right)}{n(n-1)}$ is the UMVUE of $\theta(1-\theta)$. $\boxed{5}$

(iii) The Cramer-Rao lower bound here is

$$
\frac{\left(\frac{d}{d\theta}\left(\theta(1-\theta)\right)\right)^2}{I(\theta)} = \frac{(1-2\theta)^2}{nI_{X_1}(\theta)} = \frac{(1-2\theta)^2}{\frac{n}{\theta(1-\theta)}} = \frac{\theta(1-\theta)\,(1-2\theta)^2}{n}.
$$

$\boxed{3}$

Only estimators of the form $\left\{a\sum_{i=1}^{n}X_i+b\right\}$ achieve the Cramer-Rao lower bound. So the variance of the UMVUE of $\theta(1-\theta)$ does not attain the lower bound. $\boxed{2}$

$\boxed{\text{meth seen} \Downarrow}$

2. (a) The log-likelihood function is

$$
l(\mu_1,\dots,\mu_n,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(x_i-\mu_i)^2+\sum_{i=1}^{n}(y_i-\mu_i)^2\right\}
$$

and the MLEs of the parameters are

$$
\widehat{\mu}_i = \frac{X_i+Y_i}{2} \qquad \text{and} \qquad \widehat{\sigma}^2 = \frac{1}{n}\left\{\sum_{i=1}^{n}\left(X_i-\frac{X_i+Y_i}{2}\right)^2+\sum_{i=1}^{n}\left(Y_i-\frac{X_i+Y_i}{2}\right)^2\right\}.
$$

$\boxed{6}$

(b) (i) Since $Z_i \overset{\text{i.i.d.}}{\sim} N(0, 2\sigma^2)$, the log-likelihood function based on $Z_1, \dots, Z_n$ is

$$l(\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{4\sigma^2}\sum_{i=1}^{n} z_i^2$$

and then the MLE of σ^2 is $\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} Z_i^2$. $\boxed{3}$

MLEs are consistent under the regularity conditions. Because normal distribution satisfies the regularity conditions, the MLE $\hat{\sigma}^2$ of σ^2 is consistent. $\boxed{2}$

(ii) Since $E(Z^2) = 2\sigma^2$, method of moments (MM) estimator of σ^2 is

$$2\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 \qquad \Rightarrow \qquad \hat{\sigma}^2_{MM} = \frac{1}{2n}\sum_{i=1}^{n} Z_i^2.$$

$\boxed{2}$

(iii) The family $\{N(0, 2\sigma^2) : \sigma^2 > 0\}$ has monotone likelihood ratio in $\sum_{i=1}^{n} z_i^2$. Using the Karlin-Rubin Theorem, the UMP test at level α is

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} z_i^2 \geq k \\ 0 & \text{if } \sum_{i=1}^{n} z_i^2 < k \end{cases}$$

where k can be chosen so that

$$\alpha = P_{\sigma_0^2}(\sum_{i=1}^{n} Z_i^2 \geq k) = P(\chi^2(n) \geq \frac{k}{2\sigma_0^2}).$$

We then get $k = 2\sigma_0^2\chi^2_\alpha(n)$. $\boxed{5}$

(iv) The power of the UMP test obtained in (iii) is non-decreasing in θ because of the monotone likelihood ratio property. So the UMP test here is an unbiased test since its power is not less than α. $\boxed{2}$

meth seen $\Downarrow$

3. (a) Applying the Bayes theorem, we can write the posterior distribution as follows

$$\pi(\theta|x) = \frac{\lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^{n}(x_i-2)}}{\int_0^\infty \lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^{n}(x_i-2)}d\theta} = \frac{\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{\int_0^\infty \theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}d\theta}$$
$$= \frac{\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{c}$$

where c is a constant which does not depend on θ. Because the posterior distribution is proportional to $\theta^n e^{(-K\theta)}$, where K is $\sum_{i=1}^{n}(x_i-2)+\lambda$, the posterior is $Gamma(n+1,K)$. In fact $\theta|x \sim Gamma\left(n+1,\sum_{i=1}^{n}(x_i-2)+\lambda\right)$.
Alternatively, one can obtain the above posterior distribution straightforwardly by calculation of the integral in the denominator (i.e., the constant c) as follows:

$$\pi(\theta|x) = \frac{\lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^{n}(x_i-2)}}{\int_0^\infty \lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^{n}(x_i-2)}d\theta} = \frac{\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{\int_0^\infty \theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}d\theta}$$
$$= \frac{\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{\frac{1}{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}E\left(\left(\text{Exponential}\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)\right)^n\right)}$$
$$= \frac{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{\frac{n!}{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)^n}}$$
$$= \frac{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)^{n+1}\theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{n!}$$

which is again $Gamma\left(n+1,\sum_{i=1}^{n}(x_i-2)+\lambda\right)$. $\boxed{8}$

(b) Yes, because both the prior and the posterior are Gamma distributions. Note that exponential distribution is a special case of Gamma distribution. $\boxed{3}$

(c) Under the squared error loss, the Bayes estimator is the posterior mean. Because the posterior is Gamma distribution, we can easily obtain

$$\widehat{\theta}_{\text{Bayes}} = \frac{n+1}{\sum_{i=1}^{n}(x_i-2)+\lambda}.$$

The above Bayes estimator can alternatively be obtained as follows:

$$\begin{aligned}
\widehat{\theta}_{\text{Bayes}} &= E(\theta|x) = \int_0^\infty \theta \left(\frac{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)^{n+1} \theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{n!} \right) d\theta \\
&= \frac{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)^n}{n!} E\left(\left(Exponential(\sum_{i=1}^n (x_i-2)+\lambda)\right)^{n+1}\right) \\
&= \frac{n+1}{\sum_{i=1}^n (x_i-2)+\lambda}.
\end{aligned}$$

6

(d) Because the Bayes estimator obtained in (c) is unique, therefore it is admissible. 3

meth seen ⇓

4. (a) Under the whole parameter space, the MLEs of θ_1 and θ_2 are $\widehat{\theta}_{1MLE} = \frac{1}{\bar{X}}$ and $\widehat{\theta}_{2MLE} = \frac{1}{\bar{Y}}$, respectively. And under H_0, the MLEs of θ_1 and θ_2 are

$$\left(\widehat{\theta}_1\right)_0 = \left(\widehat{\theta}_2\right)_0 = \frac{m+n}{m\bar{X}+n\bar{Y}}.$$

Hence, the likelihood ratio statistic is as follows

$$\begin{aligned}
\lambda(x,y) &= \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \frac{L\left(\left(\widehat{\theta}_1\right)_0, \left(\widehat{\theta}_2\right)_0\right)}{L\left(\widehat{\theta}_{1MLE}, \widehat{\theta}_{2MLE}\right)} \\
&= \frac{\left(\left(\widehat{\theta}_1\right)_0\right)^m e^{-\left(\widehat{\theta}_1\right)_0 \sum_{i=1}^m x_i} \left(\left(\widehat{\theta}_1\right)_0\right)^n e^{-\left(\widehat{\theta}_1\right)_0 \sum_{i=1}^n y_i}}{\left(\widehat{\theta}_{1MLE}\right)^m e^{-\widehat{\theta}_{1MLE} \sum_{i=1}^m x_i} \left(\widehat{\theta}_{2MLE}\right)^n e^{-\widehat{\theta}_{2MLE} \sum_{i=1}^n y_i}} \\
&= \left(\frac{m}{m+n} + \frac{n}{m+n}\frac{\bar{Y}}{\bar{X}}\right)^{-m} \left(\frac{n}{m+n} + \frac{m}{m+n}\frac{\bar{X}}{\bar{Y}}\right)^{-n}.
\end{aligned}$$

7

(b) We know the likelihood ratio test rejects H_0 for small values of $\lambda(x,y)$. Now, because $\lambda(x,y)$ depends only on $T = \frac{\bar{X}}{\bar{Y}}$ and we can make $\lambda(x,y)$ small by making T small or T large, so a test based on $T = \frac{\bar{X}}{\bar{Y}}$ would reject H_0 for small or large values of T. In fact, a level α test based on $T = \frac{\bar{X}}{\bar{Y}}$ rejects $H_0: \theta_1 = \theta_2$ if $T \le c_1$ or $T \ge c_2$, where c_1 and c_2 can be chosen so that $P_{H_0}(T \le c_1) + P_{H_0}(T \ge c_2) = \alpha$, where the distribution of $T = \frac{\bar{X}}{\bar{Y}}$, under H_0, is $F(2m, 2n)$. By considering equal tails of the F distribution, we can reject H_0 if $T \le F_{1-\alpha/2}(2m,2n)$ or $T \ge F_{\alpha/2}(2m,2n)$. 5

(c) Under H_0 and under regularity conditions, the asymptotic distribution of $-2\log(\lambda(x,y))$ is $\chi^2(1)$. $\boxed{2}$
The likelihood ratio level α test based on the asymptotic distribution rejects H_0 if $-2\log(\lambda(x,y)) \leq \chi^2_{1-\alpha}(1)$. $\boxed{2}$

(d) From (b), we have

$$P_{H_0:\theta_1=\theta_2}\left(F_{1-\alpha/2}(2m,2n) \leq \frac{\theta_1\bar{X}}{\theta_2\bar{Y}} \leq F_{\alpha/2}(2m,2n)\right) = 1-\alpha,$$

and hence using the connection between confidence intervals and hypothesis tests, a confidence interval for $\frac{\theta_1}{\theta_2}$ with confidence coefficient $1-\alpha$ is $\left(\frac{\bar{Y}}{\bar{X}}F_{1-\alpha/2}(2m,2n), \frac{\bar{Y}}{\bar{X}}F_{\alpha/2}(2m,2n)\right)$. $\boxed{4}$

5. (a) The likelihood equation is

part seen ⇓

$$S(\widehat{\theta}) = \sum_{i=1}^{n} \frac{2(x_i - \widehat{\theta})}{1 + \left(x_i - \widehat{\theta}\right)^2} = 0.$$

2

Because $S(\theta)$ is not monotone in θ, the equation $S(\widehat{\theta}) = 0$ may have more than one solution for given sample $x_1,...,x_n$.

2

(b) Using the Newton-Raphson method a new estimate is given by

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{S(\widehat{\theta}^{(k)})}{H(\widehat{\theta}^{(k)})}$$

where $S(\theta)$ is given in (a) and

$$H(\theta) = 2\sum_{i=1}^{n} \frac{1 - (x_i - \theta)^2}{\left(1 + (x_i - \theta)^2\right)^2}.$$

4

(c) The Fisher scoring algorithm gives a new estimate as follows

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{S(\widehat{\theta}^{(k)})}{H^*(\widehat{\theta}^{(k)})}$$

where $H^*(\theta) = E(H(\theta))$. Considering the hint, we get

$$\begin{aligned} H^*(\theta) &= E\left(2\sum_{i=1}^{n} \frac{1 - (x_i - \theta)^2}{\left(1 + (x_i - \theta)^2\right)^2}\right) = 2n\int_{-\infty}^{\infty} \frac{1 - (x_i - \theta)^2}{\pi\left(1 + (x_i - \theta)^2\right)^3} dx_i \\ &= \frac{4n}{\pi}\int_{0}^{\infty} \frac{1 - (x_i - \theta)^2}{\left(1 + (x_i - \theta)^2\right)^3} dx_i = \frac{4n}{\pi}\left(\frac{\pi}{8}\right) = \frac{n}{2}, \end{aligned}$$

and hence

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{4}{n}\sum_{i=1}^{n} \frac{x_i - \widehat{\theta}^{(k)}}{1 + \left(x_i - \widehat{\theta}^{(k)}\right)^2}.$$

5

(d) Because $E(X_i)$ is not well-defined, the sample mean may not be a good initial estimate of θ.

2

Since the density of the X_is is symmetric around θ, one may use the sample median as an initial estimate.

2

(e) The convergence of the Newton-Raphson algorithm is often faster (when both algorithms converge) because it uses the observed Fisher information rather than the expected Fisher information which needs integral calculation.

3