

McGill University

Faculty of Science

MATH 557

MATHEMATICAL STATISTICS II

Final Examination

Date: 14th April 2008

Time: 2pm-5pm

Examiner: Dr. David A. Stephens

Associate Examiner: Dr. Masoud Asgharian

Please write your answers in the answer booklets provided.

This paper contains five questions. Each question carries 20 marks. Credit will be given for all questions attempted. The total mark available is 100 but rescaling of the final mark may occur.

Candidates may take one double-sided sheet of Letter-sized (216×279 mm) or A4-sized (210×297 mm) paper with handwritten notes into the examination room.

Calculators may be used. A formula sheet is provided.

Dictionaries and Translation dictionaries are permitted.

1. (a) If $T(\underline{x}) = \sum_{i=1}^n t(x_i)$, the pmf/pdf takes the form

$$f_{\underline{X}|\theta}(\underline{x}|\theta) = h(\underline{x})\{c(\theta)\}^n \exp\{\theta T(\underline{x})\}$$

2 MARKS

- (b) $T(\underline{X})$ is a complete minimal sufficient statistic.

- *sufficient*: a statistic is sufficient if the joint pdf/likelihood of $X_1, \dots, X_n$ depends on θ only through $T(\underline{X})$.
- *minimal sufficient*: a statistic is minimal sufficient if it is a function of all other sufficient statistics.
- *complete*: A statistic T is complete if $E_{f_T}[g(T)] = 0$ implies $g(t) \equiv 0$ for all t .

3 MARKS

- (c) (i) Not EF: support depends on θ

$$f_{\underline{X}|\theta}(\underline{x}|\theta) = \left(\frac{1}{2\theta}\right)^n I_{(0, \max\{-X_{(1)}, X_{(n)}\})}$$

and the minimal sufficient statistic is $T(\underline{X}) = \max\{-X_{(1)}, X_{(n)}\}$.

3 MARKS

- (ii) Not EF: support depends on θ

$$f_{\underline{X}|\theta}(\underline{x}|\theta) = \exp\left\{-\sum_{i=1}^n x_i - n\theta\right\} I_{(0, X_{(1)})}$$

and the minimal sufficient statistic is $T(\underline{X}) = X_{(1)}$.

3 MARKS

- (iii) Not EF: not correct form

$$f_{\underline{X}|\theta}(\underline{x}|\theta) = \left(\frac{2}{1+2\theta}\right)^n \prod_{i=1}^n (x_i + \theta)$$

and the minimal sufficient statistic is $T(\underline{X}) = (X_1, \dots, X_n)^\top$.

3 MARKS

- (iv) EF:

$$f_{\underline{X}|\theta}(\underline{x}|\theta) = \left(\frac{\log \theta}{\theta - 1}\right)^n \exp\left\{\log \theta \sum_{i=1}^n x_i\right\}$$

and the minimal sufficient statistic is $T(\underline{X}) = \sum_{i=1}^n X_i$.

3 MARKS

- (d) By the Neyman-Factorization theorem, up to proportionality

$$\pi_{\theta|\underline{X}}(\theta|\underline{x}) \propto f_{\underline{X}|\theta}(\underline{x}|\theta)\pi_{\theta}(\theta) = g(\theta; T(\underline{x}))h(\underline{x})\pi_{\theta}(\theta) \propto g(\theta; T(\underline{x}))\pi_{\theta}(\theta)$$

3 MARKS

2. (a) Suppose that $X_1, \dots, X_n$ is a sample of random variables from probability model described by pmf/pdf $f_{X|\theta}$, with likelihood $L(\theta|\underline{x})$. Let $T(\underline{X})$ be an unbiased estimator of θ . Then $T(\underline{X})$ attains the Cramér-Rao lower bound if and only if, for some function $a(\theta)$,

$$a(\theta)(T(\underline{X}) - \theta) = \frac{\partial}{\partial \theta} \log L(\theta|\underline{X})$$

6 MARKS

- (b) (i) If $T(\underline{x}) = \sum_{i=1}^n x_i$, the likelihood is

$$L(\theta|\underline{X}) = \theta^{T(\underline{X})} e^{-n\theta} \left\{ \prod_{i=1}^n X_i! \right\}^{-1} \quad \therefore \quad \frac{\partial}{\partial \theta} \log L(\theta|\underline{X}) = \frac{T(\underline{X})}{\theta} - n = \frac{n}{\theta} (\bar{X} - \theta)$$

and so $\bar{X}$ is UMVUE.

2 MARKS

- (ii) If $\phi = \theta(\theta + 2)$, then $\theta = -1 + \sqrt{1 + \phi}$, as

$$\mathcal{I}(\theta) = -\mathbb{E}_{f_{X|\theta}}[\Psi(\theta, X)] = -\mathbb{E}_{f_{X|\theta}}[-X/\theta^2] = 1/\theta$$

so

$$\mathcal{I}(\phi) = \frac{\mathcal{I}(\theta)}{\{\dot{\phi}(\theta)\}^2} = \frac{1}{\theta} \frac{1}{4(\theta + 1)^2} = \frac{1}{4(1 + \phi)(\sqrt{1 + \phi} - 1)}$$

3 MARKS

- (iv) We have

$$\frac{\partial}{\partial \phi} \log L(\phi|\underline{X}) = \frac{\partial}{\partial \theta} \log L(\theta|\underline{X}) \frac{\partial \theta}{\partial \phi} = \frac{T(\underline{X}) - n\theta}{2\theta(\theta + 2)} = \frac{T(\underline{X}) - n(\sqrt{1 + \phi} - 1)}{2\phi}$$

so by the result from (a), the bound is never attained.

4 MARKS

- (v) Consider the estimator

$$T(\underline{X}) = \bar{X}^2 + \left(\frac{2n-1}{n} \right) \bar{X}.$$

Then

$$\begin{aligned} \mathbb{E}_{f_{T|\theta}}[T|\theta] &= \frac{1}{n^2} \mathbb{E}_{f_{\underline{X}|\theta}} \left[\left(\sum_{i=1}^n X_i \right)^2 \right] + \left(\frac{2n-1}{n} \right) \theta \\ &= \frac{1}{n^2} (n\theta + n^2\theta^2) + \theta = \theta^2 + \frac{\theta}{n} + \left(\frac{2n-1}{n} \right) \theta = \theta(\theta + 2) \end{aligned}$$

The Cramér-Rao bound is, from part (ii)

$$\frac{1}{n\mathcal{I}(\phi)} = \frac{4\theta(1 + \theta)^2}{n}$$

which we know is not achieved here.

5 MARKS

3. (a) **The Rao-Blackwell Theorem**

Let T be an unbiased estimator of $\tau(\theta)$, and S be a sufficient statistic for θ . Define statistic U by

$$U \equiv g(S) = E_{f_{T|S,\theta}}[T|S]$$

Then U is an unbiased estimator of $\tau(\theta)$, and for all θ $\text{Var}_{f_{U|\theta}}[U] \leq \text{Var}_{f_{T|\theta}}[T]$.

Proof: Clearly $U = g(S)$ is a valid estimator, as it does not depend on the θ ; the conditional distribution of T given S does not depend on θ by sufficiency. By iterated expectation,

$$E_{f_{U|\theta}}[U] = E_{f_{S|\theta}}[g(S)] = E_{f_{S|\theta}}[E_{f_{T|S,\theta}}[T|S]] = E_{f_{T|\theta}}[T] = \tau(\theta)$$

so U is unbiased for $\tau(\theta)$, and similarly

$$\begin{aligned} \text{Var}_{f_{T|\theta}}[T] &= E_{f_{S|\theta}}[\text{Var}_{f_{T|S,\theta}}[T|S]] + \text{Var}_{f_{S|\theta}}[E_{f_{T|S,\theta}}[T|S]] \\ &\geq \text{Var}_{f_{S|\theta}}[E_{f_{T|S,\theta}}[T|S]] = \text{Var}_{f_{S|\theta}}[g(S)] = \text{Var}_{f_{U|\theta}}[U] \end{aligned}$$

and thus U is a better estimator of $\tau(\theta)$ than T , as it has lower variance.

8 MARKS

(b) (i) The likelihood is

$$L(\theta|x, y) = \theta^2 \exp\{-\theta(x + y)\} = \theta^2 \exp\{-\theta s\}$$

so S is a sufficient statistic, and $S \sim \text{Gamma}(2, \theta)$, with cdf

$$F_{S|\theta}(s|\theta) = 1 - (1 + \theta s)e^{-\theta s} \quad s > 0$$

2 MARKS

(ii) We have for $x > 0$,

$$\begin{aligned} E_{f_{T_x|\theta}}[T_x|\theta] &= \int_0^\infty T_x(s) f_{S|\theta}(s|\theta) ds = \int_0^x f_{S|\theta}(s|\theta) ds + \int_x^\infty \frac{x}{s} f_{S|\theta}(s|\theta) ds \\ &= 1 - (1 + \theta x)e^{-\theta x} + x \int_x^\infty \theta^2 e^{-\theta s} ds \\ &= 1 - (1 + \theta x)e^{-\theta x} + x\theta e^{-\theta x} = 1 - e^{-\theta x} = \tau(\theta). \end{aligned}$$

4 MARKS

(iii) By the second corollary to the Rao-Blackwell theorem, $T_x(S)$ is a function of a complete sufficient statistic, and therefore is unique.

3 MARKS

(iv) Part (iii) implies that $T_x(S)$ is UMVUE.

3 MARKS

4. (a) **Neyman-Pearson Lemma:** Consider a parametric model $f_{X|\theta}(x|\theta)$ with parameter space $\Theta = \{\theta_0, \theta_1\}$. A test of

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

is required. Consider a test $\mathcal{T}$ with rejection region $\mathcal{R}$ that satisfies

$$\begin{aligned} f_{\underline{X}|\theta}(\underline{x}|\theta_1) &> k f_{\underline{X}|\theta}(\underline{x}|\theta_0) \implies \underline{x} \in \mathcal{R} \\ f_{\underline{X}|\theta}(\underline{x}|\theta_1) &< k f_{\underline{X}|\theta}(\underline{x}|\theta_0) \implies \underline{x} \in \mathcal{R}' \end{aligned}$$

for some $k \geq 0$, and $\Pr[\underline{X} \in \mathcal{R}|\theta = \theta_0] = \alpha$. Then $\mathcal{T}$ is UMP in the class, $\mathcal{C}_\alpha$, of tests at level α . Further, if such a test exists with $k > 0$, then **all** tests at level α also have size α (that is, α is the least upper bound of the power function $\beta(\theta)$), and have rejection region identical to that of $\mathcal{T}$, except perhaps if $\underline{x} \in A$ and

$$\Pr[\underline{X} \in A|\theta = \theta_0] = \Pr[\underline{X} \in A|\theta = \theta_1] = 0.$$

4 MARKS

- (b) Take $\theta_1 > 1/2$. Then if $T(\underline{x}) = \sum_{i=1}^n x_i^2$,

$$\frac{f_{\underline{X}|\theta}(\underline{x}|\theta_1)}{f_{\underline{X}|\theta}(\underline{x}|1/2)} = \frac{2^n \theta_1^n e^{-\theta_1 T(\underline{x})}}{e^{-T(\underline{x})/2}} = 2\theta_1 \exp\{-(\theta_1 - 1/2)T(\underline{x})\}$$

so

$$\frac{f_{\underline{X}|\theta}(\underline{x}|\theta_1)}{f_{\underline{X}|\theta}(\underline{x}|1/2)} > k \iff T(\underline{x}) < -\frac{1}{(\theta_1 - 1/2)} \log \frac{k}{2\theta_1} = k_1$$

say. Now, for this model, if $Y = X^2$, then by elementary methods, $f_{Y|\theta}(y|\theta) = \theta e^{-\theta y}$ therefore $Y \sim \text{Exp}(\theta)$, and $T(\underline{x}) \sim \text{Gamma}(n, \theta)$. Thus the α -level rejection region is $\mathcal{R} \equiv (0, q_{n,\theta}(\alpha))$, where $q_{n,\theta}(\alpha)$ is the α quantile of the $\text{Gamma}(n, \theta)$.

6 MARKS

- (c) We have

$$\dot{l}(\theta|\underline{x}) = n/\theta - T(\underline{x}) \quad \therefore \quad \hat{\theta}_n = n/T(\underline{x}).$$

Now $Y \sim \text{Exp}(\theta)$, so by the CLT

$$\sqrt{n}(T(\underline{x})/n - 1/\theta) \xrightarrow{d} Z \sim \text{Normal}(0, 1/\theta^2)$$

and thus by the Delta Method, with $g(t) = 1/t$, $\dot{g}(t) = -1/t^2$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} Z \sim \text{Normal}(0, \theta^2)$$

and thus

$$\hat{\theta}_n \sim AN(\theta, \theta^2/n)$$

6 MARKS

Thus an approximate $(1 - \alpha)$ confidence interval for θ is

$$\hat{\theta}_n \pm \frac{1.96\hat{\theta}_n}{\sqrt{n}}$$

4 MARKS

5. (a) First, note that if $\theta \neq \theta_0$,

$$T_n(\underline{x}, \theta_0, \theta) = \frac{1}{n} \frac{l_n(\theta)}{l_n(\theta_0)} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{f_{X|\theta}(x|\theta)}{f_{X|\theta}(x|\theta_0)} \right\}$$

then, as $n \rightarrow \infty$, by the weak law of large numbers (WLLN), say,

$$T_n(\underline{X}, \theta_0, \theta) \xrightarrow{p} E_{f_{X|\theta}} \left[\log \frac{f_{X|\theta}(X|\theta)}{f_{X|\theta}(X|\theta_0)} \right] = \int \log \left\{ \frac{f_{X|\theta}(x|\theta)}{f_{X|\theta}(x|\theta_0)} \right\} f_{X|\theta}(x|\theta_0) dx = -K(\theta_0, \theta) < 0$$

where $K(\theta_0, \theta)$ is the Kullback-Leibler divergence between the pdfs with parameters θ_0 and θ . Hence, by A1, $T_n(\underline{X}, \theta_0, \theta)$ converges to something negative. Thus, for all $\theta \neq \theta_0$,

$$\Pr[L(\theta_0|\underline{X}) > L(\theta|\underline{X})|\theta_0] \rightarrow 1 \quad (1)$$

as $n \rightarrow \infty$; with probability converging to 1, the likelihood at θ_0 is greater than the likelihood elsewhere in Θ .

8 MARKS

- (b) (i) Yes; all the regularity conditions are met in this model.

3 MARKS

- (ii) For this model, the sample median $\tilde{X}_n$ is a consistent and inefficient estimator.

2 MARKS

- (iii) The one-step procedure is as follows:

- Find a consistent but possibly inefficient estimator $\tilde{\theta}_n$
- Form the one-step Newton or Scoring Estimator using the formulae

$$\hat{\theta}^\dagger = \tilde{\theta}_n - \left(\ddot{l}_n(\tilde{\theta}_n) \right)^{-1} \dot{l}_n(\tilde{\theta}_n) \quad \hat{\theta}^\star = \tilde{\theta}_n + \left(\mathcal{I}(\tilde{\theta}_n) \right)^{-1} \frac{1}{n} \dot{l}_n(\tilde{\theta}_n)$$

- These estimators have the same asymptotic properties as solutions to the likelihood equations; under the regularity conditions, if $\hat{\theta}_n$ satisfies the likelihood equation,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$$

and $\hat{\theta}^\dagger$ and $\hat{\theta}^\star$ also have these properties. For the logistic model

$$\hat{\theta}^\dagger = \tilde{X}_n - \frac{\dot{l}_n(\tilde{X}_n)}{\ddot{l}_n(\tilde{X}_n)} = \tilde{X}_n + \frac{2 \sum_{i=1}^n \frac{1}{1 + e^{\tilde{X}_n - X_i}} - n}{\sum_{i=1}^n \frac{e^{\tilde{X}_n - X_i}}{(1 + e^{\tilde{X}_n - X_i})^2}}$$

4 MARKS

- (iv) This family is a location family, so the result follows by manipulation of the integral that computes $\mathcal{I}(\theta)$ to render it independent of θ .

3 MARKS