THE UNIVERSITY OF NEW SOUTH WALES

DEPARTMENT OF STATISTICS

MATH3811/MATH3911- Statistical Inference/Higher Statistical Inference

Assignment 1

A reminder that assignments count as part of your assessment. Please, write a declaration on the first page stating that the assignment is your won work, except where acknowledged. State also that you have read and understood the University Rules in respect to Student Academic Misconduct.

This assignment must be submitted to you tutor by 2pm, Friday, April 8, 2011 **at the latest.** It should represent your *individual* work.

Math3811: Attempt the first three questions. Math3911: Attempt all four questions.

1) A preliminary test of a possible carcinogenic compound can be performed by measuring the mutation rate of microorganisms exposed to the compound. An experimenter places the compound in 25 petri dishes and records the following number of mutant colonies:

$$6, 5, 6, 3, 1, 1, 2, 3, 5, 2, 5, 3, 3, 1, 3, 3, 1, 3, 3, 1, 3, 5, 4, 2, 2$$

The number X of mutant colonies is believed to follow a Poisson (λ) distribution ($P_\lambda(X = x) = f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots$).

a) Work out the Cramer-Rao lower bound for the variance of an unbiased estimator of $\tau(\lambda) = \lambda e^{-\lambda}$ (that is, of the probability that exactly one colony will emerge). Is this variance bound attainable by any unbiased estimator of $\tau(\lambda)$? Give reasons for your answer.

b) Derive the UMVUE of $\tau(\lambda) = \lambda e^{-\lambda}$.

c) What is the MLE $\hat{\tau}$ of $\tau(\lambda) = \lambda e^{-\lambda}$? What is the asymptotic distribution of this MLE (that is, state precisely the asymptotic distribution of $\sqrt{n}(\hat{\tau} - \tau(\lambda))$.

d) Using the given data, evaluate numerically the probability $\tau(\lambda)$ by substituting the estimators you have derived in (b) and (c). Compare the two numerical values. Comment.

e) Find an asymptotic 90% confidence interval for $\tau(\lambda)$ using the asymptotic distribution in c) and the available data.

2) The discrete random variable X takes values according to one of the following distribution types ($\theta \in \Theta = (0, 0.1)$):

	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$	$P(X = 4)$
Distribution I	θ	$\theta - \theta^4$	$2\theta^3 - \theta^4$	$1 + 2\theta^4 - 2\theta^3 - 2\theta$
Distribution II	$2\theta^2$	$2\theta^2$	$2\theta^3 - \theta^4$	$1 + \theta^4 - 4\theta^2 - 2\theta^3$

In each of the two cases determine whether the family of distributions $\{P_\theta(X)\}, \theta \in \Theta$ is complete. Give reasons.

3) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from a population with $f(x;\theta) = \begin{cases} e^{(\theta - x)}, & \text{if } x \geq \theta \\ 0 \text{ else} \end{cases}$

a) Show that the smallest $X_{(1)}$ is a minimal sufficient statistic for θ.

b) Show that the density of $X_{(1)}$ is given by $f_{X_{(1)}}(y;\theta) = \begin{cases} ne^{n(\theta - y)}, & \text{if } y \geq \theta \\ 0 \text{ else} \end{cases}$

and hence show that $EX_{(1)} = \theta + \frac{1}{n}$.

Hint: You may use the fact that $P(X_{(1)} \leq y) = 1 - [P(X_1 \geq y)]^n$ to find the cumulative distribution function of $X_{(1)}$ for a start.

c) Show that $X_{(1)}$ is complete. Hence find the UMVUE of θ.

4) For a fixed natural number k you can model the probability of the number of failures before the kth success in a series of independent repetitions of Bernoulli trials with probability of success $\theta \in (0,1)$ as

$$f(x;\theta) = \begin{pmatrix} k+x-1 \\ x \end{pmatrix} \theta^k (1-\theta)^x, x = 0, 1, 2, \ldots.$$

Let $\mathbf{X} = (X_1, X_2, , \ldots, X_n)$ be independent discrete random variables with the above discrete distribution.

a) Argue that $\mathbf{T} = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for θ.

b) Show that $E(X_1) = \frac{k(1-\theta)}{\theta}$.

c) Using a) and b), or otherwise, validate the claim that the information in $\mathbf{T}$ about θ is $I_T(\theta) = \frac{nk}{\theta^2(1-\theta)}$.

d) Determine the UMVUE of $\tau(\theta) = \frac{1-\theta}{\theta}$ and justify your answer. Does the variance of your UMVUE attain the Cramer-Rao bound for the variance of an unbiased estimator of $\tau(\theta)$? (You do not need to actually calculate the variance).

e) Argue that $\mathbf{T} = \sum_{i=1}^{n} X_i$ is complete. Show that $I_{\{X_1=0\}}(\mathbf{X})$ is an unbiased estimator of $\eta(\theta) = \theta^k$. Using these facts, or otherwise, find the UMVUE of $\eta(\theta) = \theta^k$.