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 own work, except where acknowledged. State also that you have read and understood the University Rules in respect of Student Academic Misconduct

This assignment must be submitted to your tutor by Thursday, 3pm, April 5, 2012 **at the latest.**

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

1) In a manufacturing process in which glass items are being produced, defects in a form of bubbles occur, and if there are many of them, the item is undesirable for marketing. The X number of bubbles is thought to follow a Poisson (λ) distribution ($P_\lambda(X = x) = f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots$). Let a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of such Poisson random variables be given.

a) Suggest a simple unbiased estimator W of the parameter $\tau(\lambda) = \frac{\lambda^3 e^{-\lambda}}{6} = P(X_1 = 3)$.

b) Use any argument to show that $T = \sum_{i=1}^n X_i$ is complete and minimal sufficient for λ. Hence derive the UMVUE of $\tau(\lambda)$.

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

d) Defects have been counted on 15 glass items giving the following result

$$5, 4, 3, 3, 1, 1, 2, 4, 5, 0, 0, 7, 2, 5, 3.$$

Find a point estimate of the probability $\tau(\lambda)$ using the data given by applying the methods in (b) and (c). Compare the two numerical values. Comment.

e) Is the variance of the proposed estimator in (b) equal to the variance given by the Cramer-Rao lower bound for an unbiased estimator of $\tau(\lambda)$? Explain. Calculate the Cramer-Rao lower bound explicitly.

f) 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	3θ	2θ	$2\theta^3 - \theta^4$	$1 + \theta^4 - 5\theta - 2\theta^3$
Distribution II	θ	$\theta - \theta^3$	$2\theta^2$	$1 - 2\theta - 2\theta^2 + \theta^3$

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 with $f(x;\theta) = \begin{cases} \frac{3\theta^3}{x^4}, \text{ if } x \geq \theta \\ 0 \text{ else} \end{cases}$, $\theta > 0$ being unknown parameter.

a) Sketch a graph of a density from this family for a fixed θ.

b) Find the cumulative distribution function $F(x;\theta)$ of X_1.

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

d) Show that the density of $X_{(1)}$ is given by $f_{X_{(1)}}(y;\theta) = \begin{cases} \frac{3n\theta^{3n}}{y^{3n+1}}, & \text{if } y \geq \theta \\ 0 & \text{else} \end{cases}$

and hence find $EX_{(1)}$.

Hint: You may use $P(X_{(1)} \leq y) = 1 - P(X_1 \geq y)]^n$ and b) to find the cumulative distribution function of $X_{(1)}$ first.

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

4. i) Let $\mathbf{X} = (X_1, X_2, \dots X_n)$ be a sample of n i.i.d observations from Bernoulli distribution with probability of success $\theta \in (0,1)$ ($f_{X_i}(x) = \theta^x(1-\theta)^{1-x}, x = \{0,1\}$). Show that no unbiased estimator of $\tau(\theta) = 1/\theta$ exists.

ii) Suppose that you are given a chance to either observe the above sample $\mathbf{X}$ or alternatively to sample n i.i.d. observations $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ each with geometric distribution:

$$f_{Y_i}(y) = \theta(1-\theta)^{y-1}, y = 1, 2, \dots,$$

i.e., each Y_i denotes the number of trials until success occurs (for a first time). Calculate and compare the Fisher informations $I_{\mathbf{X}}(\theta)$ and $I_{\mathbf{Y}}(\theta)$. If you are given a choice whether to sample $\mathbf{Y}$ or $\mathbf{X}$ and n is large enough, which sampling scheme would you prefer? (The criterion is to make a more precise inference about $\tau(\theta) = \theta$.) Explain your answer.

iii) For the geometric observations scheme, set $\theta = h(\eta) = \frac{1}{\eta}$ and parametrize the model with η instead of θ. Express $I_{\mathbf{Y}}(\eta)$ by using $I_{\mathbf{Y}}(\theta)$.

(*) Can you generalize this result to a statement saying how the information changes when a smooth parameter transformation is performed?

iv) (*) Determine the UMVUE for the parameter θ in the geometric observations scheme. Justify your answer.