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 Wednesday, 11am, April 17, 2013 **at the latest.**

Math3811: Attempt the first 4 questions. Math3911: Attempt all 5 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 $n = 25$ petri dishes and records the following number of mutant colonies:
6, 5, 5, 3, 1, 1, 2, 3, 5, 2, 5, 3, 3, 1, 3, 3, 1, 3, 3, 1, 3, 6, 4, 1, 2

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

a) Suggest a simple unbiased estimator W of the parameter $\tau(\lambda) = \frac{\lambda^2 e^{-\lambda}}{2} = P(X_1 = 2)$ (i.e., the probability that exactly two colonies will emerge).

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) For the given data, find the point estimates of the probability $\tau(\lambda)$ 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.

f) Find an asymptotic 95% 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	2θ	$2\theta^3 - \theta^4$	$1 + \theta^4 - 3\theta - 2\theta^3$
Distribution II	θ	$\theta - \theta^3$	$3\theta^2$	$1 - 2\theta - 3\theta^2 + \theta^3$

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 $f(x; \theta) = \begin{cases} e^{(\theta - x)}, & \text{if } x \geq \theta \\ 0 \text{ else} \end{cases}$, θ being unknown parameter.

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

b) Find the cumulative distribution function $F_{X_{(1)}}(y; \theta)$ and the density $f_{X_{(1)}}(y; \theta)$ of the smallest of the observations, $X_{(1)}$.

Hint: you could use $P(X_{(1)} \leq y) = 1 - P(X_1 \geq y, X_2 \geq y, \ldots, X_n \geq y) = 1 - [P(X_1 \geq y)]^n$.

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

d) Show that $EX_{(1)} = \theta + \frac{1}{n}$ holds.

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

4) Let $\mathbf{X} = (X_1, X_2, \ldots X_n)$ be a sample of n i.i.d observations from Bernoulli distribution with probability of success $\theta \in (0,1)$ ($f_{X_i}(x) = \theta^x(1-\theta)^{1-x}, x = \{0,1\}$). The odds-ratio $\tau(\theta) = \frac{\theta}{1-\theta}, 0 < \theta < 1$ is an important characteristics of this distribution.

a) Show that no unbiased estimator of $\tau(\theta)$ exists.

Hint: You could utilize the fact that as θ approaches 0, the odds- ratio can be larger than any finite constant $M > 0$.

b) The MLE of the odds-ratio is well-defined. Find the MLE, state its asymptotic distribution and suggest an asymptotic confidence interval at level 90% for the odds-ratio.

5) a) Suppose a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of i.i.d. random variables with a common density $f(x,\theta) = \theta x^{(\theta-1)}(0 < x < 1)$ is given, where the parameter $\theta > 0$ is unknown.

i) Find a minimal sufficient statistic T for θ.

ii) Show that $W = -\log X_1$ is an unbiased estimator of $\tau(\theta) = \frac{1}{\theta}$. Also argue that $\sqrt{W}$ is a biased estimator of $\xi(\theta) = \frac{1}{\sqrt{\theta}}$ and that, more precisely, $E\sqrt{W} < \frac{1}{\sqrt{\theta}}$ holds.

iii) Specify the distribution of W.

iv) Find the UMVUE of $\tau(\theta) = \frac{1}{\theta}$. Justify your answer.

(*) v) Find the UMVUE of $\bar{\tau} = \theta$. Justify your answer.

b) For a sample of n i.i.d. observations $X_1, X_2, \ldots, , X_n$ from $N(\mu, \theta^2)$ distribution with known μ, argue that the UMVUE of $\theta > 0$ is

$$\sqrt{\frac{n}{2}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)}S$$

where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

(*) Formulate and prove the corresponding statement for the UMVUE of $\theta > 0$ in the same model as above but now assuming that μ is unknown.