THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF MATHEMATICS

FINAL EXAMINATION

JUNE 2007

MATH3911
HIGHER STATISTICAL INFERENCE

(1) TIME ALLOWED – 2 1/2 Hours

(2) TOTAL NUMBER OF QUESTIONS – 5

(3) ANSWER ALL QUESTIONS

(4) THE QUESTIONS ARE **NOT** OF EQUAL VALUE

(5) TOTAL NUMBER OF MARKS – 125

(6) THIS PAPER MAY BE RETAINED BY THE CANDIDATE

All answers must be written in ink. Except where they are expressly required pencils may only be used for drawing, sketching or graphical work.

PPS

1. **[28 marks]** Let $X = (X_1, X_2, \ldots, X_n)$ $(n \geq 2)$ be i.i.d. random variables having Poisson distribution with density $f(x,\lambda) = \frac{e^{-\lambda}.\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$

a) Show that the product of indicators $I_{\{x_1=0\}}(X) \cdot I_{\{x_2=0\}}(X)$ is an unbiased estimator of the parameter $\tau(\lambda) = e^{-2\lambda}$.

b) Given that $T = \sum_{i=1}^{n} X_i$ is complete and minimal sufficient for λ, derive the UMVUE of $\tau(\lambda) = e^{-2\lambda}$. (**Hint:** you may use part (a)).

c) Does the variance of the UMVUE in b) attain the Cramer-Rao bound for the minimal variance of an unbiased estimator of $\tau(\lambda) = e^{-2\lambda}$? Give reasons for your answer.

d) Suppose now that for the same sample, the parameter of interest is $h(\lambda) = \sqrt{\lambda}$. Find the MLE of $\sqrt{\lambda}$, state its asymptotic distribution and, using this result, suggest a confidence interval for λ that asymptotically has a level $1-\alpha$. Explain why $h(\lambda) = \sqrt{\lambda}$ is called "variance stabilising transformation" and why such a transformation it useful.
Hint: For any smooth function $h(\lambda)$:

$$\sqrt{n}(h(\hat{\lambda}_{mle}) - h(\lambda_0)) \xrightarrow{d} N(0, (\frac{\partial h}{\partial \lambda}(\lambda_0))^2 I^{-1}(\lambda_0)),$$

$$I(\lambda) = E\{\frac{\partial}{\partial \lambda}[lnf(x,\lambda)]\}^2 = E\{-\frac{\partial^2}{\partial \lambda^2}[lnf(x,\lambda)]\}.$$

e) Let $n = 6$. Find the uniformly most powerful test of size $\alpha = 0.1$ for testing $H_0 : \lambda \leq 0.25$ against $H_1 : \lambda > 0.25$. Justify your answer.

2. **[24 marks]**

a) A photocopier company kept records of the number of breakdowns per year for 100 photocopiers, giving the results

Number of breakdowns	0	1	2	3	≥ 4
Frequency	8	20	37	24	11

Test at the 5% level that the above data comes from a Poisson $(\lambda = 2)$ distribution.

b) Two companies manufacture white board pens. A sample of 10 pens from one company and 8 pens from the other were tested until they were no longer readable. The results (in hours) were:

Company 1	8.9	9.9	12	7.4	6.6	8.4	7.6	10.9	11.7	10.2
Company 2	9.1	13.3	8.2	11.1	10	11.6	10.5	8.7		

Using the Wilcoxon Test for large samples, test at the 5% level the hypothesis that there is no difference between the two company's pens.

c) The following table represents data on cortisol level of women in three different groups: Group 1 delivered their babies by Caesarean section before the onset of labor, Group 2 delivered by emergency Caesarean during induced labor, and group 3 experienced spontaneous labor.

Group 1	262	307	211	323	454	339	304	154	287	356
Group 2	465	501	455	355	468	362				
Group 3	343	772	207	1048	838	687				

Use the Kruskal-Wallis test at level 0.05 to test for equality of the average cortisol level in the three groups.

3. **[23 marks]** Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample of size n from a distribution with density

$$f(x;\theta) = \begin{cases} 3\theta^3/x^4, & x \geq \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

a) Prove that the density of $\mathbf{X}$ has a monotone likelihood ratio in $X_{(1)}$.

b) Find the cumulative distribution function of $Z = X_{(1)}$. Show that the density is

$$f_Z(z;\theta) = \begin{cases} \frac{3n\theta^{3n}}{z^{3n+1}} & , Z \geq \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

c) Find the uniformly most powerful α-size test φ^* of $H_0 : \theta \leq 2$ versus $H_1 : \theta > 2$ and sketch a graph of $E\varphi^*$ as precisely as possible.

4. **[28 marks]** In a Bayesian framework, the probability of a defect $\theta \in (0,1)$ of a certain piece of a machine's output appears to be varying on various days and can be represented as a random variable with prior density

$$\tau(\theta) = 12\theta^2(1-\theta), \quad \theta \in (0,1).$$

On a certain day, $n = 16$ observations denoted $X = (X_1, X_2, \ldots, X_{16})$ are made where

$$X_i = \begin{cases} 1 \text{ if the piece is defective} \\ 0 \text{ if the piece is nondefective} \end{cases}$$

Hence $f(X|\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$.

a) Find the formula for the Bayesian estimator (with respect to quadratic loss) of the probability θ and calculate its estimated value if from the 16 pieces examined on this day, exactly three were defective.
Hint: You may use: $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$, $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, $\Gamma(\alpha) = \int_0^\infty \exp(-x)x^{\alpha-1}dx$, $\quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

b) Find the Bayesian estimator (with respect to quadratic loss) of the parameter $\eta(\theta) = \theta(1-\theta)$.

c) Using the data of part (a) (3 defective out of 16 tested), test the hypothesis $H_0 : \theta \leq 0.25$ against $H_1 : \theta > 0.25$ using the Bayesian approach and a 0-1 loss. Do you accept H_0?
(You may use: $\int_0^{0.25} \theta^5(1-\theta)^{14}d\theta = 0.00000164$).

5. **[22 marks]** Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be i.i.d., each with uniform $[0, \theta)$ density $(\theta > 0)$.

a) Find $Cov(X_{(n-1)}, X_{(n)})$.

b) Find the density of the range $R = X_{(n)} - X_{(1)}$. Using it, or otherwise, show that $ER = \frac{n-1}{n+1}\theta$ holds.

Some useful formulae

1) Wilcoxon: Two independent samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$, $W_{m+n} = \sum_{i=1}^{m} R(X_i)$. Then $\frac{W_{m+n} - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}$ has approximately a standard normal distribution.

2) The Kruskal-Wallis statistic for a K sample problem:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{K} n_i (\bar{R}_{i.} - \frac{n+1}{2})^2 = \frac{12}{n(n+1)} \sum_{i=1}^{K} \frac{R_{i.}^2}{n_i} - 3(n+1)$$

has χ^2_{K-1} as a limiting distribution under the null hypothesis.

3) The rth order statistic $(r = 1, 2, \ldots, n)$ of the sample of size n from a distribution with a density $f_X(.)$ and a cdf $F_X(.)$:

$$f_{X_{(r)}}(y_r) = \frac{n!}{(r-1)!(n-r)!} [F_X(y_r)]^{r-1} [1 - F_X(y_r)]^{n-r} f_X(y_r)$$

The joint density of the pair $(X_{(i)}, X_{(j)}), 1 \le i < j \le n$ is:

$$f_{X_{(i)}, X_{(j)}}(x, y) =$$

$$\frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(x) f_X(y) [F_X(x)]^{i-1} [F_X(y) - F_X(x)]^{j-1-i} [1 - F_X(y)]^{n-j}$$

for $x < y$.

4) Bayesian inference:

$$h(\theta|\mathbf{X}) = \frac{f(\mathbf{X}|\theta)\tau(\theta)}{g(\mathbf{X})}, \quad g(\mathbf{X}) = \int_{\Theta} f(\mathbf{X}|\theta)\tau(\theta) d\theta.$$