Data provided:
"Statistics Tables" by H.R. Neave

MAS274

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2009-2010**

Statistical Reasoning **2 hours**

RESTRICTED OPEN BOOK EXAMINATION.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator that conforms to University regulations.
All answers will be marked but credit will be given for only the best ***THREE*** *answers.*
All questions carry equal marks. Total marks 99.

1 The positive random variables X_1, X_2, $\ldots X_n$ are independent observations from the following density function

$$\frac{x^2}{2\eta^3} e^{-x/\eta} \qquad (x > 0)$$

which depends on the unknown parameter $\eta \in (0, \infty)$.

(a) Give the relationship between this distribution and one of the standard ones used in the course. Give the mean and variance of the distribution in terms of η. ***(5 marks)***

(b) Show that $\overline{X}/3$ is the maximum likelihood estimator of η, where $\overline{X}$ is the average of the X's. ***(11 marks)***

(c) Show that the k-unit likelihood region is the set of η for which

$$\left(\frac{\hat{\eta}}{\eta} - 1\right) - \log\left(\frac{\hat{\eta}}{\eta}\right) \leqslant \frac{k}{3n},$$

where $\hat{\eta}$ is the MLE of η. ***(4 marks)***

(d) Show that the maximum likelihood estimator of η is unbiased and that its variance is $\eta^2/(3n)$. ***(6 marks)***

(e) Give the maximum likelihood estimator of η^2. Find the bias of this estimator. Propose an unbiased estimator of η^2. ***(7 marks)***

2 The observations $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed and

$$Y_i \sim Ge\left(\frac{e^{-\gamma}}{1+e^{-\gamma}}\right),$$

where $\gamma \in (-\infty, \infty)$ is an unknown parameter. [Thus Y_i counts the number of failures **before** the first success in Bernoulli trials.]

(a) (i) Show that $E\overline{Y} = e^{\gamma}$, where $\overline{Y}$ is the average of the Y's. ***(3 marks)***

(ii) Show that the log-likelihood is

$$n\left[-(\overline{y}+1)\log(1+e^{-\gamma}) - \gamma\right],$$

justifying your manipulations carefully. ***(6 marks)***

(iii) Show that the (Fisher) information concerning γ is

$$\frac{n}{(1+e^{-\gamma})}.$$

Hence, give the asymptotic distribution of the MLE of γ.

(15 marks)

(b) (i) Use the Neyman-Pearson Lemma to show that the form of the best test of the null hypothesis that $\gamma = 0$ against the alternative that $\gamma = \gamma_1 > 0$ (for some specific γ_1) is 'Reject $\gamma = 0$ if $\overline{Y} \geqslant k^*$' — again you should justify your manipulations carefully. ***(7 marks)***

(ii) Suppose that the null hypothesis $\gamma = 0$ is to be tested against $\gamma > 0$. Show that there is a uniformly most powerful test. ***(2 marks)***

3 Suppose that $Y \sim N(\Theta, 5)$ and $Y = \log X$. Suppose also that $Y_1 = \log X_1$, $Y_2 = \log X_2$, $\ldots$, $Y_n = \log X_n$ are independent identically distributed random variables each with the same distribution as Y. Suppose that the prior beliefs about Θ are that Θ is $N(2, 1)$.

(a) Give EX in terms of Θ. The scientist is concerned with assessing the evidence for $EX > 99.5$. Express this as a statement about Θ.

(6 marks)

(b) If the values of the Y's are y_1, y_2, $\ldots$, y_n, give the posterior distribution for Θ. ***(5 marks)***

(c) Suppose $n = 3$, $y_1 = 4.1$, $y_2 = 7.7$ and $y_3 = 6.2$. Describe how these data change the degree of belief about the hypothesis that $EX > 99.5$.

(14 marks)

(d) Using the data just given, the scientist wants to offer a point inference about EX. He says that the loss in making the inference b is $(b - EX)^2$. What point estimate should be used? Justify the form of the estimate you propose. ***(8 marks)***

4 The random variables X_i, $i = 1, 2, \ldots, n$, are independent and

$$X_i \sim N\left(0, \frac{1}{2\nu a_i^2}\right),$$

where the a_i are known constants and ν is an unknown parameter. Let

$$T(\mathbf{x}) = \sum_{i=1}^{n} (a_i x_i)^2.$$

(a) Show that

$$\nu^{n/2} \exp(-\nu T(\mathbf{x}))$$

is a version of the likelihood, justifying your manipulations carefully. Explain why it is described as 'a version of the likelihood'. ***(10 marks)***

(b) Obtain the MLE of ν. ***(5 marks)***

(c) Suppose the prior on ν is $Ga(b+1, c)$, with $b > 0$ and $c > 0$ known constants.

(i) Show that the mode of the prior is b/c. ***(3 marks)***

(ii) Find the posterior for ν and give the posterior mode (in terms of $T(\mathbf{x})$, b and c). ***(8 marks)***

(iii) Indicate with a sketch what is meant in this context by the highest density $\alpha\%$ credible region for the posterior. What value is in this interval for every $\alpha > 0$? ***(4 marks)***

(d) Show that the posterior mode can be written as

$$w \times \text{prior mode} + (1 - w) \times \text{MLE}$$

for $w = c/(c + T(\mathbf{x}))$. ***(3 marks)***

End of Question Paper