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

The University Of Sheffield.

MAS274

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2010–2011**

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.
Attempt all questions [Q1 16; Q2 26; Q3 19; Q4 39].

1 X is a random variable which has the Poisson distribution with mean λ.

(a) (i) Give the log-likelihood and the MLE for λ based on this single observation. ***(2 marks)***

(ii) Obtain the two-unit likelihood region for λ. Show that if the observation is $x = 4$ then the two-unit likelihood region is

$$\{\lambda : (\lambda - 4\log\lambda) \leqslant c\}$$

for a suitable c. Find the numerical value of c. ***(4 marks)***

(b) Suppose now that $X_1,\ X_2,\ \ldots X_n$ are independent and identically distributed random variables with the same distribution as X.

(i) Give the MLE for λ. [You do not have to derive the MLE.] ***(1 mark)***

(ii) Give the asymptotic distribution of the MLE for λ. [You do not have to derive the distribution.] ***(2 marks)***

(iii) Give the MLE of $\phi = \sqrt{\lambda}$. Derive its asymptotic distribution. ***(5 marks)***

(iv) Give an approximate 95% confidence interval for ϕ. ***(2 marks)***

2 (a) (i) In hypothesis testing, explain what is meant by a critical region. ***(3 marks)***

(ii) Explain the usefulness of the Neyman-Pearson Lemma when a test is needed of one simple hypothesis against another. ***(4 marks)***

(iii) Explain carefully in what situation one might say that the null hypothesis is rejected at the 5% level but not at the 1% level. ***(4 marks)***

(b) Suppose that X_1, X_2, $\ldots$, X_n are independent identically distributed with the $Be(1, 1+\theta)$ distribution.

(i) Show that the density function of X_1 is $(\theta+1)(1-x_1)^\theta$ when $x_1 \in (0,1)$. ***(4 marks)***

(ii) Obtain the likelihood for the data. ***(3 marks)***

(iii) Let

$$h(\mathbf{x}) = \sum_{i=1}^{n} \log(1-x_i).$$

Find, in terms of $h(\mathbf{x})$, the form of the critical region for the likelihood ratio test of the simple null hypothesis that $\theta = 1$ against the simple alternative that $\theta = \theta_1 > 1$. ***(6 marks)***

(iv) Explain why the test you have obtained is uniformly most powerful of its size for testing the null hypothesis that $\theta = 1$ against the alternative that $\theta > 1$. ***(2 marks)***

3 (a) X_1, X_2, $\ldots$, X_n are independent observations from a normal distribution with mean zero and variance $(2\nu)^{-1}$. The prior distribution of ν is taken to be $Ga(a, b)$, where a and b are to be chosen later to reflect the views of the investigator.

(i) Write down the likelihood for ν. ***(3 marks)***

(ii) Write down the prior density for ν. ***(3 marks)***

(iii) Derive the posterior distribution for ν. ***(3 marks)***

(iv) Suppose the experimenter declares a prior mean of $1/2$ and a prior standard deviation of $1/4$. Find suitable values for a and b. Suppose now that four observations are taken, giving $x_1 = 0.4$, $x_2 = 3.8$, $x_3 = 0.7$, $x_4 = 2.1$. Find the posterior distribution, and hence the posterior mean and variance. ***(6 marks)***

(b) Explain with a suitable sketch what is meant by the 95% highest density credible interval for a unimodal posterior density on $(0, \infty)$. ***(4 marks)***

4 (a) Let R be binomial on n trials with success probability θ. Derive the MLE of θ and use your knowledge of standard distributions to give its asymptotic distribution. ***(10 marks)***

(b) A manufacturer produces bolts to the same specification on three machines, each of which produces at the same rate. The outputs from the three machines are combined in a way that mixes them thoroughly. As part of the monitoring process the manufacturer tests bolts to see whether they are within a certain tolerance of the specification. Normally, the bolts are all well within tolerance so when some bolts lie outside it will be because one of the machines needs resetting. How much the machine needs resetting depends, among other things, on what proportion, γ, of its output is now outside the tolerance. A sample of n_1 from the whole output produces r_1 that are outside tolerance. Having established which machine is the culprit, a further sample of n_2 from its output alone produces r_2 outside tolerance.

(i) What model would you propose for the number of faulty bolts in the second sample? Give the MLE of γ based on the second sample (of n_2) alone and give its asymptotic distribution. ***(3 marks)***

(ii) In the first sample, what is the probability of a bolt being outside tolerance? What model would you propose for the number of faulty bolts in the first sample? Give the MLE of γ based on the first sample (of n_1) alone and give its asymptotic distribution. ***(7 marks)***

(iii) Obtain the likelihood for the data as a whole (both samples) and hence obtain the coefficients (in terms of r_1, r_2, n_1, n_2) in the quadratic equation satisfied by the MLE of γ. By considering the quadratic at 0 and at 1 show that it has exactly one root in $(0, 1)$. Show that that root is the MLE. ***(12 marks)***

(iv) Suppose now that $n_1 = 3n_2$. Obtain the asymptotic distribution of the MLE estimator obtained in (iii). ***(7 marks)***

End of Question Paper