RESTRICTED OPEN BOOK EXAMINATION
Data provided: Neave's Statistical Tables

MAS371

The University Of Sheffield.

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

Applied Probability **2 hours**

Restricted Open Book Examination.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator which conforms to University regulations.
Marks will be awarded for your best ***three*** *answers. Total marks 60.*

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 The times (in days) before light-bulbs fail are given by i.i.d. random variables $X_1, X_2 \ldots$, each having an exponential distribution with rate parameter λ (and hence mean $1/\lambda$).

(i) Given a sample of such observations $x_1, \ldots, x_n$ (with $x_i \geq 0$ for all i), show that the log-likelihood $l(\lambda | x_1, \ldots, x_n)$ can be written in terms of n and the sample mean $\bar{x}$, and that the maximum likelihood estimate of the failure rate λ is $1/\bar{x}$. ***(4 marks)***

(ii) If $n = 25$ and $\bar{x} = 400$, show that $(0.001644, 0.003613)$ forms a likelihood interval for λ. If this is interpreted as a confidence interval for λ, calculate its approximate confidence level. With $n = 100$, and $\bar{x}$ unchanged, show that a corresponding interval is given by $(0.002042, 0.003023)$. ***(6 marks)***

(iii) Derive the form of the observed information $J(\lambda)$. By evaluating $J(\hat{\lambda})$, calculate alternative approximate confidence intervals for λ in the two cases given in (ii), at the same confidence level as in (ii). Comment on how the approximations compare, and what this suggests about the shape of the log-likelihood. ***(10 marks)***

2 A simple weather record classifies the weather on each day as Rainy, Snowy or Dry. The table below shows the numbers of pairs of days on which type of weather is followed by each possible type.

		Weather on day $t+1$		
		Dry	Rainy	Snowy
Weather	Dry	29	11	7
on day t	Rainy	9	22	6
	Snowy	8	5	3

(i) Assuming that the type of weather at this location follows a 3-state Markov chain, where state 1 represents Dry weather, state 2 represents Rainy, and state 3 represents Snowy, calculate maximum likelihood estimates of each of the transition probabilities. ***(4 marks)***

(ii) If the transition matrix has the constrained form

$$\begin{pmatrix} 1-2w & w & w \\ d & (1-d)q & (1-d)(1-q) \\ d & (1-d)(1-q) & (1-d)q \end{pmatrix}$$

show that the log-likelihood for (w, d, q) given all the data has the form

$$\begin{aligned} & n_{11}\log(1-2w) + (n_{12}+n_{13})\log(w) \\ + \quad & (n_{21}+n_{31})\log(d) + (n_{22}+n_{23}+n_{32}+n_{33})\log(1-d) \\ + \quad & (n_{22}+n_{33})\log(q) + (n_{23}+n_{32})\log(1-q). \end{aligned}$$

(3 marks)

(iii) With reference to the parameters and likelihood in (ii), explain what aspects of the dynamics of the weather could be inferred from data from an automated weather recorder that was unable to distinguish between rainy and snowy weather, if the model (ii) is correct. ***(4 marks)***

(iv) Using the full data set from (i), obtain maximum likelihood estimates of the parameters (w, d, q) in the model from (ii), and hence the restricted estimate of the transition matrix.

Explain carefully how you would test whether the model in (ii) is appropriate for the weather at the location where those data were collected, against the alternative of the more general Markov chain in (i). (You are not required to actually carry out the test.) ***(9 marks)***

3 The number of female animals in a territory evolves over time as follows. If there is at least one female present, then one 'dominant' female has the chance to breed; the probability that she produces one female offspring in any interval $(t, t + \delta t)$ is $\alpha \delta t + o(\delta t)$, and otherwise she produces none. (Male offspring are ignored in this model.) Any female, regardless of age or dominance, has a chance $\beta \delta t + o(\delta t)$ of dying in any time interval $(t, t + \delta t)$. If the number of females reaches zero, then the territory will be recolonised by a single female, after a time which has an exponential distribution with mean $1/\gamma$.

(i) Explain why this model can be formulated as a Generalized Birth-Death process, and specify the birth and death rates. ***(4 marks)***

(ii) Write down the log-likelihood for α, β, γ based on a complete record of birth, deaths and recolonizations over a time interval $(0, t)$. Show that the log-likelihood can be simplified to

$$l(\alpha, \beta, \gamma) = f \log(\alpha) + d \log(\beta) + n \log(\gamma) - \alpha(1 - p)t - \beta m t - \gamma p t$$

in terms of the following summaries of the full record: the length of observation, t; the proportion of time for which the territory is unoccupied, p; the time-average of the population, m; the number of recolonisation events n; the total number of female offspring produced, f; and the total number of deaths d. ***(8 marks)***

(iii) Derive maximum likelihood estimates of α, β, γ in terms of the summaries given in (ii). Give the approximate covariance matrix of your estimates, and comment briefly. ***(8 marks)***

4 Theory predicts that emissions detected from a radioactive sample should follow an inhomogeneous Poisson process with rate

$$\lambda(t) = \exp(\alpha - \beta t)$$

for some constants α, β, with $\beta > 0$.

(i) Given a record of the times $\{v_i, i = 1, \ldots, n\}$ of emissions over a time interval $(0, t)$, show that the log-likelihood for α and β has the form

$$(e^{-\beta t} - 1)(e^{\alpha}/\beta) + n\alpha - \beta \sum_{i=1}^{n} v_i.$$

Hence obtain the likelihood equations that must be satisfied by the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$. (You are *not* required to calculate the maximum likelihood estimates themselves.) ***(9 marks)***

(ii) If a specific value β_0 is hypothesized for the parameter β, derive the restricted maximum likelihood estimate $\tilde{\alpha}$ for α under a hypothesis of this form. ***(2 marks)***

(iii) In a particular experiment, a sample was observed for $t = 24$ hours; $n = 87$ emissions were observed, at times (measured in hours from the start of the experiment) summarised by $\sum_{i=1}^{n} v_i = 887.4$. Numerical maximisation of the unrestricted likelihood gives $\hat{\alpha} = 1.710$ and $\hat{\beta} = 0.03801$. Carry out a Generalized Likelihood Ratio test of the hypothesis that the value of β is $\beta_0 = 0.02730$, the appropriate decay constant for a particular isotope. ***(9 marks)***

End of Question Paper