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

MAS371

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2011–2012**

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.*

1 (i) Let $X_1, \ldots, X_n$ be independent Binomial(m, θ) random variables.

(a) Show that the log-likelihood for θ given an observation x_i is

$$x_i \log(\theta) + (m - x_i)\log(1-\theta) + constant.$$

Hence obtain the log-likelihood for θ given observed values $x_1, \ldots, x_n$ and show that the maximum likelihood estimate (MLE) of θ is $\sum x_i/nm$. ***(6 marks)***

(b) Derive the form of the expected information $I(\theta)$. ***(3 marks)***

(ii) The vector $\mathbf{X} = (X_1, \ldots, X_k)$ has a Multinomial$(m; \theta_1, \ldots, \theta_k)$ distribution.

(a) Show that the log-likelihood for $\theta_1, \ldots, \theta_k$ given observed values $x_1, \ldots, x_k$ is

$$\sum_{i=1}^{k} x_i \log(\theta_i) + constant.$$

By defining an appropriate Lagrangian function, derive the MLE for θ_j. ***(4 marks)***

(b) If $k = 6, m = 100, x_1 = 21, x_2 = 16, x_3 = 13, x_4 = 14, x_5 = 19, x_6 = 17$, use Wilks' Theorem to carry out a likelihood ratio test of the hypothesis that $\theta_1 = \cdots = \theta_k$. ***(7 marks)***

2 Wild animals being studied by zoologists exhibit two distinct patterns of daily behaviour: 'exploratory' behaviour, in which an animal travels long distances, and 'encamped' behaviour in which its movements are confined to a smaller area. A Markov chain model has been proposed for the sequence of behaviour by each individual animal, recorded on successive days.

(i) Animal 'A' follows the sequence of behaviours

X, N, N, N, X, N, N, N, N, X, X, N, N, N, N, X, X

where X represents a day of 'exploratory' activity and N represents a day on which the animal is 'encamped'. Fit the Markov chain with transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

giving maximum likelihood estimates and estimated standard errors for p and q. ***(3 marks)***

(ii) A further individual 'B' has a sequence of behaviour that can be summarised in the following table.

		Next behaviour X	Next behaviour N
Previous behaviour	X	6	9
	N	9	8

If 'B' has a transition matrix of the same form as 'A', investigate whether there is any evidence that transition probabilities differ between the two animals. ***(10 marks)***

(iii) (a) Show that $(q/(p+q), p/(p+q))$ is a stationary distribution for the chain in part (i). ***(3 marks)***

(b) Hence explain how maximum likelihood estimates for the parameters p and q could be found taking into account the first observation in a sequence (instead of the usual simplification of conditioning on that value). Explain what qualitative effect taking the first observation into account would have on the values calculated in part (i); you are not required to actually carry out the calculation. ***(4 marks)***

3 (i) The continuous-time Markov chain $X(t)$ is a generalized birth-death process with infinitesimal generator $G = (g_{ij})$ where

$$\begin{aligned} g_{ii} &= -(\kappa + i(\lambda + \mu)), \quad i \geq 0 \\ g_{i,i+1} &= i\lambda + \kappa, \quad i \geq 0 \\ g_{i,i-1} &= i\mu, \quad i \geq 1 \end{aligned}$$

and $g_{ij} = 0$ otherwise.

(a) Given that $X(0) = i > 0$,

$$\Pr\{X(\delta t) = i-1\} = i\mu\delta t + o(\delta t),$$

where δt is small but finite. Write down the probabilities of all other possible values of $X(\delta t)$. ***(3 marks)***

(b) Explain for which parameter values the process has an equilibrium distribution, and derive its form (up to proportionality). What shape does the distribution have for large values of $X(\cdot)$? ***(7 marks)***

(c) In the special case where $\lambda = 0$, write down the log-likelihood for μ and κ based on a complete record of the state of the process over a time interval $(0, t)$, and derive expressions for maximum likelihood estimates of these parameters. ***(4 marks)***

(ii) A 'birth process with catastrophes' is a continuous-time Markov chain with infinitesimal generator $G = (g_{ij})$ where

$$\begin{aligned} g_{00} &= -\lambda_0, \\ g_{ii} &= -(\lambda_i + \nu_i)), \quad i \geq 1 \\ g_{i,i+1} &= \lambda_i, \quad i \geq 0 \\ g_{i,0} &= \nu_i, \quad i \geq 1 \end{aligned}$$

and $g_{ij} = 0$ otherwise.

State carefully for which parameter values this process is a generalized birth-death process. In this case, explain what kind of long-run behaviour the process will exhibit. Give examples of parameter values when the process is *not* a generalized birth-death process and (a) approaches a non-degenerate equilibrium distribution; (b) has a positive probability of growing indefinitely. ***(6 marks)***

4 As an ecologist walks in a straight line through a region known to be occupied by tigers, she finds signs that indicate that a tiger has recently crossed the line she is following. She believes that the locations of these signs can be modelled by an inhomogeneous Poisson process, with rate

$$\lambda(t) = \begin{cases} \lambda_1 & 0 \le t \le t_1, \\ \lambda_2 & t_1 \le t \le t_1 + t_2, \\ \cdots & \\ \lambda_k & \sum_{j=1}^{k-1} t_j \le t \le \sum_{j=1}^{k} t_j \end{cases}$$

for some constants $t_1, \ldots, t_k > 0$, and parameters $\lambda_1, \ldots, \lambda_k > 0$, where t represents the distance she has travelled along her route.

(i) Write down the *general* form for the likelihood for an inhomogeneous Poisson process observed over an interval $[0, T]$. (You do not need to derive the result, but you should define your notation.) ***(3 marks)***

(ii) (a) In the specific model defined above, if the ecologist observes n signs altogether, at locations $v_1, \ldots, v_n$, and assuming that $t_1, \ldots, t_k$ are known, show that the likelihood for $\lambda_1, \ldots, \lambda_k$ can be written in terms of $t_1, \ldots, t_k$ and suitable counts $n_1, \ldots, n_k$ (which you should define). ***(3 marks)***

(b) Hence give expressions for the maximum likelihood estimators of $\lambda_1, \ldots, \lambda_k$ and for their estimated covariance matrix. ***(6 marks)***

(iii) The ecologist observes just 5 signs in the first 7 kilometres of her walk, in open ground, and 20 signs in the remaining 2 kilometres, in forest. Taking $k = 2, t_1 = 7, t_2 = 2$, calculate approximate 95% confidence intervals for (a) λ_1, (b) λ_2, (c) $\lambda_1 - \lambda_2$ and (d) $\lambda_1 t_1 + \lambda_2 t_2$. ***(8 marks)***

End of Question Paper