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

The University Of Sheffield.

MAS371

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2012–2013**

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 The random variables $X_1, \dots, X_n$ are independent and each follow the Geometric(θ) distribution for some $0 < \theta \leq 1$, with

$$Pr(X_i = x|\theta) = \theta(1-\theta)^{x-1}, \quad x = 1, 2, \dots$$

(i) Given observed values $x_1, \dots, x_n$, show that the log-likelihood $l(\theta|x_1, \dots, x_n)$ can be written in terms of n and the sum

$$t = \sum_{i=1}^{n} x_i.$$

(3 marks)

(ii) Find the maximum likelihood estimate of θ: (a) when $t > n$; (b) when $t = n$. ***(4 marks)***

(iii) If $n = 15$ and $t = 40$, show that $(0.2362, 0.5294)$ is a likelihood interval for θ. If this is interpreted as a confidence interval for θ, calculate its approximate confidence level. ***(5 marks)***

(iv) If $n = t = 20$, calculate a likelihood interval for θ with a confidence level of approximately 95%. ***(4 marks)***

(v) Calculate the observed information $J(\theta)$ for $0 < \theta < 1$. Given the data in (iii), evaluate $J(\hat{\theta})$ and hence calculate another approximate confidence interval for θ, with the same confidence level. ***(4 marks)***

2 As part of an investigation into social mobility, the occupations and incomes of women and their mothers, grandmothers, etc were recorded. Each woman's status was classified as Upper class, Middle class or Lower class. A proposed model is that the class defined in this way follows a Markov chain on these three states, with transition matrix P, so that each woman's class depends only on her mother's class. (The population can be assumed to be large enough that different families are independent, given the transition matrix.)

(i) The sequence of states for a sample of women in a particular town and their female ancestors can be summarised as follows.

n_{ij}		Daughter's status		
		U	M	L
Mother's	U	5	1	1
status	M	2	29	4
	L	1	7	50

The following table gives maximum likelihood estimates $\hat{p}_{ij}$ for each i, j.

	$\hat{p}_{ij}$	j = 1	2	3
	1	0.7143	0.1429	0.1429
i	2	0.0571	0.8286	0.1143
	3	$\star$	$\star$	0.8621

Calculate the missing values (indicated by $\star$) and the two corresponding estimated standard errors. ***(6 marks)***

(ii) Based on the data in (i), give approximate 90% confidence intervals for $p_{31} + p_{32}$ and for $p_{21} - p_{23}$. ***(6 marks)***

(iii) A much larger sample at a national level gives the following estimated transition probabilites:

$$Q = \begin{pmatrix} 0.713 & 0.208 & 0.079 \\ 0.097 & 0.762 & 0.141 \\ 0.079 & 0.093 & 0.828 \end{pmatrix}.$$

The sample is large enough that uncertainty on these probabilities can be ignored. Explain how to test the hypothesis that the data in (i) come from a Markov chain with the given transition matrix Q, and carry out the test. To simplify calculation, you may wish to use the fact that

$$\sum_{ij} n_{ij} \log(\hat{p}_{ij}) = -51.711, \quad \sum_{ij} n_{ij} \log(q_{ij}) = -54.786.$$

(6 marks)

Comment briefly on the interpretation of the model defined by Q, and of the differences suggested by the data from (i), if any. ***(2 marks)***

3 (i) Observations on the number of females in a captive population of a rare mammal can be summarised as follows.

Population size	Total time (in months) at this size	No. of transitions to size: 4	5	6	7
4	0.4	–	2	0	0
5	1.1	1	–	2	0
6	2.0	0	1	–	3
7	2.5	0	0	2	–

So for example, the total duration of all the times when there were 6 females was 2 months, and there were 3 births that occured while there were 6 females.

Assuming that the population can be modelled by a time-homogeneous linear birth-death process, estimate the parameters of the model based on the above data. Give a 95% confidence interval for the difference between the birth rate per individual and the death rate per individual. Comment briefly on what this model suggests about the long-term size of this population. ***(8 marks)***

(ii) A meta-population model is a simplified model of large-scale population dynamics which represents the number of 'islands' (or similar) that are occupied by a species. In a particular model of this sort, there are n islands, of which $X(t)$ are occupied at time t. The three types of events that can occur, with their rates when $X(t) = x$ are as follows.

Local extinction, reducing x by 1: rate $x\mu$.

Colonization from outside, increasing x by 1: rate $(n - x)\kappa$.

Internal colonization, also increasing x by 1: rate $x(n - x)\lambda$.

(a) Write down the probability $Pr(X(t+\delta t) = x+1|X(t) = x)$ for some small finite time-step δt. ***(2 marks)***

(b) For the case $n = 3$, write down the infinitesimal generator for the above continuous-time Markov chain, and for general n, explain why it is a generalized birth-death process. ***(5 marks)***

(c) If $n = 3, \lambda = \kappa$ and $\mu = \lambda/2$, what is the long-run proportion of the time for which all islands would be occupied? ***(5 marks)***

4 In an archaeological excavation, the depths $d_1, \ldots, d_n$ at which artefacts are found can be modelled as a point process over the interval $(0, D)$ where D is the depth in metres to which excavation has been carried out.

(i) If the point process is taken to be a homogeneous Poisson process, given an expression for the maximum likelihood estimate of the rate λ of the process in terms of n and D and for the estimated standard error on λ. ***(2 marks)***

(ii) Some kinds of artefacts are more commonly found at greater depths; their depths form an inhomogeneous Poisson process, with rate at depth d given by

$$\lambda(d) = \alpha d.$$

(a) Write down the likelihood for the parameter α given the data $d_1, \ldots, d_n$ as above. ***(4 marks)***

(b) Thus obtain an expression for the maximum likelihood estimate of α. ***(2 marks)***

(iii) A particular excavation reaches 3 metres down, and the depths (in metres) at which artefacts are found are as follows.

$$0.23, 0.31, 0.90, 1.05, 1.88, 1.95, 2.45, 2.46, 2.59, 2.70.$$

Calculate approximate 90% confidence intervals for

(a) λ in part (i), and

(b) α in part (ii). ***(5 marks)***

(iv) A more flexible model for the depths is given by an inhomogeneous Poisson process with

$$\lambda(d) = \begin{cases} \alpha d & d < \beta \\ \alpha\beta & d \geq \beta, \end{cases}$$

so that the rate increases down to some depth β $(0 < \beta \leq D)$ and is constant below that. If $m = m(d_1, \ldots, d_n)$ is defined to be the number of values d_i such that $d_i \geq \beta$, show that the likelihood for (α, β) given $d_1, \ldots, d_n$ can be written in terms of D, n, m and one other summary statistic, which you should derive. ***(5 marks)***

(v) Given a maximum depth of excavation D, explain briefly how to carry out a test of the hypothesis that the rate is proportional to depth, as in (ii), against the alternative that it takes the general form in (iv). You do not need to carry out the test. ***(2 marks)***

End of Question Paper