RESTRICTED OPEN BOOK EXAMINATION
Data provided: Table of χ^2 quantiles

MAS371

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2013–2014**

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

(i) 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)

(ii) If $k = 4, m = 25, x_1 = 5, x_2 = 9, x_3 = 7, x_4 = 4$, use Wilks' Theorem to carry out a likelihood ratio test of the hypothesis that $\theta_1 = \cdots = \theta_k$.

(8 marks)

(iii) In the case $k = 2$ (the binomial distribution), there is effectively a single parameter, θ_1. Write down the likelihood in this case, and calculate the observed information and the expected information about θ_1. Comment briefly on the dependence of the expected information on the value of m.

(8 marks)

2 Mice can be classified genetically into one of three types, labelled 0, 1 and 2. Breeding the mice in the laboratory, and observing the genetic type of one particular mouse, and then its first offspring, and so on, gives a sequence of linked observations $X_0, X_1, \ldots, X_n$ say, each taking values in $S = \{0, 1, 2\}$.

(i) A possible model for how the observations are linked is a Markov chain with transition matrix

$$\begin{pmatrix} p_{00} & p_{01} & 0 \\ p_{10} & p_{11} & p_{12} \\ 0 & p_{21} & p_{22} \end{pmatrix}$$

where the 0s denote values that are forced to be zero by the basic laws of genetics, and the other entries are all positive.

In one particular sequence of the form described above, the observations are as follows.

0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 2, 1, 0, 1, 2, 2, 1, 1.

Calculate the maximum likelihood estimate of the transition matrix for the model described above and estimated standard errors for each of the parameters. ***(6 marks)***

(ii) A particular genetic mechanism suggests that the transition matrix in (i) should have the form

$$\begin{pmatrix} p & 1-p & 0 \\ p/2 & 1/2 & (1-p)/2 \\ 0 & p & 1-p \end{pmatrix}$$

for some unknown parameter p.

Write down the likelihood for this form of the model, based on an observation of a sequence of genetic types summarised by $n_{ij}, i, j \in \{0, 1, 2\}$, where n_{ij} denotes the number of individuals of type i in the sequence whose first offspring were of type j. (You may assume that $n_{02} = n_{20} = 0$, so that the observed sequence is *possible* under this form of the model.) ***(3 marks)***

Hence show that the form of the maximum likelihood estimate of p is

$$\frac{n_{00} + n_{10} + n_{21}}{n - n_{11}},$$

where n is the total number of observations. ***(3 marks)***

(iii) Test the hypothesis that the data in (i) come from a model of the form in (ii), against the alternative of the more general form in (i). ***(8 marks)***

3 The flow of users through a library can be modelled as follows. At any instant t, the state $X(t)$ represents the number of users who are currently in the library. In a small time interval $(t, t + \delta t)$, when $X(t) = x$, the probability of a new user arriving (increasing x by one) is $\lambda\delta t + o(\delta t)$ and the probability of one of the current users departing (decreasing x by one) is $\mu x\delta t + o(\delta t)$; all other changes in x have probabilities that are $o(\delta t)$.

(i) Write down an expression for $P(X(t + \delta t) = x|X(t) = x)$. ***(2 marks)***

(ii) Explain why this model is an example of a generalized birth-death process, and specify the elements of the generator. ***(3 marks)***

(iii) Using the general results for generalized birth-death processes, derive the form of the stationary distribution for the number of users in the library. ***(4 marks)***

(iv) Given complete observation of the number of users in the library over an interval $[0, t]$, show that the likelihood for λ and μ can be written as

$$L(\lambda, \mu) = \exp(\lambda t + \mu A)\lambda^B C\mu^D$$

and define the constants A, B, C, D. Hence give the form of the maximum likelihood estimates of λ and μ. ***(7 marks)***

(v) Obtain an expression for the observed information about the parameter vector (λ, μ) and hence give approximate standard errors for $\hat{\lambda}$ and $\hat{\mu}$. ***(4 marks)***

4 The arrival times $T_1, T_2, \ldots$ of particles at a detector can be modelled as a point process over the interval $(0, T)$. For convenience in the rest of the question, times have already been scaled such that π units represents one day.

(i) If the point process is taken to be a homogeneous Poisson process

$$\lambda(t) = \kappa,$$

give an expression for the maximum likelihood estimate of the rate κ of the process in terms of T and the total number of particles recorded n, and an expression for the estimated standard error of κ. ***(2 marks)***

(ii) One theory suggests that the arrival rate depends on the time of day, with

$$\lambda(t) = \alpha(1 + \cos(t + \omega))$$

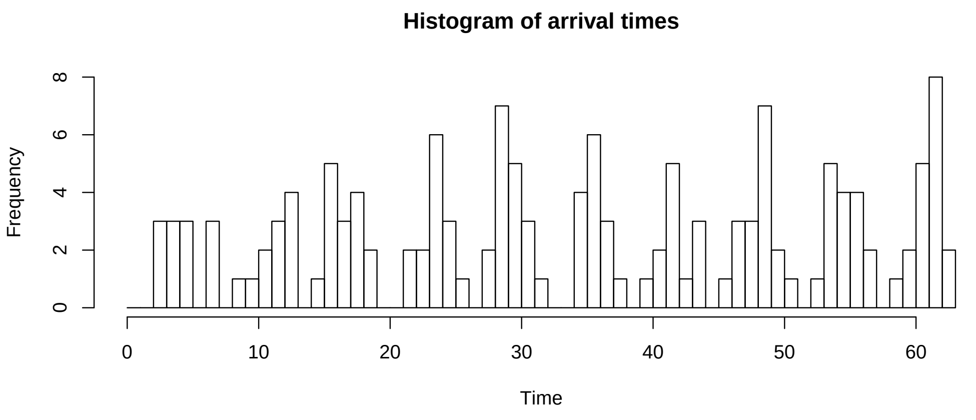
where α and ω are unknown parameters. If the total time T can be assumed to be $T = 2\pi k$ with k an integer, then show that the log-likelihood $l(\alpha, \omega)$ for α and ω based on the observations $t_1, \ldots, t_n$ is given by

$$l(\alpha, \omega) = -2\pi\alpha k + n\log(\alpha) + \sum_{i=1}^{n} \log(1 + \cos(t_i + \omega)).$$

(6 marks)

Hence, derive the likelihood equations for α and ω.
Hint: recall that $\tan(\theta/2) = \sin(\theta)/(1 + \cos(\theta))$. ***(5 marks)***

(iii) In one experiment, with $T = 20\pi$, there are $n = 147$ particles detected, as summarised in the figure.



Solving the likelihood equation for ω numerically gives $\hat{\omega} = 2.014$. Obtain the maximum likelihood estimate of α in this case, and comment on whether the data appear consistent with these parameter values. ***(7 marks)***

End of Question Paper

Table of the pth quantile of the χ^2 distribution with ν degrees of freedom, $\chi^2_{p,\nu}$

		ν								
		1	2	3	4	5	6	7	8	9
	0.10	0.016	0.211	0.584	1.064	1.610	2.204	2.833	3.490	4.168
	0.50	0.455	1.386	2.366	3.357	4.351	5.348	6.346	7.344	8.343
p	0.90	2.706	4.605	6.251	7.779	9.236	10.645	12.017	13.362	14.684
	0.95	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919
	0.99	6.635	9.210	11.345	13.277	15.086	16.812	18.475	20.090	21.666