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

The University Of Sheffield.

MAS371

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2016–2017**

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 solid line in Figure 1 represents the relative log-likelihood for a single parameter θ based on i.i.d. observations $x_1, \ldots, x_n$. The dashed lines intersect the curve at

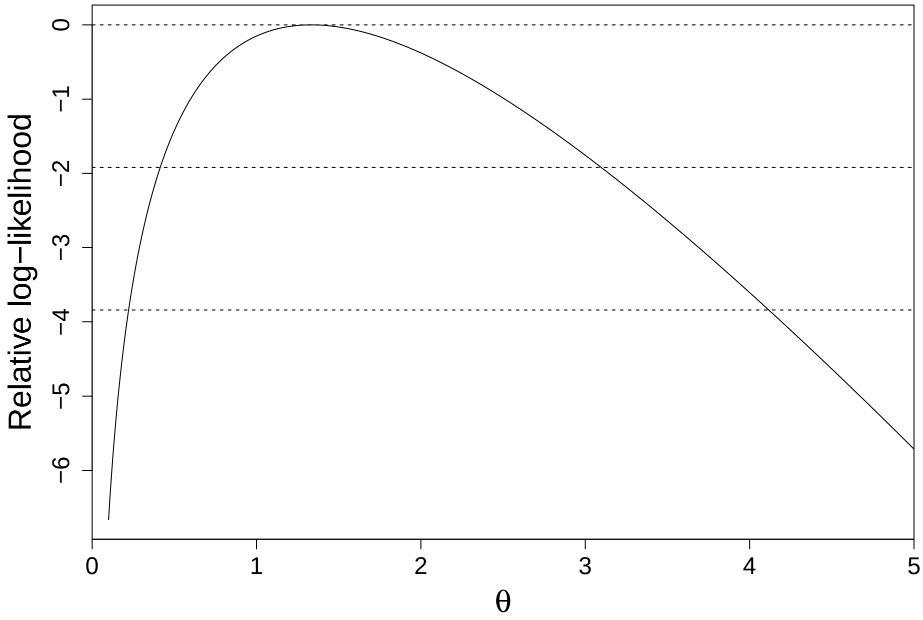

Figure 1: A relative log-likelihood function

the points $(0.222, -3.84), (0.414, -1.92), (1.333, 0.00), (3.097, -1.92), (4.117, -3.84)$.

(i) Give the maximum likelihood estimate $\hat{\theta}$ of θ. ***(1 mark)***

(ii) Give a numerical example of a likelihood region for θ based on this graph, and explain why your stated region *is* a likelihood region. ***(4 marks)***

(iii) Assuming that Wilks' Theorem holds, explain approximately what the coverage probability would be for a region constructed in the way that you used in (ii). ***(3 marks)***

(iv) The graph in Fig. 1 is actually based on n Poisson observations, so that

$$\Pr(x_i; \theta) = \exp(-\theta)\theta^{x_i}/x_i!, \; x_i = 0, 1, 2, \ldots$$

Show that the observed information on θ is given by

$$J(\theta; x_1, \ldots, x_n) = \theta^{-2} \sum_{i=1}^{n} x_i.$$

(5 marks)

(v) In Figure 1, $n = 3$ and $\sum_{i=1}^{3} x_i = 4$. Based on the observed information, calculate an approximate 95% confidence interval for θ, and explain how it compares with the corresponding likelihood region. ***(7 marks)***

2 (i) A simplified record of the daily weather at a specific location uses three categories: sunny, cloudy or wet. The table classifies the pairs of consecutive days during one summer.

		Weather on day $i+1$		
		Cloudy	Sunny	Wet
Weather	Cloudy	19	11	5
on day i	Sunny	9	30	8
	Wet	6	7	5

(a) Assuming that the type of weather in this location follows a 3-state Markov chain, calculate estimates for the probability of the weather remaining the same from one day to the next, in each of the three states. ***(3 marks)***

(b) Calculate the estimated standard error for the probability that a cloudy day is followed by a sunny one. ***(2 marks)***

(c) Calculate an approximate 95% confidence interval for the difference between the probability that a sunny day is followed by a cloudy day and the probability that sunny day is followed by a wet day. ***(4 marks)***

(ii) A computer program is intended to simulate a sequence of independent rolls of a fair 6-sided die, to enable online playing of board games. A user is interested in testing the program to check whether the simulated die rolls are fair and independent.

(a) Explain why it may be useful to represent the sequence of rolls as a Markov chain with transition matrix P, and then to test the hypothesis that

$$P^\star = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

(3 marks)

(b) Give one potential disadvantage of this approach, bearing in mind the user's aim. ***(1 mark)***

(c) The user records a long sequence of rolls, consisting of n_{ij} observations of a roll of i being followed by a roll of j, for $i, j \in \{1, \dots, 6\}$. If $n_{1\bullet} = 129, n_{2\bullet} = 125, n_{3\bullet} = 114, n_{4\bullet} = 110, n_{5\bullet} = 118, n_{6\bullet} = 120$ (where $n_{i\bullet} = \sum_{j=1}^{6} n_{ij}$) and $\sum_{i,j=1}^{6} n_{ij} \log(n_{ij}) = 2156.556$, calculate $l(P^\star)$, the log-likelihood under the hypothesis in (a), and $l(\hat{P})$, the log-likelihood at the unrestricted m.l.e., and hence carry out the test in (a). ***(7 marks)***

3 A Linear Immigration Death process is a continuous-time Markov chain $X(t)$ in which new individuals arrive at a rate α and *each* individual experiences a death rate $\beta > 0$, independently of each other and of arrivals.

(i) Write down an expression for the probability $\Pr(X(t+\delta t) = j | X(t) = j)$ for small but finite δt. ***(2 marks)***

(ii) Explain why this process is an example of a generalized birth-death process, and specify the elements of its generator. ***(4 marks)***

(iii) Explain why this process can also be thought of as an 'infinite-server queue'. ***(2 marks)***

(iv) State the equations that must be satisfied by a stationary distribution for this process. (You may use standard results for generalized birth-death processes, but they must be clearly stated.) Hence show that such a distribution always exists, and give its form. ***(5 marks)***

(v) Write down the log-likelihood for α and β based on complete observation of the process over an interval $[0, T]$. (You may use standard results for the likelihood of a continuous-time Markov chain in terms of counts n_{ij} of transitions between states and times a_i spent in each state.) ***(3 marks)***

(vi) Derive the maximum likelihood estimators of α and β, and their observed information matrix J. ***(4 marks)***

4 In an archaeological excavation at a particular site, the depths of artefacts found, $x_1, \ldots, x_n$, are recorded as values between 0 (the surface) and D metres (the depth to which excavation has been carried out).

(i) If the depths are assumed to form a homogeneous Poisson process, give an expression for the maximum likelihood estimate of the the rate of the process and for its estimated standard error. ***(2 marks)***

(ii) A more realistic model for the occurrence and survival of artefacts suggests that the Poisson process of depths may be inhomogeneous, with rate of the form

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

where β could be positive or negative. Derive an expression for the likelihood for α and β in terms of D, n and $x_1, \ldots, x_n$, and the corresponding log-likelihood. ***(5 marks)***

(iii) A large excavation was carried out to a depth of 4 metres, and 87 artefacts found; the sum of their depths was 147.9 metres. Numerical maximisation of the log-likelihood in (ii) gave $\hat{\alpha} = 32.8$ and $\hat{\beta} = -0.228$. Carry out a generalized likelihood ratio test of the hypothesis that $\beta = 0$. ***(9 marks)***

(iv) Explain carefully how you would obtain confidence intervals for α and β. (You are not required to actually calculate these intervals.) ***(4 marks)***

End of Question Paper

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

		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
q	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
		ν								
		10	20	30	40	50	60	70	80	90
	0.10	4.87	12.44	20.60	29.05	37.69	46.46	55.33	64.28	73.29
	0.50	9.34	19.34	29.34	39.34	49.33	59.33	69.33	79.33	89.33
q	0.90	15.99	28.41	40.26	51.81	63.17	74.40	85.53	96.58	107.57
	0.95	18.31	31.41	43.77	55.76	67.50	79.08	90.53	101.88	113.15
	0.99	23.21	37.57	50.89	63.69	76.15	88.38	100.43	112.33	124.12