Tutorial letter 012/0/2024

Generalized Linear Models

STA4813

Year module

Department of Statistics

ASSIGNMENT 02 QUESTIONS

ASSIGNMENT 02 **Unique Nr.: 858959** **Fixed closing date: 24 May 2024**

Instructions

(1) Use R Markdown to compile your solutions.

(2) Your solutions must have the full R codes and outputs of all questions.

(3) Discuss the relevant R outputs that are related to a specific question.

QUESTION 1 [20]

Let $X_1, \ldots, X_n$ denote independent and identically distributed random variables from a Negative Binomial distribution with parameters r and θ. The probability distribution function for each X_i is therefore

$$f(x_i, \theta) = \binom{x_i + r - 1}{r - 1} \theta^r (1 - \theta)^x_i \;\; x_i, \;\; \text{if} \;\; x_i = 0, 1, \ldots \;\; \text{and} \;\; 0 \;\; \text{otherwise}.$$

Suppose that r is known.

(a) Show that $f(x_i, \theta)$ belongs to the exponential family of probabilty mass functions. (4)

(b) Using results from Chapter 3 of the prescribed textbook or Additional Resources material of Learning Unit 2 that is available on the module web site, find $E(X)$ and $var(X)$. (4)

(c) Calculate the score statistics U and verify that $E(U) = 0$. (2)

(d) Drive the information $\mathfrak{I} = var(U)$. (2)

(e) Find the maximum likelihood estimator $\hat{\theta}$ of θ. (3)

(f) Drive the Wald statistic for making inferences about θ. (3)

(g) Use the Wald statistic to obtain an expression for an approximate 95% confidence interval for θ. (2)

QUESTION 2 [10]

Let $Y_1, \ldots, Y_n$ be a random sample from the geometric distribution which has a probability mass function

$$f(y, p) = p(1 - p)^{y-1}, \quad x = 1, 2, \ldots \quad \text{and} \quad 0 \quad \text{otherwise}$$

Here p is an unknown parameter.

(a) Show that this distribution belongs to the exponential family of probability mass functions. (3)

(b) Drive the information $\mathfrak{I} = var(U)$. (4)

(c) Find the maximum likelihood estimator $\hat{p}$ of p. (3)

QUESTION 3 **[15]**

The file `gamma_arrivals.txt` data in the Additional Resources section of the module web site contains set of gamma-rays data consisting of the times between arrivals (inter-arrival times) of 3935 photons (units are seconds). Assume the following Gamma distribution is a good model for the data:

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad \text{and} \quad 0 \quad \text{otherwise}$$

where both α and β are unknown.

(a) Draw a histogram of the inter-arrival times. Does it appear that a gamma distribution would be a plausible model? Explain. (4)

(b) Estimate the model parameters using the method of maximum likelihood estimation. Write your own R code for MLE, i.e., do not use any R built in package or function except for optimization. (7)

(c) Plot the two fitted gamma densities on top of the histogram. Do the fits look reasonable? Explain. (4)

Hint: In order to obtain the MLE, you need to maximize the likelihood function or log likelihood function. The R package provides a function `nlm()`, which can minimize an object function, therefore, you need to define the negative log likelihood function in your R code.

[45]