Bachelor's Degree in Bioinformatics
Statistical Models & Stochastic processes
Academic year 2023-2024 1st Quarter

Practical 1. Maximum Likelihood (ML) estimation

Hand-in date: 21/10/2023

Resolve the following exercise in groups of two students. Write your solution in a Word, Latex or Markdown document and generate a pdf file with your solution. Upload the pdf file with your solution to the corresponding task at the Moodle environment of the course, no later than the hand-in date.

1. (16p) ML estimation of a one-parameter distribution. Let X be a random variable with probability density $f(x \mid \beta) = \beta e^{-\beta x+1}$ with $x > 1/\beta$

   We consider a random sample of n observations of this distribution.

   a) (2p) Write down the likelihood function for a sample of n observations of this distribution
   b) (1p) Obtain the log-likelihood function.
   c) (2p) Find the stationary point(s) of the log-likelihood function analytically
   d) (1p) Determine whether the stationary point(s) are maxima or minima
   e) (1p) Download the file Sample.dat, which contains sample of observations from this probability distribution. Determine the sample size and calculate the value of the ML estimator for this sample. (See Sample.dat file)
   f) (2p) Plot the log-likelihood function, and assess graphically if your ML estimate coincides with the maximum of this function.
   g) (1p) Determine an expression for the Fisher information by calculating

   $$-E\left(\frac{\partial^2 l}{\partial \beta^2}\right)$$

   h) (1p) Use the Fisher information for obtaining an expression for the variance of the maximum likelihood estimator $\hat{\beta}_{ML}$.
   i) (1p) Using the asymptotic normality of the ML estimator, give an expression of a confidence interval for β.
   j) (1p) Calculate a 95% confidence interval for parameter β, using the dataset that you have downloaded.
   k) (1p) Do you think it is tenable that β = 1? Justify your answer.
   l) 2p) Make a histogram of the data, using function hist, using the argument freq=FALSE. Overplot the histogram with the estimated probability density f(x|β), using the maximum likelihood estimate. What do you observe?

2. (3p, 1p per question) Hardy-Weinberg law.

The Hardy-Weinberg law in genetics says that the proportions of genotypes AA, Aa, and aa are θ^2, $2\theta(1-\theta)$, and $(1-\theta)^2$, respectively, where $\theta \in [0,1]$. Suppose that in a sample of n individuals from the population (small relative to the size of the population), we observe x_1 individuals of type AA, x_2 individuals of type Aa, and x_3 individuals of type aa.

(a) What distribution do the counts (X_1, X_2, X_3) follow?

(b) Record the likelihood function, the log-likelihood function, and the score function for θ.

(c) Record the form of the MLE for θ.

3. (1p) Find an example of a published research where they study/apply/use MLE, include the link, with the retrieved data and do a summary of 5-6 text lines.