Statistical Data Analysis
Problem sheet 7
Due Monday 30 November, 2020

Exercise 1: Suppose the random variable x follows the pdf,

$$f(x;\theta) = \frac{x}{\theta^2} e^{-x/\theta} ,$$

with $x \geq 0$ and $\theta > 0$. The expectation value and variance of x are $E[x] = 2\theta$, $V[x] = 2\theta^2$. Consider a sample of n independent values $x_1, \ldots, x_n$ from this pdf, with which we want to estimate θ. For parts (a)–(c), suppose that n is a fixed constant.

1(a) [3 marks] Write down the likelihood function $L(\theta)$ and show that the maximum-likelihood estimator for θ is

$$\hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} x_i .$$

1(b) [4 marks] Show that $\hat{\theta}$ is unbiased, find its variance, and show that the variance is equal to the minimum variance bound.

1(c) [2 marks] Make a sketch of the log-likelihood function indicating the estimator $\hat{\theta}$ and indicate on the sketch how to find the standard deviation of $\hat{\theta}$.

For the rest of this question suppose that the sample size n is not fixed but rather follows a Poisson distribution with mean $\alpha\theta^2$, where α is a given constant. (Recall that the Poisson distribution for n with mean ν is $P(n;\nu) = \nu^n e^{-\nu}/n!$.)

1(d) [4 marks] Write down the full (i.e., extended) likelihood function for θ based on the Poisson distributed n and the n values $x_1, \ldots, x_n$. Show that the maximum-likelihood estimator for θ is

$$\hat{\theta} = \left(\frac{1}{2\alpha} \sum_{i=1}^{n} x_i \right)^{1/3} .$$

1(e) [3 marks] Show that the expectation value of a function a of n and $\mathbf{x} = (x_1, \ldots, x_n)$ can be written

$$E[a(n, \mathbf{x})] = E_n \left[E_{\mathbf{x}}[a(n, \mathbf{x}) | n] \right] ,$$

where E_n and $E_{\mathbf{x}}$ indicate the expectation values with respect to n and $\mathbf{x}$, respectively.

1(f) [4 marks] Using the result from (e) and the second derivative of the log-likelihood function, show that the variance of $\hat{\theta}$ can be approximated as

$$V[\hat{\theta}] = \frac{1}{6\alpha} ,$$

stating any assumptions needed. Using the fact that the expectation value of n is $\alpha\theta^2$, compare the variance found here with that found in (b) for fixed n, and comment on why they are different.

2 ([0 marks] – nothing to turn in): This is a warm-up for maximum-likelihood fitting with the minimization program MINUIT, using either its python implementation iminuit or the root/C++ version TMinuit. Please download the code and see if you can get it to run. We will return to this later on.

The programs below generate a data sample of 200 values from a pdf that is a mixture of an exponential and a Gaussian:

$$f(x;\theta,\xi) = \theta \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} + (1-\theta)\frac{1}{\xi} e^{-x/\xi} \,, \tag{1}$$

The pdf is modified so as to be truncated on the interval $0 \leq x \leq x_{\rm max}$. The program Minuit is used to find the MLEs for the parameters θ and ξ, with the other parameters treated here as fixed. You can think of θ as representing the fraction of signal events in the sample (the Gaussian component), and the parameter ξ characgterizes the shape of the background (exponential) component.

To use python, you will need to install the package iminuit (should just work with "pip install iminuit"). See:

`https://pypi.org/project/iminuit/`

Then download the program `simple_iminuit.py` from

`http://www.pp.rhul.ac.uk/~cowan/stat/python/iminuit/`

To use C++/ROOT, download the files from

`http://www.pp.rhul.ac.uk/~cowan/stat/root/tminuit/`

to your work directory and build the executable program by typing `make`. This uses the class `TMinuit`, which is described here:

`https://root.cern.ch/doc/master/classTMinuit.html`