Statistical Data Analysis
Problem sheet 6
Due Monday 23 November, 2020

Exercise 1: The number n of events observed in high-energy particle collisions having particular kinematic properties can be treated as a Poisson distributed variable. Suppose that for a certain integrated luminosity (i.e. time of data taking at a given beam intensity), $b = 3.9$ events are expected from known processes and $n_{\rm obs} = 16$ are observed.

1(a) [5 marks] Compute the p-value for the hypothesis that no new signal process is contributing to the number of events. To sum Poisson probabilities, you can use the relation

$$\sum_{n=0}^{m} P(n;\nu) \;=\; 1 - F_{\chi^2}(2\nu; n_{\rm dof})\,, \qquad (1)$$

where $P(n;\nu)$ is the Poisson probability for n given a mean value ν, and F_{χ^2} is the cumulative χ^2 distribution for $n_{\rm dof} = 2(m+1)$ degrees of freedom. This can be computed using the ROOT routine `TMath::Prob` or in Python with `scipy.stats.chi2.sf` (both of which give one minus $F_{\chi 2}$), or looked up in standard tables. If you have difficulty getting a program to return F_{χ^2}, you can simply carry out the sum of Poisson probabilities explicitly.

1(b) [2 marks] Find the corresponding equivalent Gaussian significance Z and evaluate numerically (see lecture notes).

Exercise 2: Consider N independent observations $n_1, \ldots, n_N$ of a Poisson random variable with the same unknown mean value ν.

2(a) **[3 marks]:** Write down the likelihood function for the parameter ν. (Since the Poisson distribution is not a pdf but rather a probability, here the likelihood is found directly from the joint probability for the data.) Find the maximum-likelihood estimator for ν.

2(b) **[3 marks]:** Show that the estimator is unbiased and find its variance in closed form (use the known mean and variance of a Poisson variable).

2(c) **[3 marks]:** Show that the variance of $\hat{\nu}$ is equal to the minimum variance bound (the right-hand side of the information inequality).

Exercise 3 Consider the pdf of the continuous random variable x

$$f(x) = \begin{cases} (1+\theta)x^{\theta} & 0 \le x \le 1\,, \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

Suppose the value of the parameter θ is unknown and that one is given a sample of independent observed values $x_1, \ldots, x_n$.

3(a) [3 marks] Write down the log-likelihood function for θ and show that the Maximum Likelihood (ML) estimator is

$$\hat{\theta} = -1 - \frac{n}{\sum_{i=1}^{n} \ln x_i} \; .$$

3(b) [2 marks] Make a rough sketch of what the log-likelihood function would like like assuming a large data sample. Indicate on the sketch the ML estimator $\hat{\theta}$ and show how, using the graph, to estimate its standard deviation.

3(c) [3 marks] suppose the pdf is written in terms of the parameter $\lambda = \exp[-1/(1+\theta)]$. Show that the ML estimator for λ is

$$\hat{\lambda} = \prod_{i=1}^{n} x_i^{1/n} \; .$$

G. Cowan