LIP 2018
Problems on Statistics (1)
G. Cowan / RHUL

Exercise 1: An experiment yields n time values $t_1, \ldots, t_n$, and a calibration value y, all of which are independent. The time measurements are all exponentially distributed with a mean of $\tau + \lambda$ and the calibration measurement, y, follows a Gaussian distribution with a mean λ and a standard deviation σ. Suppose that σ is known and we want to estimate τ and λ.

(a) Write down the likelihood function for τ and λ, and show that the Maximum Likelihood (ML) estimators for these parameters are

$$\begin{aligned} \hat{\tau} &= \frac{1}{n}\sum_{i=1}^{n} t_i - y \,, \\ \hat{\lambda} &= y \,. \end{aligned}$$

(b) Find the variances of $\hat{\tau}$ and $\hat{\lambda}$, and the covariance $\mathrm{cov}[\hat{\tau}, \hat{\lambda}]$. Use the fact that the variance of an exponentially distributed variable is equal to the square of its mean. (It may also be useful to note that for any random variables x, y and z, $\mathrm{cov}[x + y, z] = \mathrm{cov}[x, z] + \mathrm{cov}[y, z]$.)

(c) Show using a sketch how a contour of constant log-likelihood can be used to determine the standard deviations of $\hat{\tau}$ and $\hat{\lambda}$. Explain qualitatively how you would expect the variance of $\hat{\tau}$ to be different if the parameter λ were to be known exactly.

(d) Show that the (co)variances of $\hat{\tau}$ and $\hat{\lambda}$ obtained from the matrix of second derivatives of the log-likelihood are the same as those found in (b).

Exercise 2: The binomial distribution is given by

$$f(n; N, \theta) = \frac{N!}{n!(N-n)!}\theta^n (1-\theta)^{N-n} \,,$$

where n is the number of 'successes' in N independent trials, with a success probability of θ for each trial. Recall that the expectation value and variance of n are $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$, respectively. Suppose we have a single observation of n and using this we want to estimate the parameter θ.

(a) Find the maximum likelihood estimator $\hat{\theta}$.

(b) Show that $\hat{\theta}$ has zero bias and find its variance. Find the minimum variance bound of $\hat{\theta}$.

(c) Suppose we observe $n = 0$ for $N = 10$ trials. Find the upper limit for θ at a confidence level of CL = 95% and evaluate numerically.

Exercise 3 Consider the following two pdfs for a continuous random variable x that correspond to two types of events, signal (s) and background (b):

$$\begin{aligned} f(x|\mathrm{s}) &= 2(1-x) \,, \\ f(x|\mathrm{b}) &= 4x^3 \,, \end{aligned}$$

where $0 \le x \le 1$. We want to select events of type s by requiring $x < x_{\mathrm{cut}}$ for a specified value of x_{cut}. This can be viewed as a test of the hypothesis b, whereby an event is selected if the b hypothesis is rejected. Suppose we want a test of size $\alpha = 10^{-4}$ (i.e., a background efficiency of 10^{-4}).

(a) Find $x_{\rm cut}$ and indicate the critical region of the test. Find the power M with respect to the hypothesis s (i.e., the signal efficiency) and evaluate numerically.

(b) Suppose the prior probabilities for events to be of types s and b are $\pi_{\rm s} = 0.001$ and $\pi_{\rm b} = 0.999$, respectively. Find the purity of signal events in the selected sample, i.e., the expected fraction of selected events that are of type s. Evaluate numerically.

(c) Suppose an event is observed with $x = 0.1$. Find the probability that the event is of type b and evaluate numerically.

(d) Again for an event with $x = 0.1$, find the p-value for the hypothesis that the event is of type b and evaluate numerically. Describe briefly how to interpret this number and why it is not equal to the probability found in (c).

Exercise 4: The number 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.

(a) 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

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

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` (which gives 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.

(b) Find the corresponding equivalent Gaussian significance Z and evaluate numerically.