Statistical Methods for Particle Physics
TRISEP / 27,28 June, 2016

Statistics problem on hypothesis tests

Exercise 1: Consider the following two pdfs for a continuous random variable x that correspond to two types of events, signal (s) and background (b):

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

where $0 \le x \le 1$. We want to select events of type s by requiring $x < x_{\mathrm{cut}}$, with $x_{\mathrm{cut}} = 0.1$.

(a) Find the efficiencies for selecting signal and background, i.e., the probabilities to accept events of types s and b, $\varepsilon_{\mathrm{s}} = P(x < x_{\mathrm{cut}}|\mathrm{s})$ and $\varepsilon_{\mathrm{b}} = P(x < x_{\mathrm{cut}}|\mathrm{b})$

(b) Suppose the prior probabilities for events to be of types s and b are $\pi_{\mathrm{s}} = 0.01$ and $\pi_{\mathrm{b}} = 0.99$, respectively. Find the purity of signal events in the selected sample, i.e., the expected fraction of events with $x < x_{\mathrm{cut}}$ that are of type s and evaluate numerically.

(c) Suppose an event is observed with $x = 0.05$. Find the probability that the event is of type b and evaluate numerically.

(d) Again for an event with $x = 0.05$, 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 comment on why it is not equal to the probability found in (c).

(e) Suppose in addition to x, for each event we measure a quantity y, and that the joint pdfs for the s and b hypotheses are:

$$
\begin{aligned}
f(x,y|\mathrm{s}) &= 6(x-1)^2 y , \\
f(x,y|\mathrm{b}) &= 6x^2(1-y) .
\end{aligned}
$$

Write down the test statistic $t(x,y)$ which provides the highest signal purity for a given efficiency by selecting events inside a region defined by $t(x,y) = t_{\mathrm{cut}}$, where t_{cut} is a specified constant.

Exercise 2: The number of events n observed in an experiment with a given integrated luminosity can be modeled as a Poisson variable with a mean $s + b$, where s and b are the contributions from signal and background processes, respectively. Suppose $b = 3.9$ events are expected from background processes and $n = 16$ events are observed. Compute the p-value for the hypothesis $s = 0$, i.e., that no new process is contributing to the number of events. To sum Poisson probabilities, you can use the relation

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

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 $n_{\mathrm{dof}} = 2(m+1)$ 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.

Compute the corresponding significance, $Z = \Phi^{-1}(1-p)$. To evaluate the standard normal quantile Φ^{-1} you can use the ROOT routine `TMath::NormQuantile`.