G. Cowan / RHUL Physics
4th KMI School, Nagoya
15-17 December 2022

Exercise on Hypothesis Testing

The purpose of this exercise is to design a statistical test to discover a signal process by counting events in a detector. The exercise involves calculations by hand as well as some computations using python. The program `hypTest.py` can be used as a starting point for parts (a)–(f) and `hypTestMC.py` can be used for part (g).

Suppose a detector that looks, e.g., for Dark Matter interactions can for each event measure a quantity x with $0 \le x \le 1$. The events can be of two types: signal (s) or background (b). The probability density functions for the s and b events are

$$f(x|s) = 3(1-x)^2 , \tag{1}$$

$$f(x|b) = 3x^2 . \tag{2}$$

1(a) Suppose for each event we test the hypothesis that it is background. We reject this hypothesis if the observed value of x is less than a specified cut value $x_{\rm cut}$. Find the value of $x_{\rm cut}$ such that the probability $P(x < x_{\rm cut}|b)$ to reject the background hypothesis (i.e., accept as signal) if it is background is $\alpha = 0.05$. (The value α is the *size* or significance level of the test used to select events.)

1(b) For the value of $x_{\rm cut}$ that you find, what is the probability $P(x < x_{\rm cut}|s)$ to reject the background hypothesis (i.e., accept as a candidate signal event) with $x < x_{\rm cut}$ given that it is signal. (This is the *power* of the test of the background hypothesis with respect to the signal alternative or equivalently the signal efficiency.)

1(c) Suppose that the expected number of background events is $b_{\rm tot} = 100$ and for a given signal model one expects $s_{\rm tot} = 10$ signal events. Find the expected numbers of events s and b of signal and background events that will satisfy $x < x_{\rm cut}$ using the value of $x_{\rm cut} = 0.1$, i.e.,

$$s = s_{\rm tot} P(x < x_{\rm cut}|s) , \tag{3}$$

$$b = b_{\rm tot} P(x < x_{\rm cut}|b) . \tag{4}$$

1(d) Assuming the numbers from 1(c), the prior probabilities for an event to be signal or background are

$$\pi_s = \frac{s_{\rm tot}}{s_{\rm tot} + b_{\rm tot}} = 0.09 , \tag{5}$$

$$\pi_b = \frac{b_{\rm tot}}{s_{\rm tot} + b_{\rm tot}} = 0.91 . \tag{6}$$

Using Bayes' theorem with these values, find the probability for an event to be signal given that it has $x < x_{\rm cut}$ (the signal purity of the selected sample).

1(e) Suppose for a certain x_{cut} one has $b = 0.5$ and we find there $n_{\text{obs}} = 3$ events in the search region $x < x_{\text{cut}}$. We want to test the hypothesis that $s = 0$ (the background-only hypothesis or "b"), against the alternative that signal is present with $s \neq 0$ (the "$s + b$" hypothesis).

The actual number of events n found in the experiment with $x < x_{\text{cut}}$ can be modeled as following a Poisson distribution with a mean value of $s + b$. That is, the probability to find n events is

$$P(n|s,b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \; . \tag{7}$$

The p-value of the background-only hypothesis is the probability, assuming $s = 0$, to find $n \geq n_{\text{obs}}$:

$$p = P(n \geq n_{\text{obs}}|s=0,b) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b} \; . \tag{8}$$

Find the p-value using the values given above and from this find the *significance* with which one can reject the $s = 0$ hypothesis, defined as

$$Z = \Phi^{-1}(1-p) \; , \tag{9}$$

where Φ is the standard cumulative Gaussian distribution and Φ^{-1} is its inverse (the standard Gaussian quantile).

1(f) The expected (median) significance assuming the $s+b$ hypothesis of the test of the $s = 0$ hypothesis is a measure of sensitivity and this is what one tries to maximize when designing an experiment. It can be approximated with a number of different formulas. For $s \ll b$ one can use $\text{med}[Z_b|s+b] = s/\sqrt{b}$. If $s \ll b$ does not hold, a better approximation is

$$\text{med}[Z_b|s+b] = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right) - s\right)} \; . \tag{10}$$

Using Eq. (10), find me median significance for $x_{\text{cut}} = 0.1$. If you have time, try to write a program to find the value of x_{cut} that maximizes the median significance.

1(g) Now suppose that for each event we do not simply count the events having x in a certain region but we design a test that exploits each measured value in the entire range $0 \leq x \leq 1$. Thus there is no cut on x and in here we use $s = 10$ and $b = 100$ to refer to the total expected numbers of signal and background events. The data consist of the number n of events, which follows a Poisson distribution with mean of $s + b$, and the n values $x_1, \ldots, x_n$.

The Poisson probability to find n events can be written in terms of a strength parameter μ as

$$P(n|\mu) = \frac{(\mu s+b)^n}{n!} e^{-(\mu s+b)} \tag{11}$$

where $\mu = 0$ corresponds to the background only hypothesis and $\mu = 1$ adds to this the contribution from the signal. The joint distribution for $\mathbf{x} = (x_1, \ldots, x_n)$ given n is

$$f(\mathbf{x}|n,\mu) = \prod_{i=1}^{n} \left[\frac{\mu s}{\mu s+b} f(x_i|s) + \frac{b}{\mu s+b} f(x_i|b)\right] \; , \tag{12}$$

and the full likelihood is therefore

$$
\begin{aligned}
L(\mu) &= P(n, \mathbf{x}|\mu) = P(n|\mu) f(\mathbf{x}|n, \mu) \\
&= \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \prod_{i=1}^{n} \left[\frac{\mu s}{\mu s + b} f(x_i|s) + \frac{b}{\mu s + b} f(x_i|b) \right] \ . \qquad (13)
\end{aligned}
$$

We can define a statistic q to test the background-only hypothesis as as

$$
q = -2 \sum_{i=1}^{n} \ln \left[1 + \frac{s}{b} \frac{f(x_i|s)}{f(x_i|b)} \right] = -2 \ln \frac{L(1)}{L(0)} + C \ , \qquad (14)
$$

where C is a constant that can be dropped (the factor of -2 is conventional and could be omitted). This is a monotonic function of the likelihood ratio $L(1)/L(0)$ and thus according to the Neyman-Pearson lemma gives a test of $\mu = 0$ with the highest possible sensitivity (highest power with respect to the alternative of $\mu = 1$).

Run the program `hypTestMC.py`. This will produce histograms of q under the b and $s+b$ hypotheses, corresponding to the distributions $f(q|\mu)$ for $\mu = 0$ and $\mu = 1$, respectively. The program also finds the median q for the $s + b$ hypothesis, $\mathrm{med}[q|s + b]$.

You should add code that finds the median p-value of the b-only hypothesis for the data generated under assumption of the $s+b$ hypothesis. That is, find the fraction of experiments simulated according to b-only that have $q < \mathrm{med}[q|s + b]$. From this, find the p-value of the background-only hypothesis and the corresponding median significance $\mathrm{med}[Z_b|s + b]$ (the sensitivity). Compare to the values you found from Eq. (10).

To find the value with reasonable accuracy you will need to simulate 10^7 experiments, which could take about 30 minutes to compute (the default value of `numExp` is set to 10^6). To test your code, use at first a smaller number of experiments.