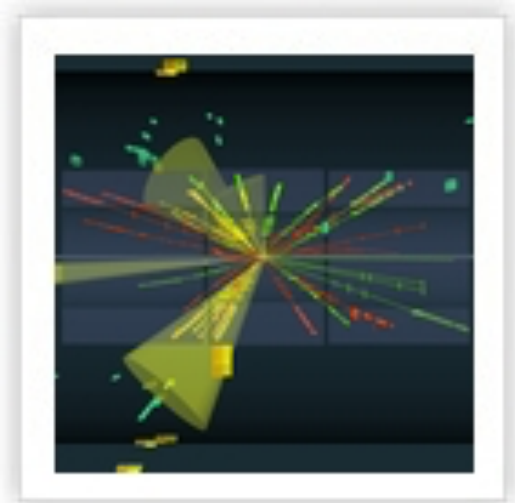


# Statistical Methods for Particle Physics

## Tutorial: statistical test for discovery

<http://benasque.org/2018tae/cgi-bin/talks/allprint.pl>



TAE 2018

Centro de ciencias Pedro Pascual

Benasque, Spain

3-15 September 2018

Glen Cowan

RHUL Physics

# Problem 1 – discovering a small signal

Materials at [www.pp.rhul.ac.uk/~cowan/stat/invisibles/](http://www.pp.rhul.ac.uk/~cowan/stat/invisibles/)

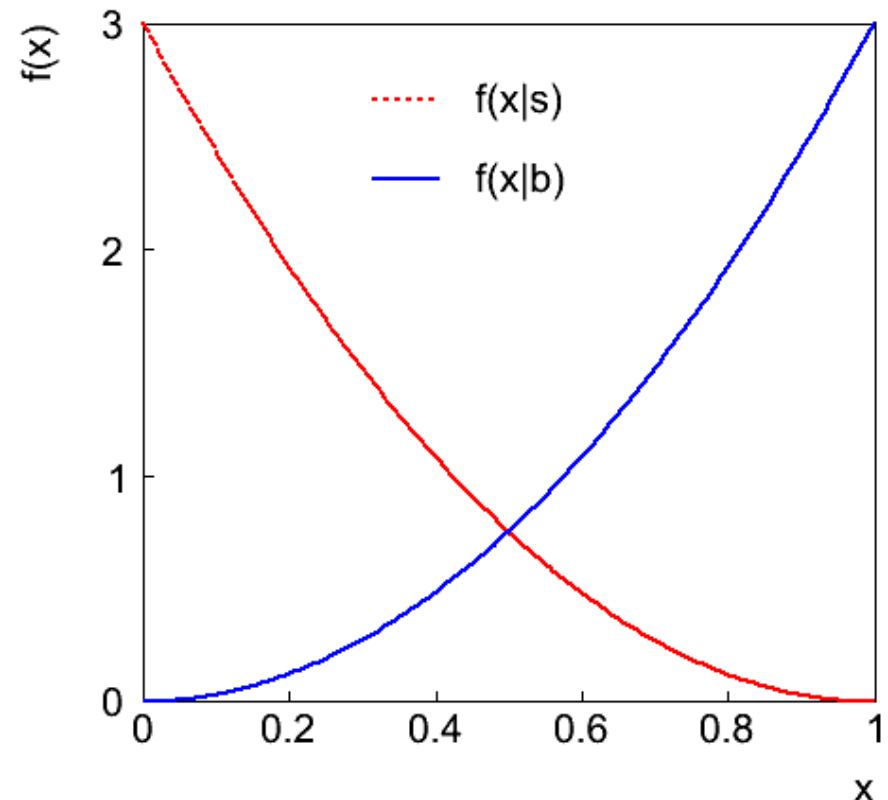
Problem concerns searching for a signal such as Dark Matter by counting events. Suppose signal/background events are characterized by a variable  $x$

( $0 \leq x \leq 1$ ):

$$f(x|s) = 3(1-x)^2,$$

$$f(x|b) = 3x^2.$$

As a first step, test the background hypothesis for each event: if  $x < x_{\text{cut}}$ , reject background hypothesis.



# Testing the outcome of the full experiment

In the full experiment we will find  $n$  events in the signal region ( $x < x_{\text{cut}}$ ), and we can model this with a Poisson distribution:

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

Suppose total expected events in  $0 \leq x \leq 1$  are  $b_{\text{tot}} = 100$ ,  $s_{\text{tot}} = 10$ ; expected in  $x < x_{\text{cut}}$  are  $s$ ,  $b$ .

Suppose for a given  $x_{\text{cut}}$ ,  $b = 0.5$  and we observe  $n_{\text{obs}} = 3$  events. Find the  $p$ -value of the hypothesis that  $s = 0$ :

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

and the corresponding significance:  $Z = \Phi^{-1}(1 - p)$

# Experimental sensitivity

To characterize the experimental sensitivity we can give the median, assuming  $s$  and  $b$  both present, of the significance of a test of  $s = 0$ . For  $s \ll b$  this can be approximated by

$$\text{med}[Z_b|s + b] = s/\sqrt{b}$$

A better approximation is:

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

Try this for  $x_{\text{cut}} = 0.1$  and if you have time, write a small program to maximize the median  $Z$  with respect to  $x_{\text{cut}}$ .

Tomorrow we will discuss methods for including uncertainty in  $b$ .

## Using the $x$ values

Instead of just counting events with  $x < x_{\text{cut}}$ , we can define a statistic that takes into account all the values of  $x$ . I.e. the data are:  $n, x_1, \dots, x_n$ . Tomorrow we will discuss ways of doing this with the likelihood ratio  $L_{s+b}/L_b$ , which leads to the statistic

$$q = -2 \sum_{i=1}^n \left[ 1 + \frac{s_{\text{tot}}}{b_{\text{tot}}} \frac{f(x_i|s)}{f(x_i|b)} \right]$$

Using [www.pp.rhul.ac.uk/~cowan/stat/invisibles/mc/invisibleMC.cc](http://www.pp.rhul.ac.uk/~cowan/stat/invisibles/mc/invisibleMC.cc) find the distribution of this statistic under the “ $b$ ” and “ $s+b$ ” hypotheses.

From these find the median, assuming the  $s+b$  hypothesis, of the significance of the  $b$  (i.e.,  $s = 0$ ) hypothesis. Compare with result from the experiment based only on counting  $n$  events.