Sensitivity studies for signal discovery

Anil Kumar, Luigi Pertoldi, Mario Schwarz

INSS18, Mainz — 31/05/2018

TOC

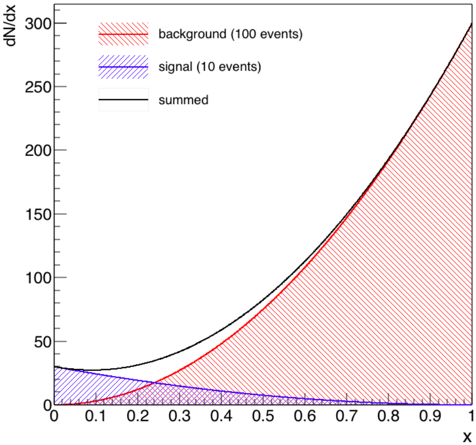
The problem

An experiment measures values of $x \in [0, 1]$.
The PDFs for background and signal events are:

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

We'll consider the following hypothesis

- H_0 : (b) 100 bkg expected
- H_1 : $(b+s)$ 100 bkg + 10 sig expected



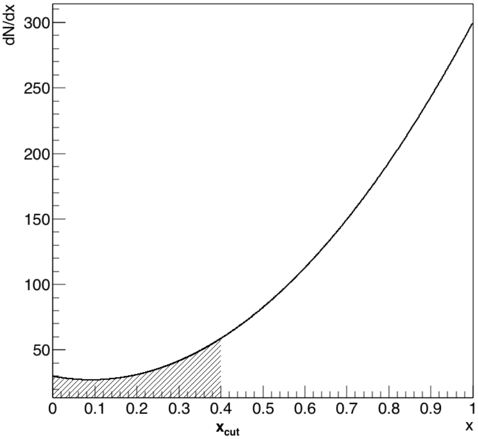
Analysis 1: basic cut

- Set x_{cut} to select the signal events
- The number of observed events n_{obs} with $x < x_{cut}$ follows a poissonian with mean $b + s$
- The p-value of the background-only hypothesis

$$p = P(n \geq n_{obs}|H_0)$$

measures the significance of an observation. For example with $b = 0.5$ and $n_{obs} = 3$

$$p = 0.014 \quad (2.2\sigma)$$



Analysis 1: basic cut

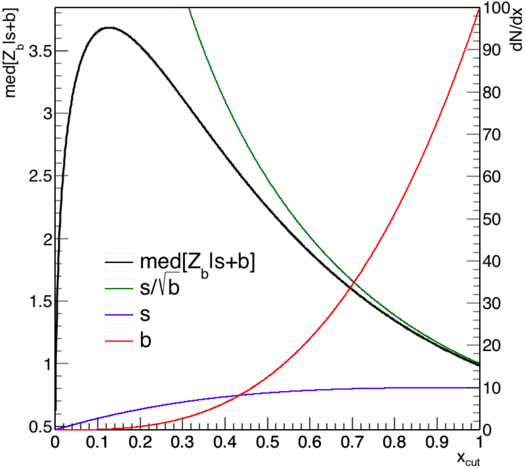
The expected (median) significance of the test of the null (H_0) hypothesis, assuming the H_1 hypothesis as true, is a measure of sensitivity.

To a good approximation (Wilks theorem):

$$\text{med}[Z_b|H_1] = \sqrt{2\left[(s+b)\log\left(1+\frac{s}{b}\right) - s\right]}$$

that reaches its maximum at

$$(x_{\text{cut}} = 0.127, \text{med}[Z_b|H_1] = 3.7\sigma)$$



Analysis 2: unbinned data

We try to use now a test that takes into account each single measured x_i with $i = 1, \ldots, n$

$$L_{H_0} = P(n, \mathbf{x}|H_0) = \frac{B^n}{n!} e^{-B} \prod_{i=1}^{n} f(x_i|b)$$

$$L_{H_1} = P(n, \mathbf{x}|H_1) = \frac{(B+S)^n}{n!} e^{-(B+S)} \prod_{i=1}^{n} \left[\frac{S}{B+S} f(x_i|s) + \frac{B}{B+S} f(x_i|b) \right]$$

Where B and S are the *total* number of expected bkg and signal events. We are interested in the ratio L_{H_1}/L_{H_0}, some algebra gives:

$$\frac{L_{H_1}}{L_{H_0}} \propto \prod_{i=1}^{n} \left[1 + \frac{S}{B} \frac{f(x_i|s)}{f(x_i|b)} \right]$$

Analysis 2: unbinned data

We are interested in the ratio L_{H_1}/L_{H_0}, some algebra gives:

$$\frac{L_{H_1}}{L_{H_0}} \propto \prod_{i=1}^{n} \left[1 + \frac{S}{B} \frac{f(x_i|s)}{f(x_i|b)} \right]$$

Therefore, the following test can be adopted:

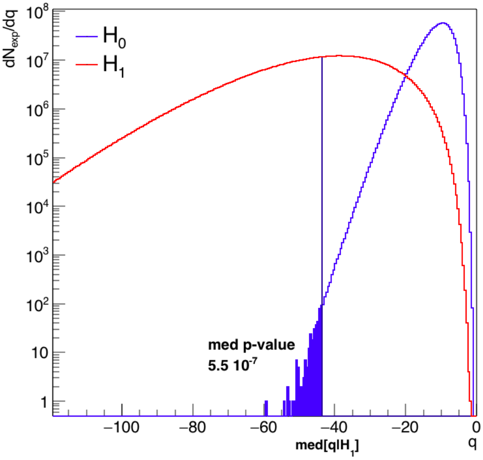
$$q = -2 \sum_{i=1}^{n} \log \left[1 + \frac{S}{B} \frac{f(x_i|s)}{f(x_i|b)} \right]$$

Analysis 2: unbinned data

1. Generate $2N$ experiments (H_0 and H_1)
2. compute $q \to$ histogram
3. compute the fraction of q values, assuming H_0 true, with $q < \text{med}[q|H_1]$

This is the median p-value of the background-only hypothesis assuming H_1 true

$$\text{med}[p|H_1] = 5.5 \cdot 10^{-7} \quad (5\sigma)$$



[Extra] Analysis 3: binned data

We now histogram the data in N bins. This leads to the following likelihood function:

$$L(\mu) = \prod_{i=1}^{n} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}$$

The following test statistics can be defined:

$$q_0 = -2 \log \lambda(0) \Theta(\hat{\mu}) \qquad \lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})}$$

where $\hat{\mu}$ maximizes $L(\mu)$. It can be shown (Wilks theorem) that:

$$f(q_0|H_0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-\frac{q_0}{2}} = \frac{1}{2}\delta(q_0) + \frac{1}{2}\chi_1^2$$

[Extra] Analysis 3: Wilks theorem verification

We wrote a small program to compute $f(q_0|H_0)$:

1. simulate experiment with 10^3 bkg-only counts
2. find $\hat{\mu}$ minimizing $-\log L(\mu)$
3. histogram values of q_0
4. check shape $\rightarrow$ OK!

Julia code:

`https://tinyurl.com/wilksth-jl`