limit-setter

A frequentist hypothesis testing tool

Abstract

This short note provides a brief description of a python-based binned hypothesis-testing tool, based on the CL_s method. The code can be found at `https://github.com/MichaelRussell2/limit_setter`.

Contents

1 Introduction

Most limit-setting in particle physics boils down to choosing between two hypotheses to describe the data: a *null* hypotheses H_0 (typically that the data x is described by the background model alone $H_0 = b$) and a *test* hypothesis H_1 (that the data is described by signal plus background $H_1 = s + b$). The Neyman-Pearson lemma states that the most powerful test statistic, i.e. the one which minimises the false positive rate for a given false negative rate, is the likelihood ratio,

$$Q = \frac{L(x|H_1)}{L(x|H_0)}. \tag{1}$$

In the absence of systematic uncertainties, the likelihood ratio is exactly described by a ratio of Poisson probabilities for obtaining d data events in a distribution with Poisson means $s+b$ and b, respectively.

$$Q = \frac{P(d|s+b)}{P(d|b)} = \frac{e^{-(s+b)}(s+b)^d/d!}{e^{-b}b^d/d!} = e^{-s}\left(1+s/b\right)^d. \tag{2}$$

If instead of a single measurement, the data consists of several bins or channels i, the combined distribution is then a product of Poisson distributions $Q = \prod_i Q_i$. In practice, to avoid issues of numerical precision in the limit of very small or large d, the logarithm of this quantity is usually taken as the test statistic

$$q \equiv -2\log Q = -2\sum_i \left(-s_i + d_i \log\left(1+\frac{s_i}{b_i}\right)\right), \tag{3}$$

where the factor of -2 is a convention. The confidence levels for excluding the $s+b$ and b-only hypotheses are

$$CL_{s+b} = P_{s+b}(q > q_{obs}), \qquad CL_b = P_b(q > q_{obs}). \tag{4}$$

These represent, respectively, the probability that the test statistic q would be less than that observed in the data, given the hypothesised number of signal and background events $s+b$ or background-only events b.

In practice, we numerically evaluate these p-values by generating a large number of Monte Carlo pseudo-experiments, with CL_{s+b} being the fraction of pseudo-experiments that generate *at least as many* events as observed in the data. Instead of excluding regions for which $CL_{s+b} \leq 0.05$, we take the modified frequentist CL_s procedure, which only excludes this hypothesis if $CL_{s+b}/1 - CL_b \leq 0.05$, thus providing robustness against spuriously high sensitivity in regions of parameter space where both s and b are small, at the price of over-conservatism in other regions.

2 Description

2.1 Algorithm

Typically, to set exclusion limits for phenomenological purposes, one just assumes the "data" is exactly equal to the background in each bin, then computes the exclusion probability P_{s+b}. Assuming we have two histograms s_i and b_i, (typically $s_i \ll b_i \ \forall\ i$), the algorithm for doing this is then:

- For each bin i:
  - Evaluate the log-likelihood ratio $q_{obs} = -2(-s_i + d_i \log\left[1 + s_i/b_i\right])$, taking $d_i = b_i$.

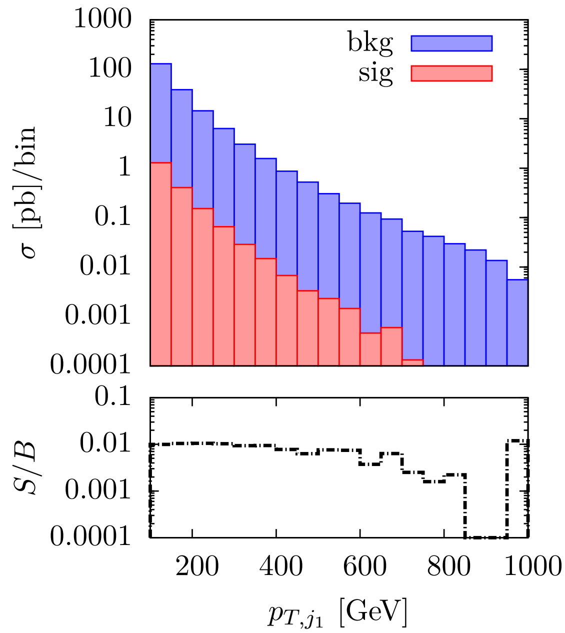

Figure 1: The histogram used in this example: p_T of the leading jet in H+jets events generated with MadGraph+Pythia (signal, red) and in $Z(\to \nu\nu)$+jets generated with Sherpa (background, blue). The lower panel plots the ratio.

- Initialise two counters, N_{null} and N_{test} to zero.
- Run a large number (at least 10^4) of Monte Carlo "pseudoexperiments". For each pseudoexperiment:
  * Generate pseudo-data from Poisson distributions with mean $\mu = b_i$ for the null hypothesis $d^{MC}_{null} = Pois(b_i)$ and $\mu = s_i + b_i$ for the test hypothesis $d^{MC}_{test} = Pois(s_i + b_i)$.
  * Calculate the log-likelihood ratios q^{MC}_{null} and q^{MC}_{test} corresponding to d^{MC}_{null} and d^{MC}_{test}.
  * If $q_{obs} < q^{MC}_{null}$, iterate N_{null} by 1. Similarly, if $q_{obs} < q^{MC}_{test}$, increase N_{test} by 1.
- For this bin, the p-value to exclude the test hypothesis is given by $CL_{s+b} = N_{test}/N_{MC}$, i.e. the fraction of pseudoexperiments which have a test statistic *at least as large* as the one observed. Likewise for the null hypothesis.

- Assuming the bins are independent, the combined p-values CL_{s+b} and CL_b are the products of the individual p-values for each bin. Then $CL_s = CL_{s+b}/CL_b$.

2.2 A physics example

As a test case, we use the p_T of the leading jet in H+jets events as a signal (with the Higgs decaying invisibly with $BR_{\text{inv}} = 50\%$), and $Z(\to \nu\nu)$+jets as background, plotted in

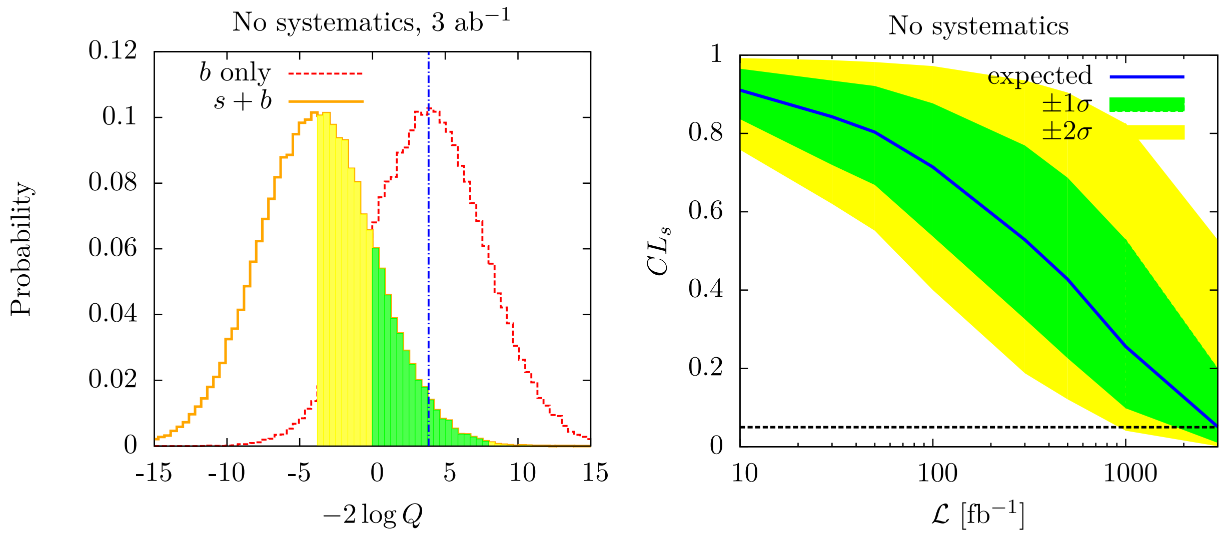

Figure 2: Left: Log-likelihood distributions for the signal plus background (orange, solid) and background only (red, dashed) hypotheses, using the first bin of the distributions of Fig. **??**. The data is taken to be the background, i.e. the median of the background-only distribution, represented by the blue line. Also shown are the "Brazil bands" obtained by fluctuating the data by 1 and 2 σ. Right: Corresponding CL_s versus luminosity.

Fig. **??**. Firstly, we consider the case of no background systematics at all (i.e. only statistical uncertainties). The distributions for the log-likelihood ratio for signal plus background and background only are shown in Fig. 2, using only the first bin of the p_T distribution in Fig. **??**, based on running 10^4 pseudoexperiments. We see the characteristic Poisson curves for the two hypotheses, along with the point corresponding to the actual "data" at the median of the null hypotheses. This figure makes it easier to visualise CL_{s+b} as the integral of $s+b$ distribution from the blue line to $x=\infty$, in the limit of large statistics. Similarly, CL_b is the corresponding integral of the background only hypothesis. If we take the data as exactly equal to the background, $CL_b = 0.5$ by construction.

Often, we are interested not just in the central value of CL_s, but how sensitive it is to random fluctuations in the data. This is typically quantified by plotting the bands obtained by fluctuating the data d up and down by 1 and 2 standard deviations; since we take the data to be Poisson distributed this is simply $\pm(2)\sqrt{d}$. The corresponding "Brazil bands" are also shown in Fig. 2. In the absence of any systematic uncertainties, the two distributions will separate more with more data. This is shown on the right of Fig. 2, where we plot the central CL_s and Brazil bands for this hypothesis test, as a function of luminosity. We see CL_s tends to zero, just as the cut-and-count significance $S/\sqrt{B}$ tends to infinity with luminosity. We also see that the Brazil bands are not necessarily symmetric around the central value, because the integrals of the likelihood distributions are not necessarily symmetric around the likelihood from the data. In real applications, other sources of uncertainty (systematics) exist, and the floor in sensitivity is set by the point when they dominate over statistical errors. The next section describes the modelling

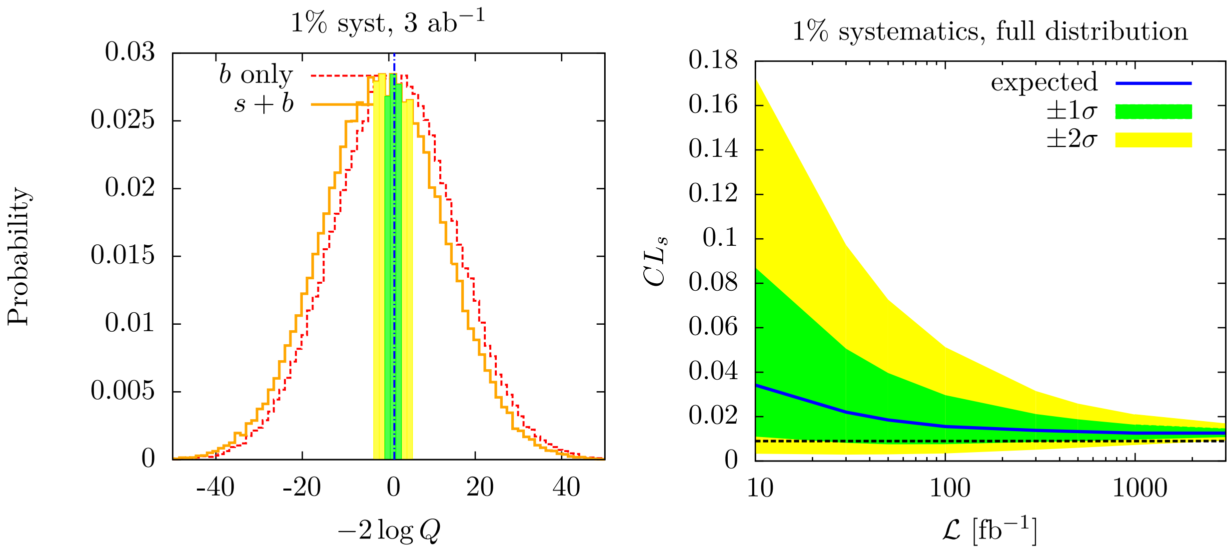

Figure 3: Left: As in Fig. 2 but with a 1% systematic uncertainty on the background.

of these sources of uncertainty.

2.3 Systematics

There is no a priori correct way of dealing with systematic uncertainties in distributions and several prescriptions are common in high-energy physics. In this tool we adopt a Bayesian approach; treating each source of systematic as a nuisance parameter by convoluting the individual Poisson likelihoods in Eq. (2) with Gaussians with standard deviation σ_b and σ_s, as advocated in the Feldman-Cousins procedure [1]. Effectively this means the probability distribution for signal+background becomes:

$$P(d|s+b) = \frac{1}{2\pi\sigma_b}\int_0^\infty db' P(d|s+b') e^{-\frac{(b-b')^2}{2\sigma_b^2}} \tag{5}$$

Uncertainties on the signal are usually neglected since typically $s \ll b$. This Bayesian marginalisation procedure typically reduces the sensitivity by "smearing out" the log-likelihood distributions for the two hypotheses, thus reducing the distinction between s and $s+b$. The corresponding log-likelihood distributions for $s+b$ and b-only in the presence of a 1% background uncertainty are shown on the left of Fig. 3. The median of the distributions stays the same, but the convolution in Eq. (5) broadens the distributions and considerably reduces the sensitivity.

On the right of Fig. 3 we plot the corresponding CL_s versus luminosity. This time the sensitivity in the CL_s central-value tails off at around 100 fb^{-1}. The upper Brazil bands, obtained by fluctuating the data upwards, saturate more slowly because statistical errors will dominate for longer if S/B is smaller. On the other hand, counter-intuitively, the lower bands appear to slightly worsen with luminosity, because the broadening of the likelihood distributions pushes the 1 and 2 σ bands closer to the central value. We emphasise here that the Brazil bands are a representation of *statistical uncertainties only*, and are not particularly useful in cases where systematics dominate, as in this one.

3 Code details

References

[1] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998) doi:10.1103/PhysRevD.57.3873 [physics/9711021 [physics.data-an]].