The CL_s method: information for conference speakers

This note provides a brief description of the CLs procedure used for setting upper limits. More information can be found in the original references for the method [1]. This is one of the three methods for setting limits mentioned in the Review of Particle Physics by the PDG [2], and has been widely used in HEP in recent years. The primary motivation for using CLs at this time in ATLAS is to allow for comparison with other experiments (CMS and Tevatron).

As with all (frequentist) upper limits, those from the CLs method are desiged to be greater than the true value of the parameter with a probability at least equal to the stated confidence level (CL), taken by convention to be 95%. The CLs method is conservative in the sense that this coverage probability can, depending on the true value of the parameter, be greater than 95% (see below).

Upper limits from the CLs procedure are the same as those from the Bayesian method in two important special cases, namely, for limits on the mean value of a Poisson or Gaussian distributed measurement. In both cases, a Bayesian limit based on a constant prior for the mean leads to the same limit as CLs.

Background information

We assume that the analyst has constructed a test statistic q used to distinguish between the hypothesis that the data contain signal and background $(s+b)$ and that of background only (b). These correspond to the distributions $f(q|s+b)$ and $f(q|b)$, as indicated in Fig. 1. For the moment we leave open the details of how the test statistic q is defined.
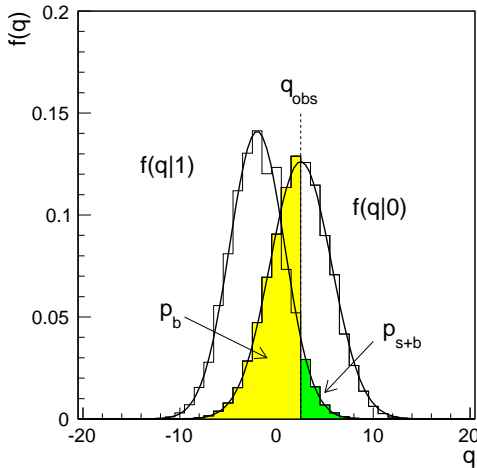

Figure 1: Distributions of the test variable q under the $s+b$ and b hypotheses (see text).

Suppose the actual data result in a value q_{obs} of the test variable. The p-value of the $s+b$ hypothesis is defined as the probability, under assumption of this hypothesis, to find a value of q with equal or lesser compatibility with the $s+b$ model relative to what is found with q_{obs}. As the background-only distribution $f(q|b)$ is here shifted to the right, one takes the p-value of $s+b$ to be the probability to find q greater than or equal to q_{obs}, under assumption of the $s+b$ hypothesis, i.e.,

$$p_{s+b} = P(q \geq q_{\mathrm{obs}}|s+b) = \int_{q_{\mathrm{obs}}}^{\infty} f(q|s+b)\, dq \ . \tag{1}$$

In a similar way, one takes the p-value of the background-only hypothesis to be

$$p_b = P(q \leq q_{\mathrm{obs}}|b) = \int_{-\infty}^{q_{\mathrm{obs}}} f(q|b)\, dq \ . \tag{2}$$

In what is called the "CL$_{s+b}$" method, one carries out a standard statistical test of the $s+b$ hypothesis based on its p-value, p_{s+b}. The signal model is regarded as excluded at a confidence level of $1 - \alpha = 95\%$ if one finds

$$p_{s+b} < \alpha \ , \tag{3}$$

where, e.g., $\alpha = 0.05$. A confidence interval at confidence level $\mathrm{CL} = 1 - \alpha$ for the rate of the signal process can be constructed from those values of the rate s (or cross section) that are not excluded, and the upper limit s_{up} is the largest value of s not excluded. By construction, the interval $[0, s_{\mathrm{up}}]$ will cover s with a probability of at least 95%, regardless of the value of s.

The problem with the CL$_{s+b}$ procedure is that one will exclude, with probability close to α (i.e, 5%) hypotheses to which one has little or no sensitivity. This corresponds to the case where the expected number of signal events is much less than that of background. Such a scenario is illustrated in Fig. 2, and corresponds to having the distributions of q under both the b and $s+b$ hypotheses almost overlapping with each other.

If, for example, the expected numbers of signal and background events are s and b, respectively, and one has $s \ll b$, then if the observed number of events has a sufficient downward fluctuation relative to $s+b$ (which is approximately equal to b), then this value of s will be excluded. In the limit where $s \ll b$, one might want intuitively this exclusion probability to go to zero, but in fact in the CL$_{s+b}$ procedure it approaches $\alpha = 5\%$. Given that one carries out many tests for different signal models, it is not desirable that one out of twenty searches where one has no sensitivity should result in exclusion.
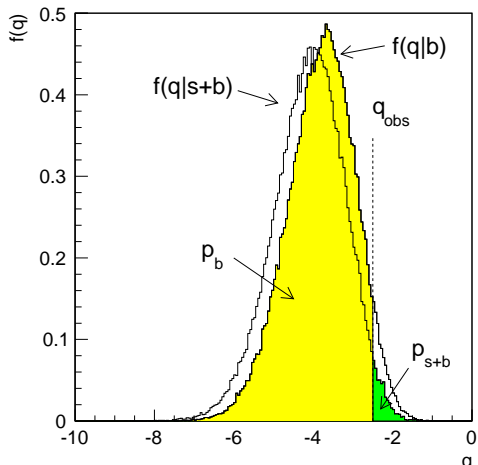

Figure 2: Distributions of the test variable q under the $s+b$ and b hypotheses in an example where one has very little sensitivity to the signal model.

To protect against excluding models to which one has little or no sensitivity, in the CL$_s$ procedure a signal model is regarded as excluded if one finds

$$\mathrm{CL}_s \equiv \frac{p_{s+b}}{1 - p_b} < \alpha \, . \tag{4}$$

That is, the p-value is effectively penalized by dividing by $1-p_b$. If the two distributions $f(q|b)$ and $f(q|s+b)$ are widely separated, then $1 - p_b$ is only slightly less than unity, the penalty is slight, and thus exclusion based in CL_s is similar to that obtained from the usual p-value p_{s+b}. If, however, one has little sensitivity to the signal model, then the two distributions are close together, $1 - p_b$ becomes small, and thus the p-value of $s + b$ is penalized (increased) more. In this way one is prevented from excluding signal models in cases of low sensitivity. As previously, one takes the upper limit to be the largest value of the parameter (e.g., the signal rate s) not excluded.

From the definition (4), one can see that CL_s is always greater than the p-value p_{s+b}. Thus the models excluded by requiring $\mathrm{CL}_s < \alpha$ are a subset of those excluded by the usual criterion $p_{s+b} < \alpha$, and the upper limit from CL_s is therefore higher (weaker). In this sense the CL_s procedure is conservative.

One can compute, e.g., using Monte Carlo, the coverage probability by generating data according to the $s + b$ model and for each simulated experiment one can see if the assumed value of s is above or below the upper limit obtained. The example shown in Fig. 3 shows the coverage probability of the CL_s interval for the case of a Gaussian distributed measurement with a mean $\mu \geq 0$ and (known) standard deviation σ, here taken to be unity. That is, in this example the parameter μ takes on the role of s above.
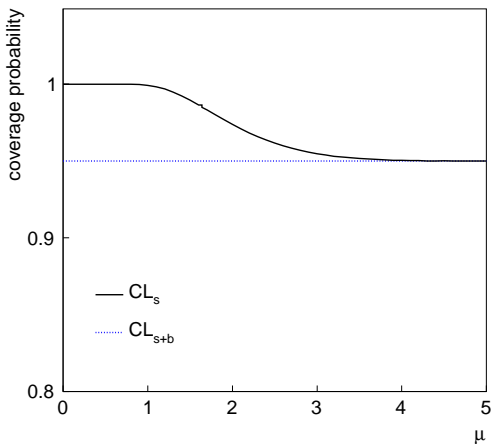

Figure 3: The coverage probability of the CLs upper limit for the mean of a Gaussian distributed measurement (see text).

In the CL_s procedure must define the statistic q. In ATLAS, it is recommended to use either the statistic q_μ or $\tilde{q}_\mu$ as described in [3], both of which are based on the profile likelihood ratio. By using the profile treatment of nuisance parameters one incorporates systematic uncertainties. This is similar but not identical to the so-called hybrid method, where the nuisance parameters are treated in a Bayesian fashion.

When using q_μ or $\tilde{q}_\mu$, the parameter on which one sets a limit is μ, defined as the cross section of the signal process divided by the predicted cross section of the nominal signal model. For a sufficiently large data sample, the distributions $f(q_\mu|\mu)$ or $f(\tilde{q}_\mu|\mu)$ needed to compute p-values can be written down in closed form as described in Ref. [4], and a simple formula for the CLs-modified p-value $p_s/(1 - p_b)$ valid for the large-sample case can be found in Ref. [3]. Otherwise the p-values must be determined by Monte Carlo. Note that the distributions of q_μ and $\tilde{q}_\mu$ are not Gaussian as in the examples of Figs. 1 and 2, but rather are related to the chi-square distribution.

3

References

[1] A.L. Read, J. Phys. G **28**, 2693 (2002); T. Junk, Nucl. Instrum. Methods Phys. Res., Sec. A **434**, 435 (1999).

[2] K. Nakamura et al. (Particle Data Group), J. Phys. G 37, 075021 (2010); `pdg.lbl.gov`.

[3] ATLAS Statistics Forum, *Frequentist Limit Recommendation* from `https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTTools/`.

[4] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, Eur. Phys. J. C 71 (2011) 1-19.