DRAFT 0.1

Glen Cowan
9 June, 2011

Some Properties of Limits

This note summarizes some properties of several limit procedures that have come under consideration for use in LHC searches. It is not in its current form intended to represent a consensus view, but rather only provide a basis for discussion. For now only frequentist limits are described. Bayesian limits (and their frequentist properties) will be considered at a future point.

1 Introduction and prototype analysis

As a useful prototype analysis, we consider a measured quantity x is modeled as following a Gaussian distribution with mean μ and standard deviation σ. Taking parameter $\mu = 0$ corresponds to the null or "no-signal" or "background-only" model. Often we consider signal models where μ cannot be negative.

Rejecting the background-only model is the key step towards claiming discovery of new physics. In addition to testing the $\mu = 0$ hypothesis, one may also want to test nonzero μ and see which values can be rejected. The largest (smallest) values not reject may be reported as upper (lower) limits on the parameter.

To carry out the procedure outlined above, one defines for each value of μ a statistical test. This can be done by taking a fixed significance level α and finding the corresponding critical region. If the data fall in the critical region, the parameter value is rejected. Those parameter values not rejected form a confidence interval with confidence level $1 - \alpha$.

Equivalently one may for every hypothesized parameter value specify a rule by which the data outcomes are ordered in terms of their level of compatibility with the hypothesis. The p-value of the hypothesized parameter is the probability, under assumption of that hypothesis, to find an outcome in the region of equal or lesser compatibility relative to what was observed in the real data. If the p-value of a hypothesized parameter value μ is found below α, then this corresponds to finding the data in the critical region and μ is rejected. Thus the edges of the confidence interval are those values of the parameter for which the p-value is just at the threshold α.

2 Limits for different kinds of analyses

The test chosen as the basis for the confidence interval will depend in general on the nature of the analysis. We could consider, e.g., the following scenarios:

1. The signal is not yet established and physically restricted to $\mu > 0$. (Here "established" could be taken in a loose sense meaning "broadly recognized as a real process" or could correspond to a specific numerical criterion such as a 5σ discovery threshold.)

2. The signal is not yet established and there is no restriction on the sign of μ.

3. The signal is well established. That is, values of μ both above and below some range have been excluded, and in particular $\mu = 0$ has been rejected in convincing enough fashion that there is a general consensus that the signal exists.

4. The signal is not well established and the signal model is characterized by more than one parameter of interest, e.g., a strength parameter μ and a particle mass m_H, or the case of neutrino oscillations with a mass difference squared Δm^2 and a mixing parameter $\sin^2 2\theta$.

3 Criteria to consider when constructing limits

When constructing a limit, one may wish to consider the follow criteria:

1. Coverage. The probability for the confidence interval to include the true parameter value should be greater than or equal to the confidence level $1 - \alpha$.

2. Power. One may wish to construct the test of each parameter value so as to maximize the probability to reject it under assumption of some stated alternative.

3. Choice of alternative for test. For example, when testing a value μ_0 one may be regard the alternative to be $\mu \neq \mu_0$, or one might have a one-sided alternative, e.g., $\mu < \mu_0$.

4. Conditional coverage. One may wish to construct the interval so that it does not contain identifiable data outcomes such that the maximum coverage probability over all values of μ and conditional on the data outcome is less than the stated confidence level (which would indicate the presence of negatively biased relevant subsets [4]).

5. Adapted conditional coverage. One may wish to require that the maximum conditional coverage not be below the confidence level for a specified subset of parameter values.

6. Bayesian correspondence. One may wish to require that the resulting confidence interval matches, at least approximately, the interval that would result from a Bayesian analysis for some prior.

7. Flip-flopping. One would like the procedure to be such that there is no violation of the stated coverage properties as a result of a choosing what result is reported depending on the data outcome.

4 Properties of limits

Limits from the PCL [1], CLs [2] and Feldman-Cousins [3] methods are shown in Fig. 1(a), and the corresponding coverage probabilities are shown in Fig. 1(b). The properties of these limits are discussed below.

4.1 Properties of CLs

The CLs upper limit for the Gaussian problem is shown in Fig. 1(a). The coverage of CLs is everywhere greater than then nominal confidence level, as can be seen in Fig. 1(b).

The CLs procedure corresponds to a one-sided upper limit. That is, both the numerator, CL_{s+b}, and denominator, CL_b are based on (different) one-sided tests.

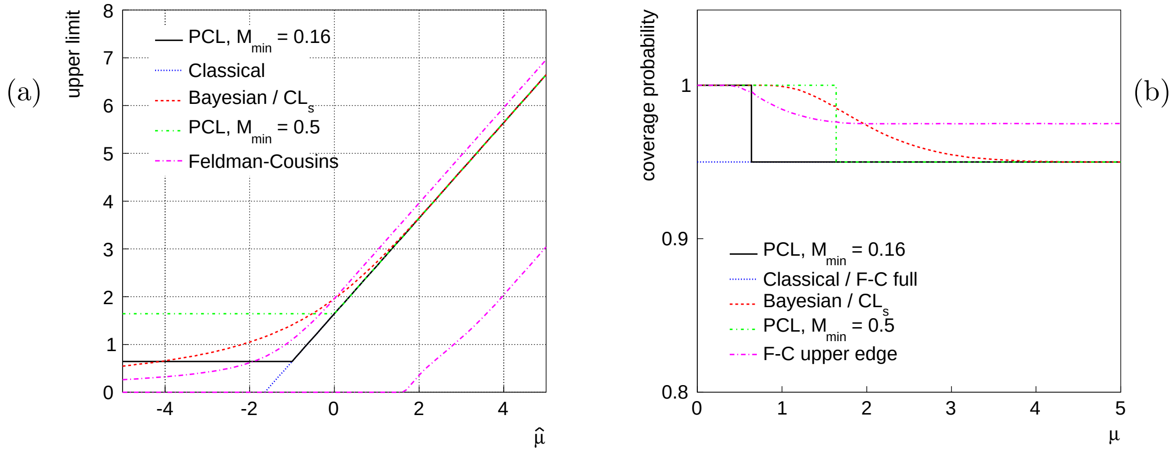

Figure 1: Plots of (a) several types of limits and (b) their coverage as a function of μ.

For the Gaussian problem, CLs coincides with the Bayesian limit obtained with a constant prior for $\mu \geq 0$.

Flip-flopping is potentially an issue, but could be handled by always quoting an upper limit. “Always” in this context could be taken to mean in all analyses that are carried out in the search phase of the phenomenon, i.e., before the existence of the process is well established.

4.2 Properties of PCL

The Power-Constrained Limits are shown in Fig. 1(a) for the Gaussian problem.

The unconstrained confidence interval is first constructed using a likelihood ratio with respect to a one-sided alternative. Values of μ rejected by the test may then be re-included in the interval if the power of the test with respect to the $\mu = 0$ alternative is below the power threshold $M_{\min}$.

The choice of $M_{\min}$ is a matter of convention. Formally it should be large compared to α, and initially the value $M_{\min} = \Phi(-1) = 0.16$ was proposed. This choice could be revisited. The PCL interval for $M_{\min} = 0.50$ is also shown in Fig. 1. This choice would have the advantage that one would see fewer instances where the PCL and CLs limits differ numerically by large amounts, reducing the amount of “surprise” in the community. The ability to state easily in what regions PCL has 95% and 100% coverage is retained.

The coverage of PCL intervals is 100% for μ less than some value $\mu_{\min}$. This corresponds to the value for which the power is just equal to $M_{\min}$. For $\mu > \mu_{\min}$ the coverage is $1 - \alpha = 95\%$.

Because of the nature of the initial likelihood-ratio test, the point $\mu = 0$ is never excluded. Therefore, formally, even the unconstrained interval does not have negatively biased relevant subsets.

After application of the power constraint, the region of μ values with 95% coverage, i.e., $\mu > \mu_{\min}$ has negatively biased subsets, if this concept is adapted to allow conditioning on the true value μ.

The PCL interval does not naturally map onto a Bayesian interval.

Flip-flopping is potentially an issue, but could be handled by always quoting an upper limit. "Always" in this context could be taken to mean in all analyses that are carried out in the search phase of the phenomenon, i.e., before the existence of the process is well established.

PCL intervals require that one be able to produce the distribution of unconstrained limits. In some cases this can be problematic. For example, to be conservative an experimenter may intentionally over-estimate the systematic uncertainties by not constraining nuisance parameters as tightly as one might believe is realistic. As a result, the $\pm 1\sigma$ bands of the distribution of unconstrained limits becomes broader, and in particular the lower edge of the -1σ band decreases. If the unconstrained limit fluctuates low, then the value at which the power constraint is applied would be lower, and thus the quoted limit is lower. So by being conservative (with the systematics) one can wind up being more aggressive (with the limit). This behaviour would be absent if one chose the minimum power to be 50%, in that a shift in the width of the limit distribution does not affect its median.

As with CLs, flip-flopping is potentially an issue, but could be handled by always quoting an upper limit. As noted above, "always" could be taken to mean always in analyses that are carried out as a search, not for a measurement of a well-established signal process.

4.3 Properties of Feldman-Cousins

The interval is based on a likelihood-ratio test with respect to a two-sided alternative.

The Feldman-Cousins intervals are never zero length. If the data value x is observed in the negative range, then the upper edge of the interval stays positive, and approaches zero when x becomes increasingly negative. In such cases, the values of μ below the limit are not excluded because the critical region for their test also includes data outcomes where x fluctuates high relative to the rate predicted by μ.

If the upper-edge of the F-C interval is used as an upper limit, then its coverage probability exceeds everywhere $1 - \alpha/2$ (i.e., 97.5%), approaching this value when μ becomes large.

The power of the test with respect to a one-sided alternative is not as high as that used for PCL.

There is no flip-flopping problem to the extent that one always quotes the Feldman-Cousins interval.

The interval can effectively be used to establish discovery at the 5σ level by choosing $\alpha = 2.9 \times 10^{-7}$.

References

[1] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, *Power-Constrained Limits*, arXiv:1105.3166.

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

[3] Gary J. Feldman and Robert D. Cousins, Phys. Rev. D 57 (1998) 3873–3889.

[4] Robert D. Cousins, Draft note on Negatively Biased Relevant Subsets and references therein.