DRAFT 1.3

Glen Cowan, Kyle Cranmer,
Eilam Gross, Ofer Vitells
13 April, 2011

Power Constrained Limits

Abstract

We propose a method for setting limits that avoids excluding parameter values for which the sensitivity falls below a specified threshold. These "power-constrained" limits (PCL) address the issue that motivated the widely used CL_s procedure [1], but do so in a way that makes more transparent the properties of the statistical test to which each value of the parameter is subjected. A case of particular interest is for upper limits on parameters that are proportional to the cross section of a process whose existence is not yet established. The basic idea of the power constraint can easily be applied, however, to other types of limits.

1 Introduction

In particle physics experiments one often tests specific models that predict new phenomena. Some regions of a model's parameter space may be rejected by these tests; in other regions the model may be deemed compatible with the data. This is often done in the framework of a frequentist statistical test, which is inverted to determine a confidence interval. This formalism is reviewed in Sec. 2.

It often happens that for some regions in a model's parameter space, the magnitude of the predicted effect with respect to the background-only model is extremely small. That is, one has effectively no experimental sensitivity to those parts of the model's parameter space. Nevertheless, procedures based on frequentist tests may exclude these values. We discuss how this can occur and how it has been dealt with in the past in Sections 3 and 4.

In Sec. 5 we introduce a new method for constraining confidence intervals in a way that prevents one from excluding parameter values to which one does not have sufficient sensitivity. As the measure of sensitivity is based on the power of a statistical test, we refer to the bounds established by these modified intervals as power-constrained limits (PCL).

Section 6 illustrates the procedure for the case of an upper limit derived from a Gaussian measurement. Section 8 discusses how the procedure can be applied in cases where there are additional nuisance parameters, beyond the parameters of interest, that must be fitted using the data. A summary and conclusions are given in Sec. 9.

2 Confidence intervals from inverting a statistical test

In this section we review the formalism of inverting a frequentist statistical test to obtain a confidence interval. A more thorough treatment can be found in many texts, such as Ref. [2].

We consider a test for a parameter μ, which here represents the signal strength (or any parameter proportional to the rate) of a certain process. A test of a given μ is carried out by specifying a region of data outcomes called the *critical region*, which are disfavoured, in a sense discussed below, under assumption of μ. The data outcome could be, for example, the number of events observed in a given region of phase space, or it could represent a larger set of numerical values. Here we will use $\mathbf{x}$ to represent the data, and w_μ to denote the critical region.

The critical region is chosen to such that the probability to observe the data in it, under assumption of the hypothesized μ, is not greater than a given constant α, called the *size* or *significance level* of the test, i.e.,

$$P(\mathbf{x} \in w_\mu | \mu) \leq \alpha \; . \tag{1}$$

Often by convention $\alpha = 0.05$ is used. If the data are observed in the critical region, the hypothesis μ is rejected. It is necessary in general to specify Eq. (1) as an inequality because the data may be discrete (e.g., an integer number of events), and so there may not exist a subset of the possible data values for which the summed probability is exactly equal to α.

It is convenient to construct from the data a test statistic q_μ, such that greater q_μ reflects an increasing level of incompatibility between the data and the hypothesized parameter value μ. In this way the boundary of the critical region in data space is given by a surface of constant q_μ, with the critical region containing the data that give the greatest values of q_μ. Once such a function has been defined, one can for any observed value $q_{\mu,\mathrm{obs}}$ compute a p-value, i.e., the probability under assumption of μ to find data with equal or greater incompatibility with μ,

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu) \, dq_\mu \; , \tag{2}$$

where $f(q_\mu|\mu)$ represents the probability density function (pdf) of q_μ assuming a data distribution with strength parameter μ. Thus the test can be equivalently formulated by rejecting μ if its p-value is found less than α.

A test of size α can be carried out for all values of μ. The set of values not rejected constitutes a *confidence interval* for μ with confidence level $1 - \alpha$. This interval will by construction include the true value of the parameter with a probability of at least $1 - \alpha$.

The procedure described above for constructing a confidence interval by inverting a test is not unique, however, because there are (often infinitely) many different subsets of the data space that could be chosen for the test's critical region w_μ. This is usually selected such that the probability to find $\mathbf{x} \in w_\mu$ is large if a given alternative hypothesis (or set of alternatives) is true. The *power* of the test with respect to an alternative value of the parameter μ', which we denote here as $M_{\mu'}(\mu)$, is

$$M_{\mu'}(\mu) = P(\mathbf{x} \in w_\mu | \mu') \; . \tag{3}$$

If the test of μ is formulated using a p-value, such that finding $p_\mu < \alpha$ is equivalent to finding $\mathbf{x} \in w_\mu$, then the power can be written equivalently as

$$M_{\mu'}(\mu) = P(p_\mu < \alpha | \mu') \; . \tag{4}$$

Often the power with respect to certain alternatives is used as the criterion according to which one chooses the critical region of a test. Confidence intervals obtained from inverting the test thus depend on this choice. For the present discussion, however, we will assume that the test has been defined, and the power will be used only to modify the resulting confidence interval so that it does not exclude parameter values to which one does not have sufficient sensitivity. This concept is defined more quantitatively in the following section.

3 Spurious exclusion

When testing a hypothesized strength parameter μ, it may be that the magnitude of the signal implied by μ is extremely small — so small, that the probabilities for the data are very close to what they would be in the absence of the signal process, i.e., $\mu = 0$. In such a case one has little or no sensitivity to the given value of μ.

For example, Fig. 1 illustrates a situation where there is only a very small level of sensitivity to a given strength parameter μ. The plot shows the pdfs of the statistic q_μ under assumption of strength parameters μ, and also assuming $\mu = 0$, i.e., $f(q_\mu|\mu)$ and $f(q_\mu|0)$. If the observed value of the statistic is found in the critical region corresponding to the top 5% of $f(q_\mu|\mu)$, then the hypothesized μ is rejected. But as the two pdfs almost coincide, the probability to reject μ if the true strength parameter is zero is also close to $\alpha = 0.05$.

(a) (b)

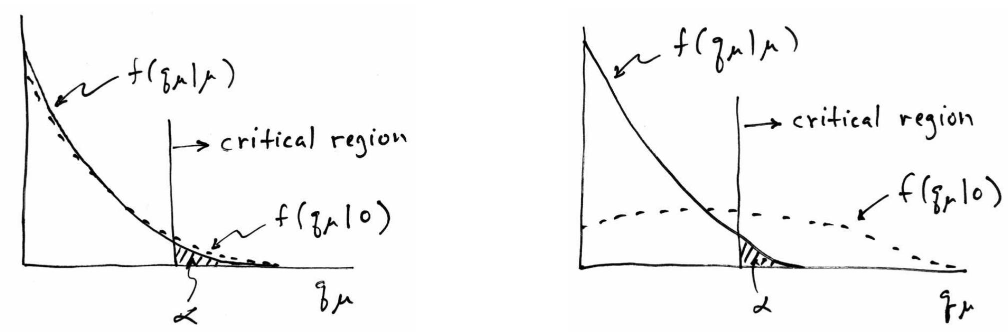

Figure 1: Illustration of statistical tests of parameter values μ for the cases of (a) little sensitivity and (b) substantial sensitivity (see text).

Figure 1(b) shows the same distributions as (a) but for a different value of μ. The size of the test is, as in (a), equal to α. Here, however, the distribution of q_μ under the assumption of $\mu' = 0$ leads to a substantially greater probability to reject μ, i.e., to find q_μ in the critical region.

The sensitivity of a test of μ can be quantified using the power of the test with respect to a stated alternative μ', which we will take here to be the no-signal hypothesis, $\mu' = 0$. In the case where the pdfs $f(q_\mu|\mu)$ and $f(q_\mu|0)$ coincide, the probability to reject μ assuming the alternative $\mu' = 0$ approaches the significance level of the test, α.

In the context of a search for a new phenomenon, this means that with probability not less than α one will exclude hypotheses to which one has little or no sensitivity, which we refer to here as spurious exclusion. The hypothesis might indeed be false, but if it is excluded, this is more naturally interpreted as a data fluctuation away from the region favoured under assumption of μ. This could result, for example, in a search for a hypothetical particle with a mass far above the range where it would have a noticeable impact on the data. Particle Physics experiments often carry out many searches covering a broad parameter range for many signal models, and so spurious exclusion is in fact a problem that can arise often.

4 Previous methods that address spurious exclusion

The problem of spurious exclusion, or equivalently, having a "lucky" statistical fluctuation lead to an anomalously strong limit, has been known in the particle physics community for many years. The note by Highland [3] reviews the problem and proposes several possible solutions; further discussion can be found in the review by the Particle Data Group [4].

The problem received particular focus during searches for the Higgs Boson at the LEP Collider in the 1990s, and led to a procedure called "$\mathrm{CL_s}$" [1]. Here one forms the ratio

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_0} \,, \tag{5}$$

where p_μ and p_0 are the p-values of the hypothesized strength parameter values μ and 0, respectively. In the $\mathrm{CL_s}$ procedure, μ is deemed to be excluded if one finds $\mathrm{CL}_s < \alpha$. Because CL_s is aways greater than p_μ, the probability of exclusion assuming μ is necessarily less than α. Thus the quoted upper limit from the $\mathrm{CL_s}$ procedure will be greater than the upper limit according to the method of Sec. 2, and in this sense the $\mathrm{CL_s}$ procedure is said to be conservative. This is illustrated in the example described in Sec. 6.

Because of this conservatism, the frequentist coverage probability of the $\mathrm{CL_s}$ upper limits (i.e., the probability under assumption of μ that the interval will contain μ) is not equal to α, but is in general larger. Although the exact coverage probabilities of $\mathrm{CL_s}$ intervals can be found as a function of μ, this requires additional computational effort and is not often done in practice.

5 Power Constrained Limits

Here we propose an alternate procedure for producing intervals whose coverage properties are easily apparent for all values of μ. To do this we break the range of μ to be tested into two categories based on the power $M_0(\mu)$ of a test of μ with respect to the no-signal alternative, $\mu' = 0$. If this power is below a specified threshold $M_{\min}$, one's sensitivity to this parameter is deemed to be too low and the point is not regarded as testable. If the power is greater than or equal to the threshold, then the test of size α is carried out. A value of μ is excluded if

(a) one has sufficient sensitivity to μ, i.e., $M_0(\mu) \geq M_{\min}$, and

(b) the value μ is rejected by the test, i.e., $\mathbf{x} \in w_\mu$ or equivalently $p_\mu < \alpha$.

An interval is constructed from the values of μ not excluded. The coverage probability of the interval is 100% for μ values that have power below $M_{\min}$, and α for those values with power greater than or equal to the threshold. When reporting the result it is recommended to indicate which parameter values were above and which below the power-constraint threshold, and in this way one can easily see what the coverage probability is for all values of μ.

The choice of the minimum power threshold is a matter of convention. We prefer to use $M_{\min} = 0.16$, or more precisely, $M_{\min} = \Phi(-1) = 0.1587$, where Φ is the standard normal cumulative distribution (i.e., the cumulative distribution for Gaussian with a mean of zero and unit standard deviation). As shown below, this corresponds to applying the power constraint if the data fluctuate one standard deviation below the expected background.

This procedure bears some similarity to one introduced recently in the astrophysics community in Ref. [5], although there the power refers to a test of the background-only ($\mu = 0$) hypothesis, and furthermore the result is not used in quite the same way as what we propose here. Note also in Ref. [5], "upper bound" is similar to what we call an upper limit, and their term "upper limit" is taken to refer to the sensitivity threshold.

Formally, to construct the interval for μ one begins by finding the power for a test of each μ with respect to the alternative $\mu' = 0$,

$$M_0(\mu) = P(\mathbf{x} \in w_\mu | 0) = P(p_\mu < \alpha | 0) \ . \tag{6}$$

In some problems this can be found in closed form; otherwise it can be obtained using a Monte Carlo calculation, in which one for every value of μ calculates the distribution of p_μ using data generated according to $\mu = 0$. The value $M_0(\mu)$ is then found simply by integrating each distribution from zero up to the desired significance level α (e.g., 0.05).

An equivalent and in ways simpler procedure is first to carry out the statistical test without the power constraint, and invert this to find the unconstrained confidence interval for μ. Some of the parameter values that are excluded from this interval may be found to have a power below the required threshold, and they are then re-included in the power-constrained interval, which is thus by construction larger than the unconstrained one.

For example, one may be interested in finding an an upper limit, μ_{up}, i.e., the largest value of μ not excluded. By inverting the test, one determines μ_{up} as a function of the data. One can therefore determine the distribution of μ_{up}, e.g., by simulating the experiment many times under assumption of $\mu = 0$ and constructing a histogram of μ_{up} for each outcome. Then for each value of μ one determines the corresponding power. This is the probability, under assumption of the background-only ($\mu = 0$) hypothesis, to reject μ, i.e., to find μ outside of the unconstrained confidence interval. In the case of an upper limit this is

$$M_0(\mu) = P(\mu_{\mathrm{up}} < \mu | 0) \ . \tag{7}$$

One should note the following caveat: It can be that for certain data outcomes, all values of μ are excluded by the test, in which case μ_{up} is not defined. In such cases one must count the outcomes as contributing to the probability that μ is outside the confidence interval.

With this in mind, one can then find the smallest value of μ for which the power $M_0(\mu)$ is at least equal to the minimum value M_{min}, denoted here as μ_{min}. The Power-Constrained Limit μ^*_{up} is given by the larger of the unconstrained limit μ_{up} or the minimum value to which one has sensitivity, μ_{min}:

$$\mu^*_{\mathrm{up}} = \max(\mu_{\mathrm{up}}, \mu_{\mathrm{min}}) \ . \tag{8}$$

6 PCL for an upper limit based on a Gaussian measurement

Often the test of μ is based on a Gaussian distributed measurement. For example, for a sufficiently large data sample and under conditions often satisfied in practice, the distribution of the Maximum Likelihood Estimator $\hat{\mu}$ has a Gaussian form with standard deviation σ and is centred about the true μ. Here we will assume this is the case and further take σ to be known.

For the case of an upper limit, we define the critical region to contain the lowest values of $\hat{\mu}$ such that the probability to find $\hat{\mu}$ there is equal to α. For Gaussian distributed $\hat{\mu}$ with mean μ and standard deviation σ, one defines the critical region therefore to be

$$\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha) \;, \tag{9}$$

where Φ^{-1} is the inverse of the standard Gaussian cumulative distribution (the standard normal quantile). For example, $\alpha = 0.05$ gives $\Phi^{-1}(1-\alpha) = 1.64$.

Rejecting μ if the data are in the critical region gives the unconstrained upper limit,

$$\mu_{\rm up} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha) \;. \tag{10}$$

The power of the test of μ with respect to the alternative $\mu' = 0$ is

$$M_0(\mu) = P\left(\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha) | 0\right) \;. \tag{11}$$

Because $\hat{\mu}$ here follows a Gaussian distribution, the power can be written

$$M_0(\mu) = \Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right) \;. \tag{12}$$

This is illustrated in Fig. 2 for $\alpha = 0.05$ and $\sigma = 1$. Since the cumulative distribution Φ is monotonically increasing and furthermore $\Phi(1-\alpha) = -\Phi(\alpha)$, Eq. (12) gives $M_0(0) = \alpha$ and $M_0(\mu) > \alpha$ for all $\mu > 0$, as can be seen in the figure.

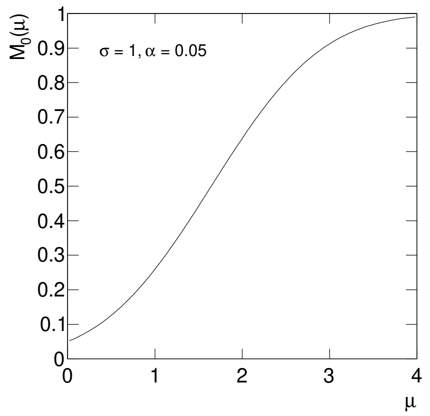

Figure 2: The power function $M_0(\mu)$ for a test of μ with respect to the alternative $\mu' = 0$ (see text).

Requiring the power $M_0(\mu) \geq M_{\rm min}$,

$$\Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right) \geq M_{\rm min} \;, \tag{13}$$

implies that the smallest μ to which one is sensitive is

$$\mu_{\rm min} = \sigma\left(\Phi^{-1}(M_{\rm min}) + \Phi^{-1}(1-\alpha)\right) \;. \tag{14}$$

By combining Eqs. (10) and (14), one sees that $\mu_{\rm up}$ is below $\mu_{\rm min}$ if one finds

$$\hat{\mu} < \sigma\Phi^{-1}(M_{\rm min}) \;. \tag{15}$$

Thus one finds the following expression for the power-constrained upper limit:

$$\mu_{\rm up}^* = \begin{cases} \sigma\left(\Phi^{-1}(M_{\rm min}) + \Phi^{-1}(1-\alpha)\right) & \hat{\mu} < \sigma\Phi^{-1}(M_{\rm min})\ , \\ \hat{\mu} + \sigma\Phi^{-1}(1-\alpha) & \text{otherwise}\ . \end{cases} \tag{16}$$

This is shown as a function of $\hat{\mu}$ in Fig. 3(a).

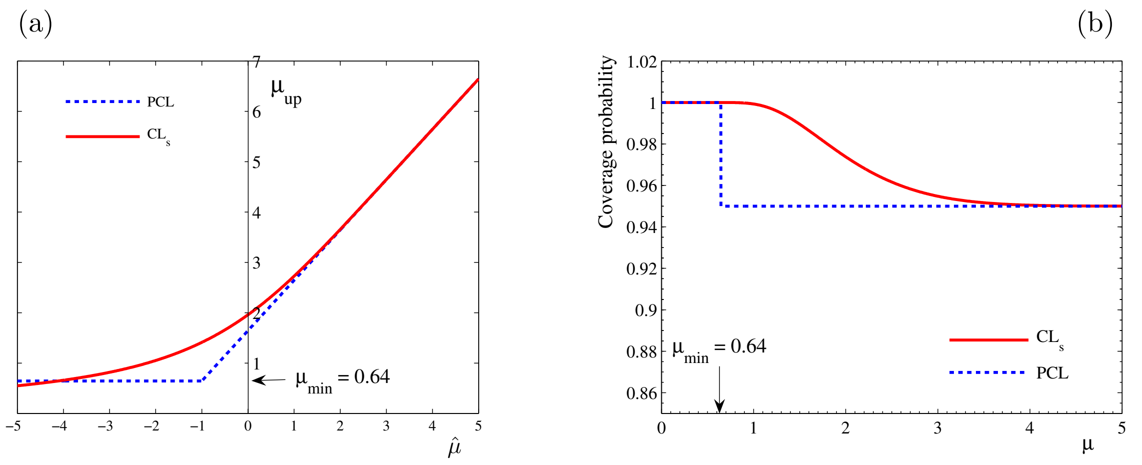

Figure 3: (a) Upper limits from the PCL (dashed) and $\mathrm{CL_s}$ (solid) procedures as a function of $\hat{\mu}$), which is assumed to follow a Gaussian distribution with unit standard deviation. (b) The coverage probabilities from the PCL (dashed) and $\mathrm{CL_s}$ (solid) procedures as a function of μ.

For comparison, Fig. 3(a) also shows the upper limit obtained from the $\mathrm{CL_s}$ procedure. Figure 3(b) shows the coverage probability of the upper limits from PCL and $\mathrm{CL_s}$. For PCL, this is 100% for $\mu < \mu_{\rm min} = \sigma(\Phi^{-1}(M_{\rm min}) + \Phi^{-1}(1-\alpha)) = 0.64$, and 95% otherwise. For $\mathrm{CL_s}$, the coverage probability is everywhere greater than 95%, approaching 95% as μ increases.

7 Distribution of upper limit and choice of minimum power

As mentioned above, we prefer to take the minimum power threshold $M_{\rm min} = \Phi(-1) = 0.1587$. From Eq. (15) one can see that if $\mu_{\rm up}$ follows a Gaussian distribution, this choice of $M_{\rm min}$ corresponds to applying the power constraint if the data fluctuate below their expected value, under assumption of $\mu = 0$, by more than one standard deviation. Here we will refer to a fluctuation at this level as 1σ (downward), regardless of the distribution of $\mu_{\rm up}$. In fact, the distribution of $\mu_{\rm up}$ often is close to Gaussian so the terminology is natural and convenient.

This choice of $M_{\rm min}$ can be motivated by the idea that a sufficiently small fluctuation should not result in spurious exclusion of the type that the PCL and $\mathrm{CL_s}$ procedures are intended to prevent. If, for example, one were to require $M_{\rm min} = 0.5$, then one would impose the power constraint whenever the observed limit is found below the median, i.e., half of the time, which is not consistent with the notion of accepting small fluctuations. Therefore we feel requiring a power of 50% is too extreme.

On the other hand, for any (unbiased) test, the power is always greater than or equal to the significance level α. So if one were to take $M_{\rm min} \leq \alpha$ then the result is the same as the unconstrained limit. Since one often takes $\alpha = 0.05$, taking $M_{\rm min} = 0.05$ would correspond to a 1.64σ downward fluctuation (i.e., $\Phi(-1.64) = 0.05$). We therefore believe $M_{\rm min} = \Phi(-1) \approx 0.16$ is a natural choice, as it allows for fluctuations up to the one-sigma level before imposing the power constraint.

In many searches for new phenomena, one may carry out the analysis for a range of parameters in the signal model. For example, when searching for the Higgs boson one may carry out the analysis for each value of the mass m_{H}. In this situation one can simply repeat the power-constraint procedure for each value of the signal model's parameters, as is illustrated in Fig. 4.

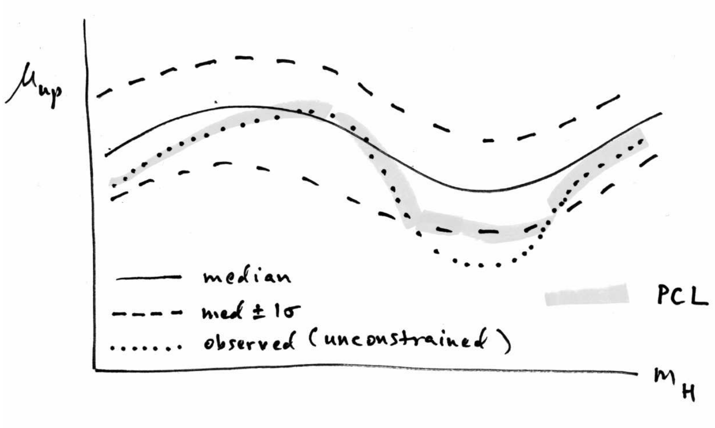

Figure 4: Illustration of the power-constrained limit as a function of a model parameter such as the Higgs boson mass m_{H} (see text).

In Fig. 4, the solid line represents the median value of the unconstrained upper limit μ_{up}, and the lower and upper dashed curves are the 0.16 and 0.84 quantiles of the distribution of μ_{up}. The dotted curve in Fig. 4 represents a possible outcome for the unconstrained limit μ_{up}. The minimum power is taken to be $M_{\mathrm{min}} = \Phi(-1) = 0.16$, and thus the power-constrained limit is the greater of the dotted and lower dashed curves, as indicated by the shaded curve.

8 Treatment of nuisance parameters

In many analyses, the probability model that describes the data is not uniquely specified by the parameter (or parameters) of interest, but rather also contains nuisance parameters. That is, the values of these parameters are not known a priori and they must be fitted using the data. For concreteness suppose the model is characterized by a strength parameter μ and a set of nuisance parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$.

The nuisance parameters complicate the present problem in two ways. First, they make it difficult to construct an unconstrained interval for the parameter of interest that has the correct coverage probability for all values of $\boldsymbol{\theta}$. This problem has been widely discussed in recent years, e.g., Ref. [6]. Many of the proposed procedures give intervals with correct coverage for some values of $\boldsymbol{\theta}$, but approximate coverage elsewhere. For example, an approximate solution based on the profile likelihood ratio test is discussed in Refs. [7]. For the present discussion we will assume that a test procedure that gives an unconstrained interval has been chosen. Its coverage probability may or may not be exactly equal to the nominal confidence level for all values of $\boldsymbol{\theta}$.

Of more direct concern for the present paper is the fact that the power of the test of μ with respect to the no-signal alternative will depend in general on the nuisance parameters $\boldsymbol{\theta}$. As the power is intended to represent the probability, under assumption of the no-signal model, to reject a given value of μ, we take the values of $\boldsymbol{\theta}$ that are in best agreement with the actual data under assumption of $\mu = 0$. We denote these as $\hat{\hat{\boldsymbol{\theta}}}(0)$, i.e., they are the conditional estimators for $\boldsymbol{\theta}$ under assumption of $\mu = 0$.

As a consequence of this choice, the power $M_0(\mu)$ becomes a function of the actual data, since the data are used to determine values for the nuisance parameters. Thus the range of μ

values where one has sufficient sensitivity also depends to some extent on the data. This may seem counter-intuitive, since the power of a specific test, i.e., at a given point in $(\mu, \boldsymbol{\theta})$-space, is independent of the data. But there is a certain power $M_0(\mu)$ for every point in $\boldsymbol{\theta}$-space, and one uses the data to choose the point at which one quotes the power.

Alternatively, one could require that the power is greater than or equal to the minimum threshold for all values of the nuisance parameters in a specified range. In this way the set of μ values for which one has sufficient sensitivity would not depend on the data. As this would entail considerable computational effort, however, we prefer to define the power using a specific point in the nuisance-parameter space as described above.

9 Summary and conclusions

We propose a power-constraint procedure for modifying confidence limits so that parameter values to which one has little or no sensitivity are not excluded. The sensitivity is measured using the power of the test of the parameter with respect to the no-signal alternative. The coverage probability of the resulting limits is equal to the nominal confidence level (e.g., 95%) for parameter values to which one's sensitivity is above a given threshold, and 100% if the sensitivity is below the threshold. This can be contrasted with the $\mathrm{CL_s}$ procedure, for which the coverage probability is always greater than the nominal confidence level by an amount that varies continuously as a function of the assumed parameter value.

The power used for the sensitivity threshold is a matter of convention, but recommend taking this to be $M_{\min} = \Phi(-1) \approx 0.16$. This is consistent with allowing for reasonably small downward fluctuations of the data by drawing the boundary at the one-sigma level. Allowing more than 1.64σ fluctuations would mean the power constraint is never imposed (for a 95% confidence level limit), and requiring $M_{\min} = 0.5$ would impose the power constraint half of the time, including cases with only an infinitesimal downward fluctuation.

The PCL procedure is easily extended to problems with nuisance parameters. There we define the power with respect to the background-only ($\mu = 0$) model using the conditional estimates of the nuisance parameters given $\mu = 0$.

The PCL procedure is particularly useful in cases where spurious exclusion is problematic, such as when a one-sided test is inverted to give an upper limit. It can be applied, however, to any confidence interval, including those based on inversion of a likelihood-ratio test (i.e., Feldman-Cousins intervals [8]).

When reporting results, we recommend to show both the constrained and unconstrained limits. In this way one can know whether a given parameter value is not rejected because the data are in good agreement with it, or rather because it is a value to which the sensitivity is deemed to low to allow exclusion.

Acknowledgements

We thank...

References

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

[2] A. Stuart, J.K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2A: *Classical Inference and the Linear Model* 6th Ed., Oxford Univ. Press (1999), and earlier editions by Kendall and Stuart.

[3] Virgil L. Highland, *Estimation of Upper Limits from Experimental Data*, July 1986, Revised February 1987, Temple University Report C00-3539-38.

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

[5] Vinay L. Kashyap et al., *On Computing Upper Limits to Source Intensities*, Astrophysical Journal, 719, 900-914 (2010); arXiv:1006.4334.

[6] Louis Lyons and Müge Karagöz Ünel (eds.), *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, Proceedings of PHYSTAT05, Imperial College Press (2006); see, e.g., the section *Nuisance Parameters/Limits/Discovery.*

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

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