ATLAS statistics forum
Draft 1.1, Feb 7th, 2011

Frequentist Limit Recommendation

1 Introduction

We summarize the recommended procedure needed for computing frequentist exclusion limits based on profile likelihood ratio tests. We consider testing a hypothesized signal strength μ, defined such that $\mu = 0$ is the background-only model, and $\mu = 1$ corresponds to the nominal signal model. The result of the significance test is a p-value, p_μ. If one finds $p_\mu < 0.05$, then this value of μ is excluded at 95% confidence level. The upper limit on μ is the highest value of μ not excluded, in practice found by solving $p_\mu = 0.05$ for μ. The recommended limit procedure is based on toy Monte Carlos which can be supplemented and partially validated with simple and fast asymptotic formulas. The asymptotic formulas and the definitions are extracted from Ref. [1], where more details can be found. Even though the recommendation given here is based on power constrained limits (PCL) we also recommend to derive the CL_s limit [2] which is based on a ratio of p-values, to allow comparisons with the TEVATRON limits.

2 The Recipe

We hereby give a recipe 'in a nut shell' to find the observed and expected limits . Note that to derive the observed limits, one uses the observed data, while to derive expected limits, one generates simulated data sets via toy Monte Carlos.

1. Construct the likelihood function $\mathcal{L}(\mu, \boldsymbol{\theta})$ where μ is the signal strength and $\boldsymbol{\theta}$ represent the nuisance parameters. An example of a likelihood function is given in section 5.
2. Construct the test statistic $\tilde{q}_\mu$ based on the $\tilde{\lambda}(\mu)$ likelihood ratio:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ \frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \end{cases} \tag{1}$$

Here $\hat{\hat{\boldsymbol{\theta}}}(0)$ and $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ refer to the conditional ML estimators of $\boldsymbol{\theta}$ given a strength parameter of 0 or μ, respectively.

The test statistic $\tilde{q}_\mu$ is given by

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\ , \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu\ , \\ 0 & \hat{\mu} > \mu\ . \end{cases} \tag{2}$$

3. Find the observed test statistic for the tested μ, $\tilde{q}_{\mu,obs}$.

4. Generate toy Monte Carlo experiments to construct the pdf of $\tilde{q}_\mu$ under signal (with strength μ) and background only experiments $f(\tilde{q}_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs}))$ and $f(\tilde{q}_\mu|0, \hat{\hat{\boldsymbol{\theta}}}(0, \text{obs}))$. Here the $\hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs})$ and $\hat{\hat{\boldsymbol{\theta}}}(0, \text{obs})$ are the conditional MLEs based on the *observed data*. Also note, that the nuisance parameters are fixed to their conditional MLEs for generating the toy Monte Carlo, but are allowed to float in fits needed to evaluate the test statistic. In the asymptotic limit the distribution $f(\tilde{q}_\mu|\mu, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$.

   Important: Conditional vs. Unconditional Ensembles –
   When generating toy Monte Carlo pseudo-experiments, there is a subtle issue associated with constraint terms on nuisance parameters that come from auxiliary measurements. These often include constraints on luminosity, identification efficiencies, jet energy scale, etc. that are not constrained by the main measurement. Consider a constraint on the luminosity $G(L_0|L, \Delta_L)$, where L is the *nuisance parameter* associated with the true, unknown luminosity and L_0 is an *auxiliary measurement* that provides the nominal value of the luminosity. The nuisance parameter L is one of the components of $\boldsymbol{\theta}$, and it has a particular value $\hat{\hat{L}}(\mu, \text{obs})$ inside $\hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs})$. The issue here is whether or not one randomizes the *auxiliary measurement* L_0 about $\hat{\hat{L}}(\mu, \text{obs})$ when generating the pseudo-experiments. In the "unconditional ensemble" L_0 is randomized according to $G(L_0|\hat{\hat{L}}(\mu, \text{obs}), \Delta_L)$. In the "conditional ensemble", one would keep a fixed value of the auxiliary measurement L_0 for each pseudo-experiment (although the nuisance parameter L would still float in each fit). Both are valid frequentist constructions, but the asymptotic results correspond to the unconditional ensemble, so that is the recommendation. There is an explicit example in Sect. 5.

5. From the constructed distribution of $\tilde{q}_\mu$ for the signal+background, $f(\tilde{q}_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs}))$, find the p-value of the observation

$$p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu, \text{obs}))\, d\tilde{q}_\mu \tag{3}$$

6. Find (by iteration or any other way) μ_{up} which satisfies $p_{\mu_{up}} = 5\%$.

7. To find the median sensitivity, generate background-only toy MC experiments, for each one of them, find μ_{up} (steps 5 and 6). Draw the μ_{up} distribution and find its median.

8. To find the ± 1 and ± 2 σ bands (green and yellow), use the above generated μ_{up} pdf and derive the 68% and 95% bands.

9. The 2σ band allows to exclude signals with very low cross sections, such that the experiment is not sensitive to. To protect against this one could use the CL_s technique (see next item) or construct the power constrained limit (PCL). We recommend to use a power of 16% and not allow the observed limit to go below the -1σ expected limit. This means practically that if the observed limit goes below the -1σ band, the quoted limit is $\mu_{up} - 1\sigma$.

   In terms of the green/and/yellow plot the outer band would stretch down to zero, and it should be dropped, colored white. The observed data should not be hidden - if we have a downward fluctuation, show it. But when it passes the power constraint, mark the power constraint solid and the downward fluctuation dotted so that the result we are using is clear.

10. Calculate the CL_s upper limit. To do this calculate $1 - p_b$ where

$$p_b = 1 - \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\hat{\boldsymbol{\theta}}}(0, obs)) d\tilde{q}_\mu \tag{4}$$

Define p'_μ as a ratio of p-values,

$$p'_\mu = \frac{p_\mu}{1 - p_b} \tag{5}$$

Follow steps 6-9 by solving for $p'_{\mu_{up}} = 5\%$ (replacing p_μ by p'_μ).

11. Use the approximate formulas in section 3 and/or section 4 and compare with your derivations. If you find big differences, make sure you understand them.

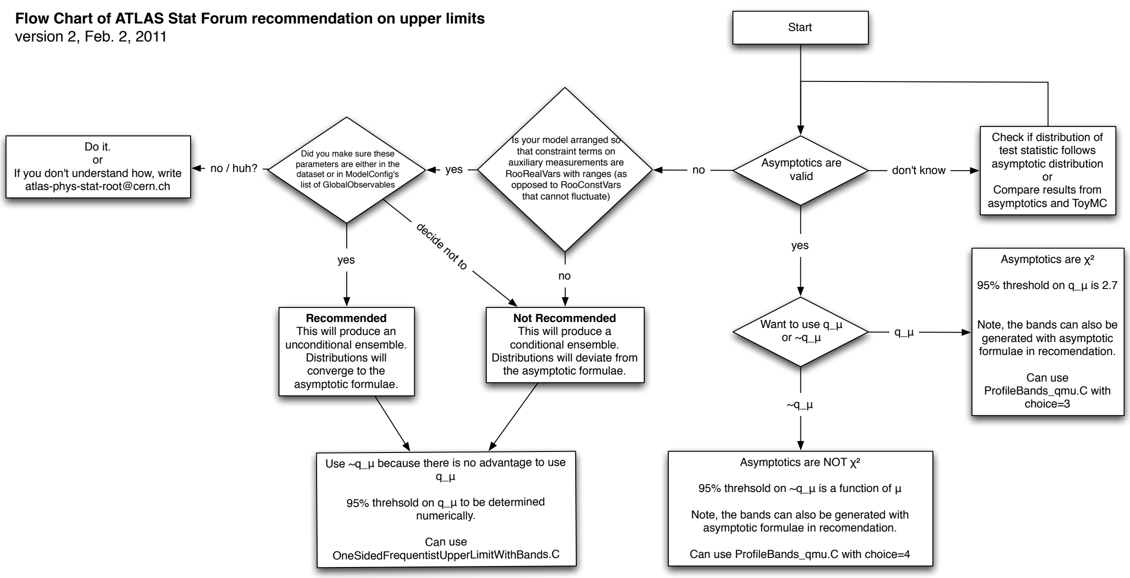

Figure 1: A flow chart for the recommendations..

3 Asymptotic Formulas for q_μ

There are two possible definitions here for the Profile Likelihood. q_μ and $\tilde{q_\mu}$. They both allow $\hat{\mu} < 0$ but $\tilde{q}_\mu$ treats it in a special manner (see Equation 2). Both are asymptotically equivalent as shown in [1]. The simplest way is to define the profile likelihood is by

$$\begin{aligned} q_\mu &= -2\ln\frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}; \quad \hat{\mu} < \mu \\ q_\mu &= 0; \quad \hat{\mu} > \mu \end{aligned} \tag{6}$$

It is recommended to verify that q_μ is distributed like $f(q_\mu|\mu) \sim \chi^2_1$ (Wilks theorem), this will usually be the case, in particular when combining channels. The approximation works best when Wilks theorem is satisfied.

Following Wilks theorem, the upper limit on the signal strength μ_{95} is given by solving for $q_{\mu_{95}} = 1.64^2$.

$$p_{\mu_{95}} = 1 - \Phi(\sqrt{q_\mu}) = 1 - \Phi(1.64) = 0.05 \tag{7}$$

If $\mu_{95} < 1$ the point is excluded (the 95% Confidence Interval does not contain $\mu = 1$).

3.1 Expected Limit and Error Bands "(CL_{s+b})"

To find the expected limit, one should plug in the Asimov data which is the expected background (with no fluctuations). The signal strength is set to zero. One then gets $q_{\mu,A}$ and the corresponding μ_{up}^{med} is given by solving $q_{\mu_{up}^{med},A} = 1.64^2$ The error bands are given by

$$\mu_{up+N} = \sigma(\Phi^{-1}(1-\alpha) + N) \tag{8}$$

with

$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}} \tag{9}$$

$\alpha = 0.05$, μ can be taken as μ_{up}^{med} in the calculation of σ.

3.2 Expected Limit and Error Bands "(CL_s)"

To avoid setting limits when the experiment is not sensitive to the signal, one might use the modified p-value defined above, "p'_s"

$$p'_s = \frac{p_s}{1 - p_b} \tag{10}$$

We find

$$p'_\mu = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})} \tag{11}$$

The median and expected error bands will therefore be

$$\mu_{up+N} = \sigma(\Phi^{-1}(1 - \alpha\Phi(N)) + N) \tag{12}$$

with

$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}} \tag{13}$$

$\alpha = 0.05$,μ can be taken as μ_{up}^{med} in the calculation of σ..

Note that for $N = 0$ we find the median limit

$$\mu_{up}^{med} = \sigma\Phi^{-1}(1 - 0.5\alpha) \tag{14}$$

The expected μ and the expectation for error band N is shown in Figure 1. one can clearly see the condensation when $N \to -\infty$

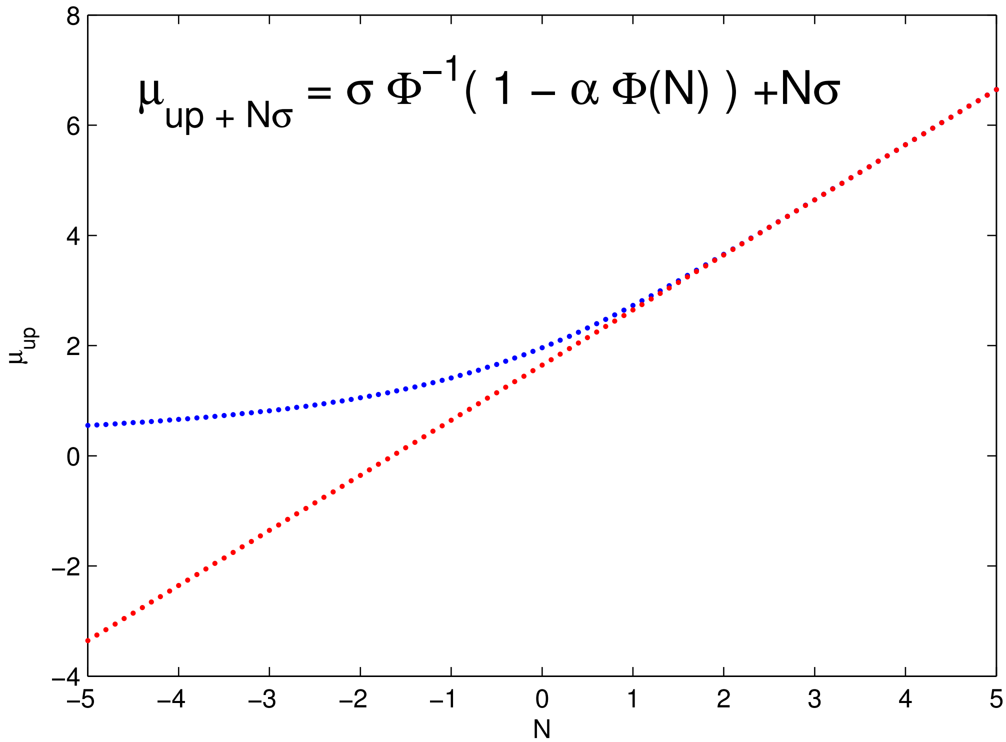

Figure 2: $\mu_{up+N\sigma}$ as a function of N (in units of σ). Red is based on p_s blue is based on p'_s (CL_s).

4 Asymptotic Formulas for $\tilde{q}_\mu$

Large values of $\tilde{q}_\mu$ (Equation 2) corresponding to increasing disagreement between the data and the hypothesized μ. For a sufficiently large data sample, the pdf $f(\tilde{q}_\mu|\mu)$ is found to approach

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2}\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}e^{-\tilde{q}_\mu/2} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \ , \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} \exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu+\mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2 \ . \end{cases} \tag{15}$$

This is the asymptotic formula that should be used for $\tilde{q}_\mu$ (and not a chi squared).

Equation (15) requires the standard deviation σ of $\hat{\mu}$, under assumption of a signal strength μ. This can be found by the Asimov data set

$$\sigma \sim \frac{\mu}{\sqrt{\tilde{q}_{\mu,A}}} \tag{16}$$

or more accurately with the covariance matrix. To this end we estimate the covariance matrix from the matrix of second derivatives of the log-likelihood function, evaluated with the Asimov data set that corresponds to the strength parameter μ that is being tested.

We denote the likelihood evaluated with the Asimov data values as $L_{\mathrm{A}}(\mu)$.

For the inverse covariance matrix one finds (see [1] Eq. (28)),

$$V_{jk}^{-1} = -\frac{\partial^2 \ln L_{\mathrm{A}}}{\partial\theta_j\partial\theta_k} \tag{17}$$

In Eq. (17) the parameter μ is regarded as one of the θ_i (say, θ_0). To find σ, evaluate the derivatives of $\ln L_{\mathrm{A}}$ numerically, use this to find the inverse covariance matrix, and then invert and extract the variance of $\hat{\mu}$. One can see directly from Eq. (17) that this variance depends on the parameter values assumed for the Asimov data set, in particular on the assumed strength parameter μ.

The cumulative distribution for $\tilde{q}_\mu$ corresponding to the pdf (15) is

$$F(\tilde{q}_\mu|\mu) = \begin{cases} \Phi\left(\sqrt{\tilde{q}_\mu}\right) & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \ , \\ \Phi\left(\frac{\tilde{q}_\mu+\mu^2/\sigma^2}{2\mu/\sigma}\right) & \tilde{q}_\mu > \mu^2/\sigma^2 \ . \end{cases} \tag{18}$$

The p-value of the hypothesized μ is as before given by one minus the cumulative distribution,

$$p_\mu = 1 - F(\tilde{q}_\mu|\mu) \ , \tag{19}$$

and therefore the corresponding significance is

$$Z_\mu = \begin{cases} \sqrt{\tilde{q}_\mu} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \ , \\ \frac{\tilde{q}_\mu+\mu^2/\sigma^2}{2\mu/\sigma} & \tilde{q}_\mu > \mu^2/\sigma^2 \ . \end{cases} \tag{20}$$

The upper limit on μ at confidence level $1-\alpha$ is found by setting $p_\mu = \alpha$ and solving for μ, which gives

$$\mu_{\rm up} = \hat{\mu} + \sigma \Phi^{-1}(1-\alpha) \; . \tag{21}$$

Note that because σ depends in general on μ, Eq. (21) must be solved numerically.

The error bands are given by Equation 8.

5 An Example of a Likelihood Function

We treat two kinds of systematic errors. The common ones, which we assume (for simplicity) are fully correlated and the channel specific ones. The common correlated systematics include the Luminosity, the Jet Energy Scale (JES) and Acceptance. We perform a Profile Likelihood statistical analysis . Some measurements are data driven with a scale factor τ. We assume all systematics are Gaussian (as an example).

The likelihood (for one bin, or for the global counting analysis) is given by

$$\mathcal{L}(\mu, \beta^0_{j(j\in SB)}; \delta_{\epsilon^s}, \delta_{\beta_j}, \delta_i) = \mathrm{Pois}(n|\mu_T)\; N(m_{\delta_s}|\delta_{\epsilon^s}) \prod_{j\in SB} \mathrm{Pois}(n_j|\beta^0_j) \prod_j N(m_{\delta_{\beta_j}}|\delta_{\beta_j}) \prod_i N(m_{\delta_i}|\delta_i) \tag{22}$$

where j is an index over background processes, $j \in SB$ are background channels which are measured via Side Bands (or control regions), i is an index over systematic effects, μ_T is the total number of expected events given by

$$\begin{aligned} \mu_T \; &= \textstyle\sum_l \mu L \; \sigma_l (1 + \epsilon^s_l \delta_{\epsilon^s}) \prod_i (1 + \epsilon^s_{li}\delta_i) \\ &+ \textstyle\sum_j L \beta^0_j (1 + \epsilon^b_j \delta_{\beta_j}) \prod_i (1 + \epsilon^b_{ji}\delta_i), \end{aligned} \tag{23}$$

- n is the number of events in the signal region,
- $m_{\delta_s}, m_{\delta_{\beta_j}}, m_{\delta_i}$ represent auxiliary measurements of the corresponding δ systematic uncertainties. When generating toy Monte Carlo experiments, the m_δ should fluctuate around the value of δ in $\hat{\hat{\boldsymbol{\theta}}}(\mu, obs)$ or $\hat{\hat{\boldsymbol{\theta}}}(0, obs)$. If they are not randomized, this corresponds to a conditional ensemble in which the distribution of the test statistic departs significantly from the asymptotic distributions. $\boldsymbol{\theta}$ is the vector of nuisance parameters given below.
- $n_{j(j\in SB)}$ is the number of events measured in the control sample which is scaled by an extrapolation coefficient τ to estimate the number of events in the signal region. Since τ itself has uncertainty, we standardize it by writing $\tau_{j,j\in SB} = 1 + \epsilon_{\beta_j}\delta_{\beta_j}$,
- L is the nominal integrated luminosity,
- μ is the one parameter of interest, the signal strength,
- σ_l is the effective cross section (in pb) for signal events in channel l,
- ϵ^s_l is relative uncertainty on the efficiency of the channel l ,
- β^0_j is the nominal effective cross section (in pb) for background j,
- ϵ^b_j is the relative uncertainty on the effective cross section for background j,
- ϵ^s_{li} is the relative change in the effective cross-section due to the i^{th} systematic effect on signal channel l, and

- ϵ^b_{ji} is the relative change in the effective cross-section due to the i^{th} systematic effect on channel j.

The nuisance parameters are $\boldsymbol{\theta} = (\beta^0_{j(j \in SB)}; \delta_{\epsilon^s}, \delta_{\beta_j}, \delta_i)$ and the δ are constrained by the normal distribution $N(m_\delta|\delta) = G(m_\delta|\delta, 1)$.

References

[1] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, accepted by EPJC; arXiv:1007.1727.

[2] Read, Alexander L.,Presentation of search results: The CL(s) technique,J. Phys., G28, 2002, 2693-2704.