Multichannel number counting experiments

V. Zhukov*, M. Bonsch
KIT, Institute für Experimentelle Kernphysik, Universität Karlsruhe

Abstract
The confidence intervals calculated with different statistical methods for the combined model of number counting experiments have been compared. The Bayesian approach with flat prior provides the most conservative limit in a vast range of model parameters. The limits calculated with the Feldman-Cousins method are sensitive to the systematics. The results of Hybrid CLs calculations for the combined model can be the most optimistic but tend to undercover.

1 Introduction

The search for new phenomena in HEP if often reduced to the combination of many number counting experiments with different observed statistics, background expectations, signal sensitivity and systematic uncertainties [1]. Although such combination in one global fit is attractive from the physics perspective, it is challenging from the statistical point of view [2]. The combination of very different channels becomes sensitive to the ranges of parameter of interest, systematic uncertainties, correlations hence resulting in different predictions from different statistical methods.

In this study the upper confidence limits (95% CL) have been calculated with Profile Likelihood Calculator(PLC), Feldman-Cousins(FC) [9], Hybrid(using LR statistics and CLs ratio) [10] and Bayesian(with flat prior) methods. The modeling is done within RooStats package of Root v5.28 [3]. For Bayesian calculation the Markov Chain MC implemented in BAT has been used [4].

2 Statistical model

The combination of many(N_{ch}) exclusive channels can be written as a product of individual Likelihood functions $L = \prod_k^{Nch} L_k$ where each function may have two components: 1) the statistical term $Poiss(n_k|\mu_k)$ and 2) the systematics term $\prod_i^{Nnuis} G(\delta_i|\delta_0^i, \sigma_\delta^i)$ which is a set of pdf's for the vector of nuisances parameters $\vec{\delta}$ with mean $\vec{\delta}_0$ and $\vec{\sigma}_\delta$ affecting expectation value $\mu_k(\vec{\delta})$. There are usually a few common sources of systematic uncertainties for all channels which can be factorized as $(1 + f_k\delta_i)\mu$, where δ_i is the i-nuisance and f_k is the scaling factor for the k-channel depending on the amplitude of variations. For example the background part for k-channel can be written as: $B_k = b\prod_i^{Nnuis}(1 + f_k^i\delta_i)$ where b is the total background in this channel. Similarly the signal part is: $S_k = s\nu_k\prod_i^{Nnuis}(1+f_k^i\delta_i)$, here s is a common signal strength for all search channels and ν_k is the signal yield for k-channel. Such factorization of systematics allows to account for correlations in a simple way, otherwise the full covariance matrix has to be evaluated and the limit calculation becomes difficult for complicated models.

Although the shape of systematics pdf's can almost always be transformed into the standard distributions by replacement of variables it is common to keep natural observables in the statistical model. Then the distribution of systematics usually is a compromise between a simple analytical form ensuring good performance of statistical methods and the most realistic distribution of uncertainties describing the data which in turn depends on the nature of uncertainties and can be roughly divided onto two main categories [5].

First are the systematic uncertainties originating from statistical errors in some auxiliary measurements. The second type of systematic uncertainties is related to missing or incomplete knowledge, for

*On leave from SINP MSU

example model uncertainties. With the only statistical errors from auxiliary measurements the combined Likelihood can be written as: $Poiss(n|s + \tilde{c}\tau)Poiss(n_c|\tilde{c})$. Here the systematics term is replaced by an extra Poisson for control measurement n_c with an expectation $\tilde{c}$ treated as a nuisance parameter. The relation between control region c and the background in the signal region b is defined by the $\tau = b/c$ factor which can have its own uncertainty usually of the second type. Without this uncertainty the model with auxiliary measurements is equivalent to the simple Poisson model with the Gamma distributed systematics on the background, where the Gamma skewness and σ_Γ are defined by the τ factor: $Poiss(n|s + b')Gamma(b'|b, \tau)$ [7]. The uncertainty in the τ factor introduces yet another nuisance parameter which breaks this equivalence. However if this τ uncertainty can be described by Gaussian with σ_τ and convoluted with the Gamma distribution, the resulting distribution will still have Gamma like shape with $\sigma_b^2 \approx \sigma_\Gamma^2 + \sigma_\tau^2$. Such distribution can be used in the systematics term of the Likelihood function instead of the full form with two Poissons. Here different systematics pdf's has been used for the comparison.

The use of a Gamma distribution introduces a relation between the standard deviation and the mean value which complicates the factorization of correlated systematic uncertainties in multichannel case. Some simplification can be achieved with a Lognormal distribution which has a similar shape and allows factorization when it is written as: $(1 + \sigma_k)^{\delta_i}$, where σ_k is the relative uncertainty for k-channel and δ_i is the normally distributed i- nuisance parameter.

The combination of identical channels with the same signal to background ratio (S/B) without systematics uncertainties is equivalent to the splitting of Poisson statistics in one search; this property was used to verify the accuracy for the combined model. For channels with different S/B or different systematics the results depend upon observed statistics, especially in the limiting case when some channels have zero observations or the range of systematics is very different.

The results also depend upon the internal accuracy of the method and its implementation. There is no unique algorithm for estimating these internal uncertainties. For Profile Likelihood the intrinsic errors are related to the maximization of likelihood ratio(LR). The method based on Neyman construction, like FC, depends upon binning in the scan of parameter of interest and on the treatment of nuisance parameters. Bayesian integration uses either numerical integration or Markov Chain MC, in both cases the accuracy degrading fast with increased dimensionality. The Hybrid limits are estimated from MC toy experiments and the accuracy depends on the number of these toys which is tuned to have 0.5% accuracy in limits. However this does not guarantee that the whole range of nuisance parameter is explored.

3 Confidence limits for single channel and the combined model

The single number counting experiment with the anticipated relative systematic on the background σ_b has been used as a reference. The 95% CL upper limits versus expected backgrounds (N_{bkg}) and observations (N_{obs}) calculated with different methods are shown in Fig. 1. On this and the following plots only relatively small values of $N_{obs} - N_{bkg}$ are important for upper limits, for larger values the confidence belt becomes two sided but study of the flip-flop problem [9] is beyond the scope of this paper. The Bayesian approach produces the most conservative limits for a single channel. The Hybrid CLs limit is not defined for $N_{bkg} >> N_{obs}$ where $CL_b \to 0$. It is important to notice the behavior at $N_{obs} \approx 0$. The Bayesian and Hybrid CLs limits are independent on the background while the FC limits improve with larger background even for larger N_{obs}. The PLC fails at zero observation when the Likelihood ratio is collapsed to $lnQ = s$ and Wilk's theorem is not valid. The Hybrid CLs limits should be similar to the Bayesian credible intervals for this simple model and the difference can be related to the accuracy of implementations.

The systematics pdf have different effect in different methods, see Fig. 2. The FC is the most sensitive to the shapes because the nuisance parameter minimization is performed for each value of the scanned parameter of interest.

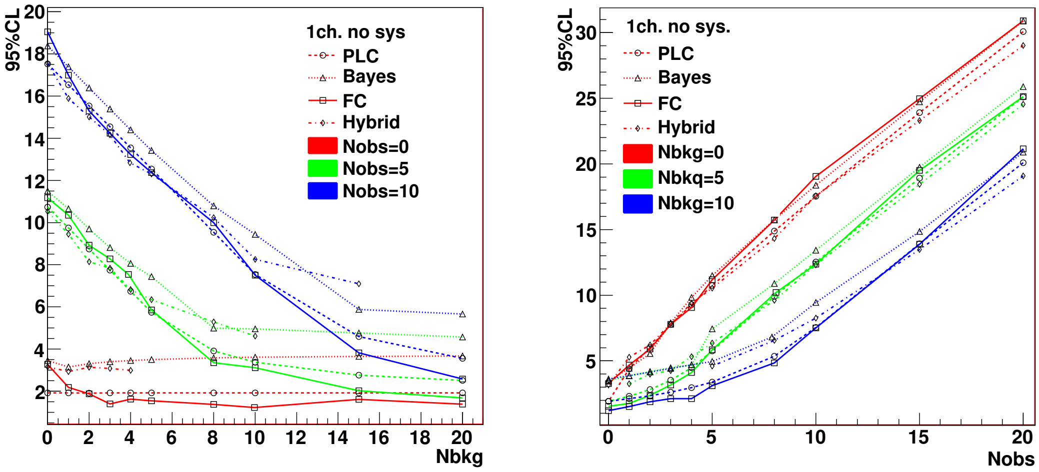

Fig. 1: 95% CL upper limits versus background and observations for single channel calculated with different methods.

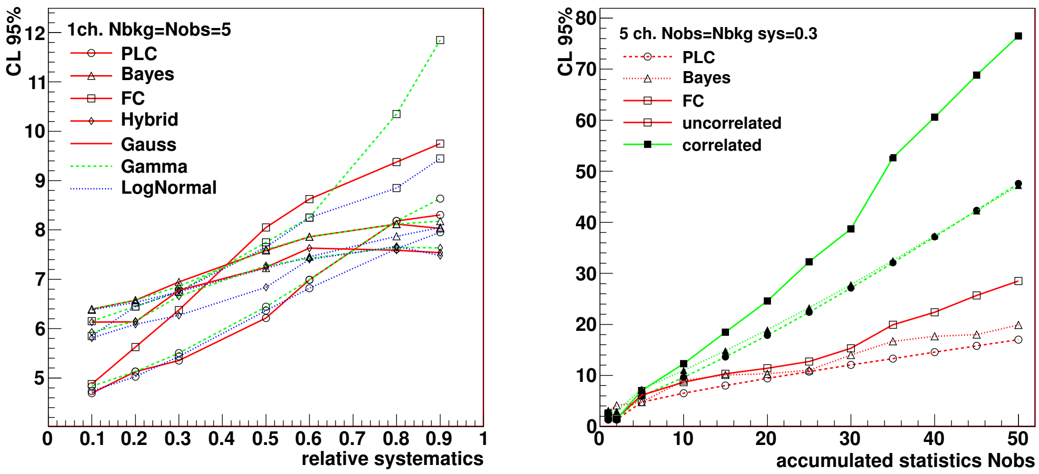

Fig. 2: Left: 95% CL upper limits for single channel versus relative systematics on background (σ_b) with different systematics pdf's. Right: limits versus accumulated statistics ($N_{obs} = N_{bkg}$) for the combined model of five channels with correlated and uncorrelated systematics $\sigma_b = 0.3$

There are some differences in limits even for this trivial case and the question is - which method to use? The obvious flaws in modeling can be spotted with the coverage test [9]. Figure 5 shows results of the coverage test for a single channel with relative systematics uncertainties on the background σ_b=0.3. All methods show no undercoverage with some overcoverage especially for the PLC. This coverage test can be done relatively easily for the simple model but becomes increasingly difficult or impossible for the multichannel search with many nuisance parameters when one has to guarantee no undercoverage for all possible combinations of nuisance parameters. The FC method should have correct coverage, or at least no undercoverage by construction. But for the PLC and Hybrid CLs methods coverage is not guaranteed and has to be checked. For the Bayesian credible limits the frequentist coverage test does not have a clear statistical interpretation and depends on the prior.

The combined model, here with five channels and uncorrelated Lognormal systematics with relative error $\sigma_b = 0.3$, is shown in Fig. 3. The Bayesian limit remains the most conservative while the

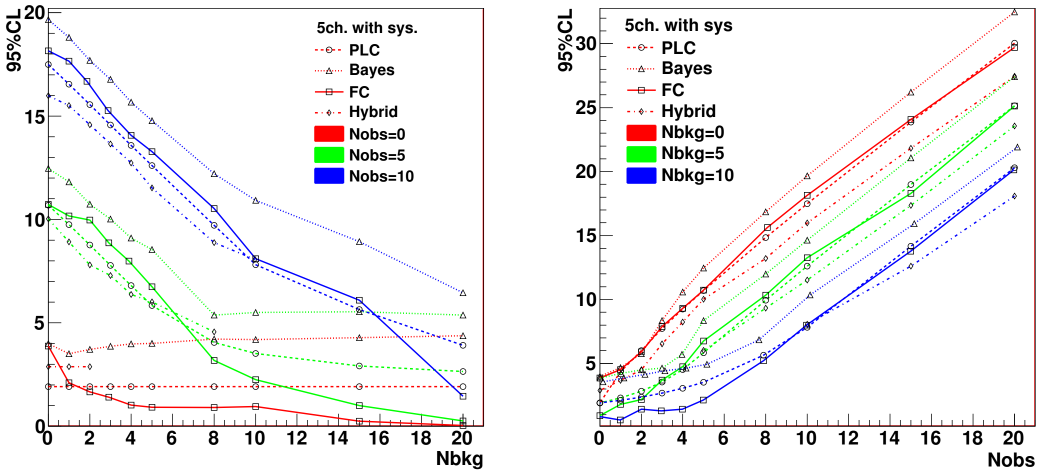

Fig. 3: 95% CL upper limits for the combined model of five identical channels with uncorrelated Lognormal systematics σ_b=0.3

Hybrid CLs becomes even more optimistic than PLC, which apparently is too low for small N_{obs}.

The effect of correlations is demonstrated in Fig. 2 at different statistics. The correlations reduce the number of independent nuisance parameters and usually degrade the limits. However in some cases the combination can be beneficial, that is, some channels with large systematics and low signal sensitivity can serve as a control measurement for the channels with higher sensitivity.

In reality the channels can have different S/B, different observed statistics and different ranges of systematics. Two typical cases are considered. The first model has one channel with five times higher signal sensitivity than the other four but zero observation. The second model with correlated uncertainties has one channel with three times larger systematics. The upper limits calculated for these two cases are shown in Fig. 4 versus difference N_{obs} minus N_{bkg}: $\Delta = N_{obs} - N_{bkg}$. The biggest difference is between the Bayesian and the PLC, FC and Hybrid limits for the first model. The Hybrid limits are the most optimistic at relatively large Δ. For smaller Δ the FC limits are very low, comparable with PLC.

In spite of large differences in the calculated limits, all methods, except the Hybrid CLs, show relatively good coverage, see Fig. 5.

4 Summary

The upper limits calculated with different statistical methods for the model comprised of several number counting experiments can have large variations depending on model parameters.

The Bayesian intervals with flat prior are the most conservative limits for all considered parameters, especially at high background expectation and combined models with very different channels.

The Feldman-Cousins upper limits are close to the Bayesian at small background but are getting more optimistic for larger background expectations. The results are also dependent on the steps in the scan of the parameter of interest and distribution of systematics which can result in empty intervals for some configurations.

The Profile Likelihood delivers the most optimistic limits for low observations and has to be avoided with zero or small observation in some of the channels.

The results of Hybrid CLs calculations are difficult to predict. While for a simple model it is rather similar to the conservative Bayesian limits, for the combined models the Hybrid limits become the most optimistic with some non zero observations. For low observation and downward fluctuation the Hybrid

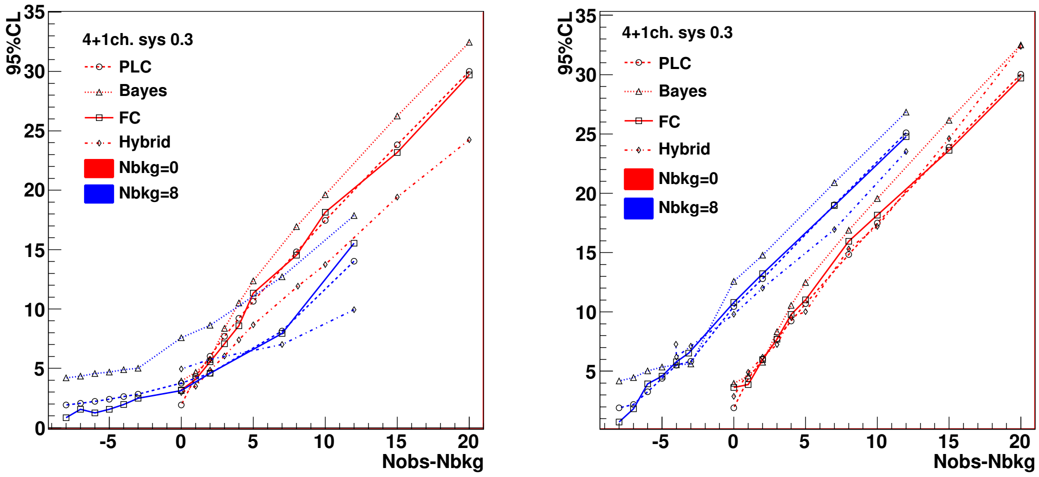

Fig. 4: Left: Upper limits versus $Nobs - Nbkg$ for the combined model of five channels with one five times more sensitive but with zero observation. All channels have Lognormal uncorrelated systematics with σ_b=0.3. Right: Combined model with correlated systematics σ_b=0.3 and one channel with σ_b=0.9

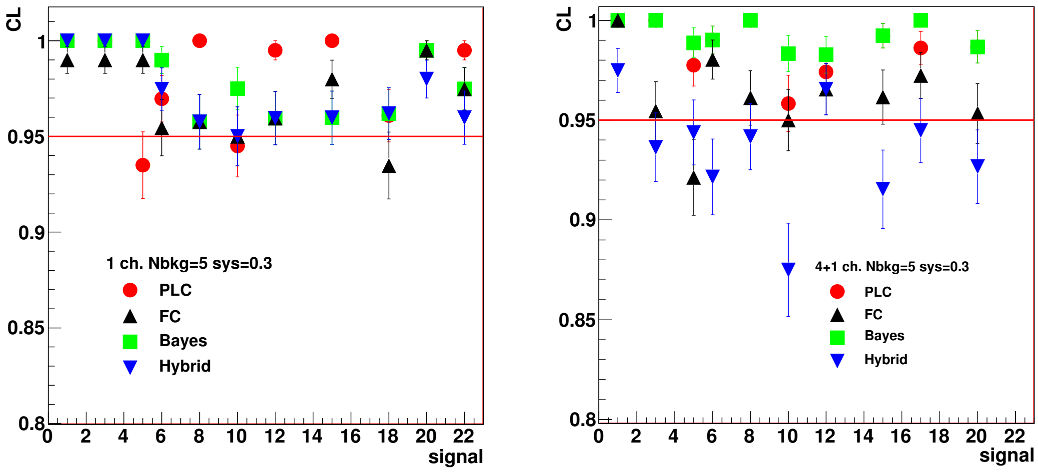

Fig. 5: Coverage of 95% CL upper limits calculated with different methods for single channel (σ_b=0.3) and combined model with five channels where one has five times higher signal sensitivity and zero observation (N_{MCtoys}=500)

limits are protected by the CLs ratio. Moreover the Hybrid method tends to undercover for combined models of very different channels.

The final choice of statistical method remains rather subjective. Apart of the conservative Bayesian limits, the usage of frequentist methods always have some drawbacks. For complicated models with many different channels the Feldman-Cousins limits looks the most attractive due to its intrinsic coverage.

In conclusion, the RooFit/RooStats package offers an excellent tool for the statistic modeling in LHC analysis.

5 Acknowledgments

The authors thanks the PHYSTAT2011 Organizers, VZ thanks R.Cousins, J.Linnemann and L.Lyons for the inspiring discussions and help.

References

[1] CMS PAS SUS-10-008, 2011
[2] K.Cranmer, PhysStat2007, 261, 2007
[3] L.Moneta et al. arXiv:1009.1003.438, 2010
[4] A. Caldwell, D. Kollar, K. Kroeninger, Comp. Phys. Comm. 180, 2197, 2009
[5] P.Sinervo, PhyStat2003,122,2003
[6] J.Heinrich, PhyStat2003,52 , 2003
[7] J. Linnemann arXiv:ph/0312059, 2003
[8] T.Junk, arXiv:hep-ex/9902006, 1999
[9] G.J. Feldman and R.D. Cousins, Phys.Rev. D 57, 3873, 1998
[10] R.D. Cousins and V.L. Highland, NIM A320, 331, 1992