The BIRS confidence interval problem treated with a hybrid Bayesian method

J. Conrad
KTH-Stockholm, Particle Physics Section, AlbaNova University Center, SE-10691 Stockholm

Abstract

Presented here the result of applying a hybrid-Bayesian method to the standard problem proposed by the BIRS committee. The method consists of a Bayesian treatment of nuisance parameters and a Neyman construction with likelihood ratio ordering for the parameter of prime interest. Results are presented for the single and double (2 channel) experiment as well as coverage.

1 Introduction

A popular technique in particle physics used to calculate confidence intervals (CI) is the technique suggested by Feldman & Cousins [1]. This method is based on a Neyman construction where as ordering principle the likelihood ratio:

$$R(s,n)_{\mathcal{L}} = \frac{\mathcal{L}(n|s+b)}{\mathcal{L}(n|s_{best}+b)} \tag{1}$$

where s is the hypothesis, n the experimental outcome, b the expected background, s_{best} is the hypothesis most compatible with n and $\mathcal{L}$ the Likelihood function. In the originally proposed method by Feldman & Cousins, only the presence of background was considered and it was assumed to be exactly known. In 1992 Cousins & Highland [2] proposed a method which is based on a Bayesian treatment of the nuisance parameters. The main idea is to use a probability density function (pdf) in which the average is taken over the nuisance parameter:

$$P(n|s,\epsilon) \longrightarrow \int P(n|s,\epsilon')P(\epsilon'|\epsilon)d\,\epsilon' := q(n|s,\epsilon) \tag{2}$$

where ϵ' is the true value of the nuisance parameter, ϵ denotes its estimate and s and n symbolize the signal hypothesis and the experimental outcome respectively. The method has since been generalized [3] to operate with the Feldman & Cousins ordering scheme and taking into account both efficiency and background uncertainties as well as correlations.

2 The BIRS problem

We consider a Poisson process with nuisance parameters ϵ (efficiency) and b (background) measured in subsidiary measurements. ϵ_{true} and b_{true} denote the true value of the nuisance parameters. The uncertainties in the nuisance parameter is in our approach described by a Gaussian PDF, $G(b|b_{true}, \sigma_b)$ and $G(\epsilon|\epsilon_{true}, \sigma_\epsilon)$. We interpret the uncertainties given in the BIRS problem formulation as being the width of the Gaussian distributions centered on the true value. We then (assuming an appropriate prior) retrieve a posterior distribution of a Gaussian with the same width but centered on the measured value of the nuisance parameter. The PDFs used in the Neyman construction are consequently given by:

$$q(n|s,\epsilon,b) = \int_0^\infty \int_0^\infty P(n|\epsilon_{true}s + b_{true})G(\epsilon_{true}|\epsilon,\sigma_\epsilon)G(b_{true}|b,\sigma_b)d\,\epsilon_{true}d\,b_{true} \tag{3}$$

where P denotes the Poisson distribution.

The typical values we examine as suggested: b = 3.0 $\pm$ 0.3 and ϵ = 1.0 $\pm$ 0.1. We consider intervals for experimental outcomes n=0,.....20, as well as b=0,...6,5. for the cases n=0 and n=3 for the background dependence. We present results for the two-channel experiments and coverage for the values presented by the BIRS committee. Further results on coverage, power and combined experiments can be found in [4][5].

The code (some of it however not released) used to do the present calculations (and the results presented in [4] and [5]) can treat multi-channel experiments with correlated uncertainties and Gaussian, flat and log-Normal parameterizations of pdfs and can be used in a mode optimized for coverage calculations. It also features a modified hybrid method (I usually call it the hybrid-hybrid method) suggested in a comment by [6].

3 Results

3.1 The intervals for n=0,...,20

The intervals for number of observed events ranging from 0 to 20 can be found in table 1.

n	interval	n	interval	n	interval	n	interval
0	[0.00, 0.94]	5	[0.00, 7.37]	11	[3.02, 15.32]	17	[7.38, 7.39]
1	[0.00, 1.86]	6	[0.13, 8.42]	12	[4.07, 16.35]	18	[8.43, 24.25]
2	[0.00, 3.01]	7	[0.87, 9.89]	13	[4.48, 17.81]	19	[9.47, 25.27]
3	[0.00, 4.47]	8	[1.49, 11.36]	14	[5.49, 18.84]	20	[9.90, 26.73]
4	[0.00, 5.90]	9	[1.87, 12.39]	15	[5.91, 20.30]		
5	[0.00, 7.37]	10	[2.63, 13.86]	16	[6.96, 21.54]		

Table 1
CIs calculated with the hybrid Bayesian method. The efficiency is assumed to be $\epsilon = 1 \pm 0.1$ and background, $b = 3 \pm 0.3$

3.2 Background dependence

The dependence of the CI on the expected background can be seen in figure 1. As in the case without uncertainties in nuisance parameters, the upper limit decreases (the interval width becomes smaller) with increasing background expectation.

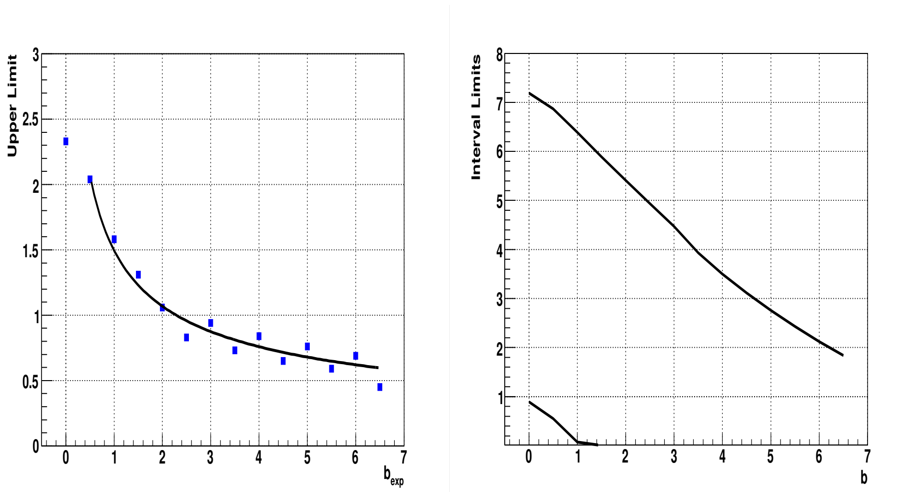

Fig. 1. Left plot: the background dependence of the upper limit as calculated with the presented method for n=0. Just for fun, the line is a fit of the function A/b^{α}, where best-fit $\alpha \sim 0.5$. Right plot: the background dependence of the CI in case n=3.

The coverage as a function of signal parameter s for the different combinations of the nuisance parameters [ϵ,b] can be found in figure 2. In general a modest (2-3 %ish) over-coverage can be observed. There is no visible difference between the considered cases.

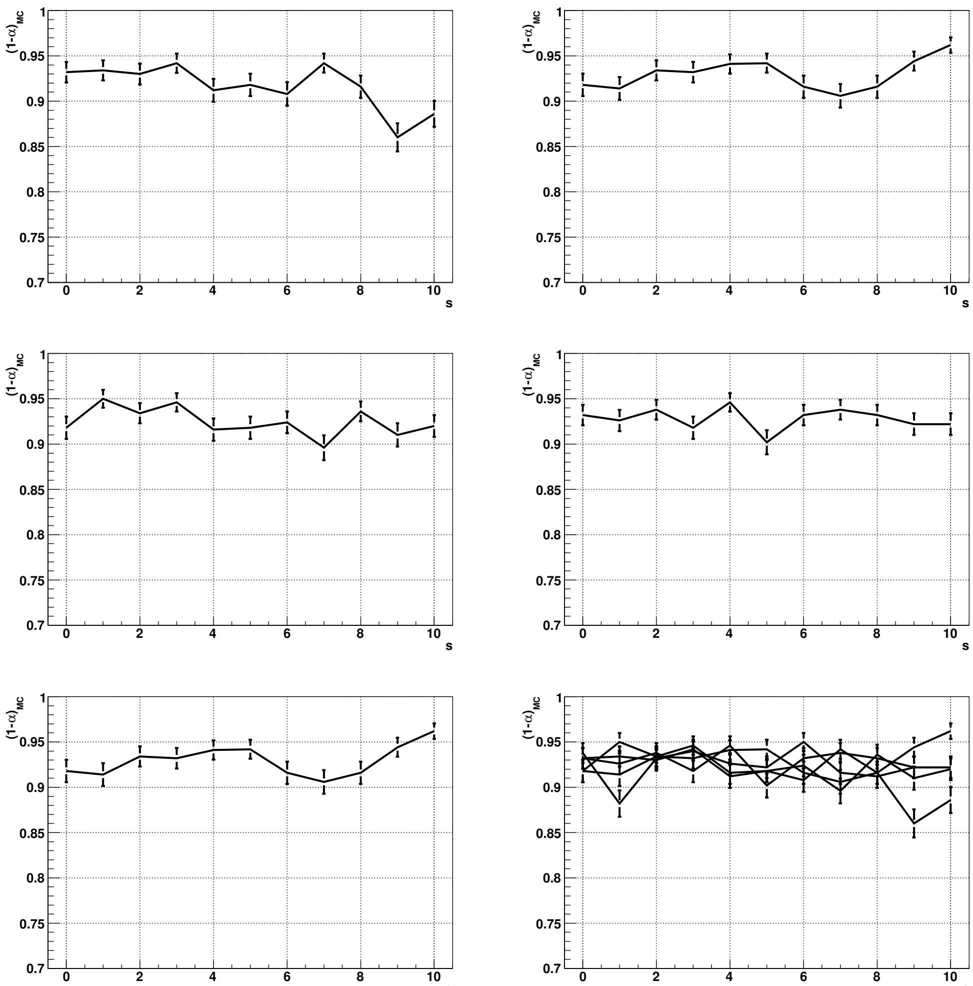

Fig. 2. Coverage versus s at (counting from left to right), [ϵ,b] = [1,3],[1.1,3],[0.9,3],[1,3.3] [1,2.7] and all coverage curves superimposed. The nominal coverage is 0.9

3.4 Two Channel Case

As suggested in the exercise, we present here results for a combination of two channels. The method is described in [5]. Uncertainties in nuisance parameters are assumed uncorrelated.

n_1	n_2	interval
0	0	[0 , 0.94]
0	1	[0 , 1.86]
1	1	[0 , 2.98]
1	2	[0 , 4.12]
1	3	[0 , 5.07]
2	2	[0 , 5.22]
3	3	[0.59 , 7.47]

Table 2
CIs calculated for combination of two independent measurements. The efficiency is assumed to be $\epsilon = 0.5 \pm 0.1/\sqrt{2}$ and background, $b = 1.5 \pm 0.3/\sqrt{2}$

It can be observed that the limits obtained are always (except for zero and 1 observed event) are more stringent than if the same number of events had been measured in one single channel (with doubled background and $\sqrt{2}$ times the uncertainties).

4 Some remarks

For the problem at hand the Bayesian treatment of nuisance parameters results in moderate over-coverage for a flat prior (which is the implicit assumption here). More detailed coverage studies for different cases can be found in [4] and [5]. There we also describe cases with different posterior PDFs. In particular, if the uncertainties in the efficiency become larger a more reasonable model is the log-Normal distribution. The coverage of the multi-channel case has not been studied. At PHYSTAT 05 it was shown that in the entirely Bayesian approach a flat prior in the nuisance parameter leads to under-coverage for the multi-channel case [7], which is probably true also for the hybrid method presented here.

From a conceptual point of view a legitimate question is what the benefit is of treating nuisance parameters Bayesian while treating the parameter of interest frequentistically. In some sense the method suffers from the drawbacks of both worlds: it still gives decreasing limits for increasing background (a feature which is found unintuitive by most physicists, for statisticians this is not a problem) and still you have to put in some prior believe, this time for your nuisance parameter. A valid argument, in my mind is that for some systematic uncertainties (as for example the uncertainty produced by theory), there is no other way than to use a Bayesian PDF to describe the uncertainties (see also Zech's mail). This would in some sense suggest a fully Bayesian treatment of the problem.

The reason why the hybrid method presented here is relatively popular ([3] has 68 citations since 2003, [2] has 259 citations since 1992) is that physicists are looking for an extension to the *en vogue* Feldman & Cousins proposal and simple to use code is readily available and one would hope that they feel comfortable since the coverage of the method has been comparably well studied (though judging from the number of citations for the coverage studies, my suspicion is that most of the people care much more that code is available than for the fact that the method's coverage has been studied).
Fully frequentist extensions to Feldman & Cousins have been presented but are as far as I know less well studied and documented (Feldman's profile likelihood Neyman-construction, see for example [8],[9] and [10]).

5 Acknowledgements

Thanks to F. Tegenfeldt for help with some of the computations.

References

[1] G. J. Feldman and R. D. Cousins, Phys. Rev. **D57**, 3873, (1998)

[2] R. D. Cousins and V. L. Highland, Nucl. Instrum. Meth. A **320** (1992) 331

[3] J. Conrad, O. Botner, A. Hallgren and C. P. de los Heros, Phys. Rev. **D67** (2003) 012002

[4] F. Tegenfeldt and J. Conrad, Nucl. Instrum. Meth. A **539** (2005) 407

[5] J. Conrad and F. Tegenfeldt, arXiv:physics/0511055.

[6] G. C. Hill, Phys. Rev. D **67** (2003) 118101 [arXiv:physics/0302057].

[7] J. Heinrich, Proceedings of PHYSTAT 2005.

[8] G. J. Feldman, Fermilab workshop on Confidence Limits (2000)

[9] K. S. Cranmer, eConf **C030908** (2003) WEMT004 [arXiv:physics/0310108].

[10] G. Punzi, arXiv:physics/0511202.