Statistical analysis of detection of, and upper limits on, dark matter lines

ROYAL INSTITUTE OF TECHNOLOGY

J. Conrad[1], J. Scargle[2] and T. Ylinen[1]

[1]Department of Physics, Royal Institute of Technology (KTH), Stockholm, Sweden
[2]NASA Ames Research Center, Moffett Field, CA, USA

Contact: conrad@particle.kth.se, jeffrey@cosmic.arc.nasa.gov, tomiy@particle.kth.se

Ames Research Center

Introduction

The detection of a line signal from dark matter annihilations over a background from conventional astrophysical sources is one of the most important statistical problems faced by GLAST. The simplest approach is to divide the data into a signal region (where signal and background is supposed to be present) and a background region (where only background is supposed to be present), from which the contribution from the background to the signal region counts is estimated. The estimated background is then a random variable which follows a Poisson distribution.

In this simple approach, the likelihood model under the two hypotheses M1 and M0 for the measurement is given by:

$$M1: L(n_{S+B}, n_B \mid S+B) = \frac{(S+B)^{n_{S+B}} e^{-(S+B)}}{n_{S+B}!} \cdot \frac{B^{n_B} e^{-B}}{n_B!}$$

$$M0: L(n_{S+B}, n_B \mid B) = \frac{B^{n_{S+B}} e^{-B}}{n_{S+B}!} \cdot \frac{B^{n_B} e^{-B}}{n_B!}$$

where n_{S+B} is the observed number of counts in the signal region, n_B is the estimated number of background counts in the signal region, S and B are the signal and background parameters in the Poisson process respectively. In this poster contribution we use this simple model as a benchmark to compare three different methods for calculating upper limits and claim discovery.

Bayesian approach

In Bayesian theory, a test statistics can be defined by taking the ratio of the Bayes factors for the two hypotheses:

$$B_{fact,M1} = \iint (S+B)^{n_{S+B}} B^{n_B} e^{-(S+2B)} P(S) P(B) dS dB$$

$$B_{fact,M0} = \int B^{n_{S+B}+n_B} e^{-2B} P(B) dB$$

The ratio measures the probability that a signal is present independent of the signal and background strength.

Frequentist approach

Profile Likelihood

To fit the model to the data we might wish to use the log-likelihood function, multiplied by a factor -2 so that it behaves asymptotically like the chi-square. In this approach an uncertainty in the nuisance parameter can be treated by maximizing the log likelihood over the nuisance parameters. After this the likelihood function ("profile likelihood") is a function of the parameters of interest alone. In the present case, the maximization can be done by requiring:

$$\frac{\partial}{\partial B} L(n_{S+B}, n_B; S, B) = \frac{n_{S+B}}{S+B} + \frac{n_B}{B} - 2 = 0$$

and setting

$$pl(S) = L(S, \hat{B}(S), n_{S+B}, n_B)$$

Feldman & Cousins

A popular technique to calculate confidence intervals in recent years is the technique suggested by Feldman & Cousins [2]. The method consists of constructing an acceptance region for each possible hypothesis (in the way as proposed by Neyman [3] and fixing the limits of the region by including experimental outcomes according to rank which is given by the likelihood ratio. Throughout this note we consider Poisson distributions with experimental outcome *n*, hypothesis parameter *S* and (possibly not exactly) known background *B*:

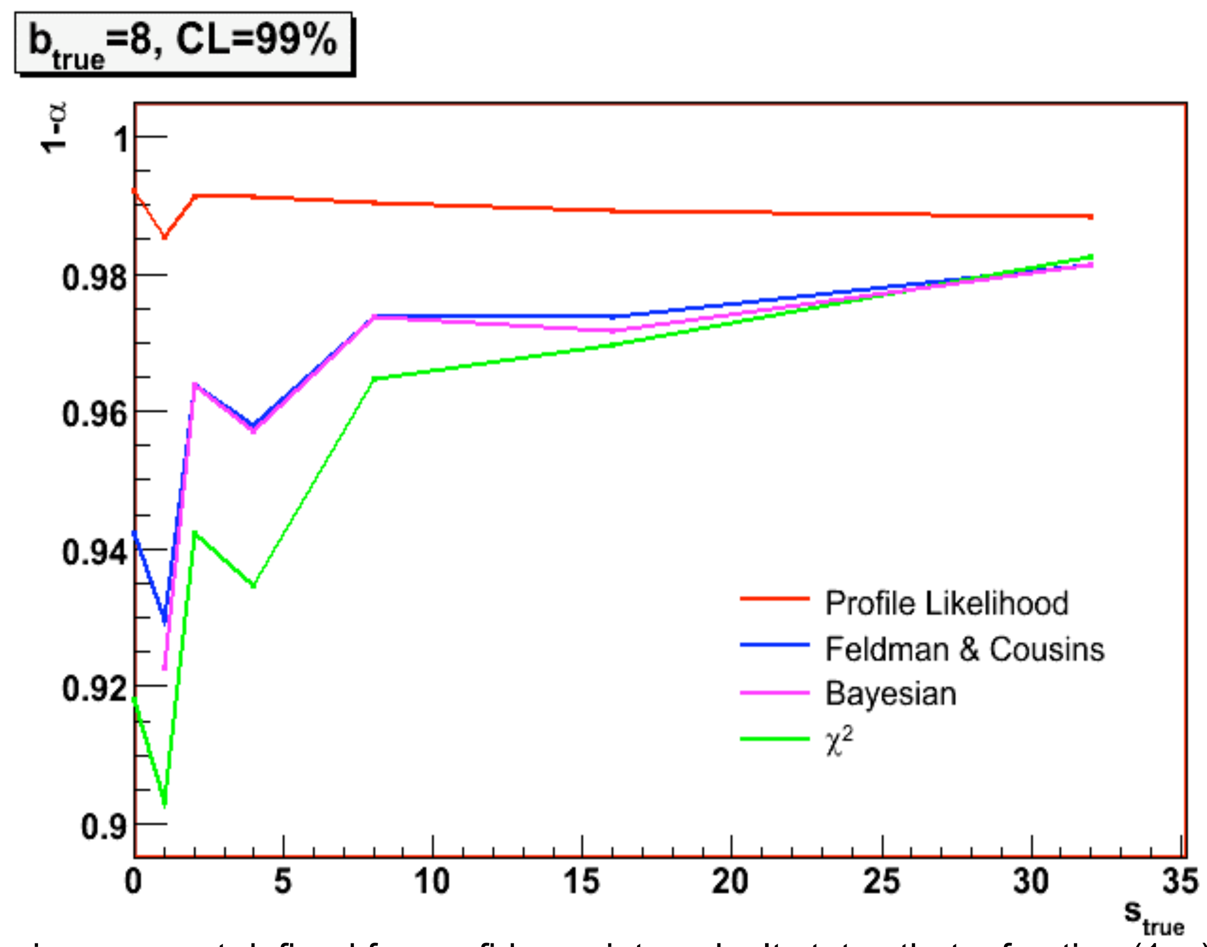
$$R(S, n_{S+B})_L = \frac{L(n_{S+B} \mid S+B)}{L(n_{S+B} \mid S_{best} + B)}$$

Where *S* is the hypothesis, n_{S+B} the experimental outcome, *B* the expected background, S_{best} is the hypothesis most compatible with n_{S+B} and *L* the Likelihood function. Note that in this method it is assumed that the expected background (also called nuisance parameter) is perfectly known.

Detection and Upper Limit

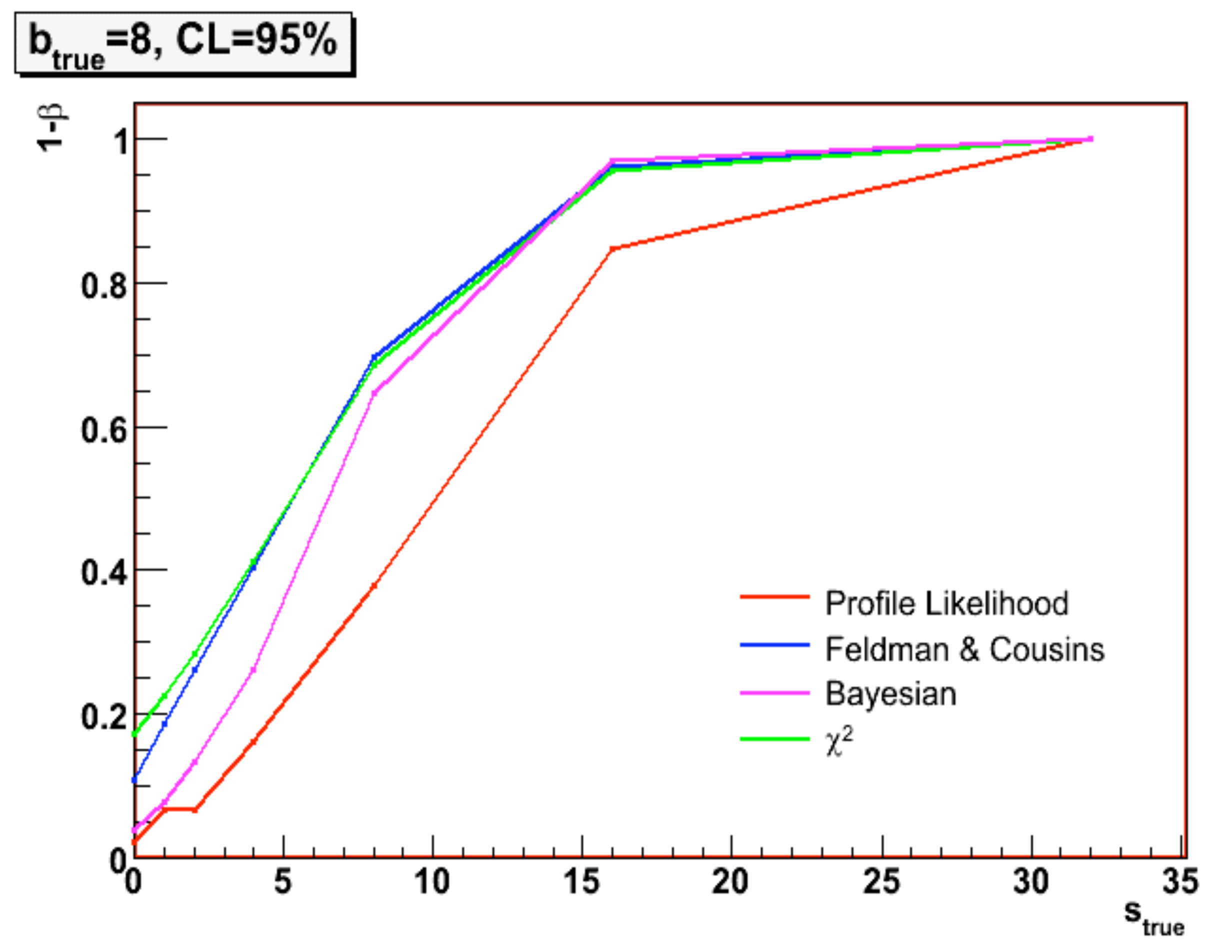
The question of presence of signal (detection) and calculation of confidence intervals are in general different topics in mathematical statistics (see e.g. [4]). The Bayesian method described above represents a hypothesis test, the frequentist methods represent confidence interval calculation methods. Also confidence intervals can be used for claiming detection in requiring that the null hypothesis (S=0 in our case) is not contained in the calculated confidence interval.

Coverage



Coverage is a concept defined for confidence intervals. It states that a fraction (1-α) of an infinite set of confidence intervals obtained from an infinite number of identical experiments should contain the true value of the parameter to be estimated. In the context of hypothesis testing, 1 - coverage corresponds to the type 1 error, i.e. the probability that the null hypothesis gets rejected though it is true.

Power



Power is a concept defined for hypothesis tests. The power of a test is the probability that the null hypothesis is rejected under the condition that the alternative hypothesis is true (power = 1 – probability for a type II error, or β). In using confidence intervals for hypothesis testing, power is just the fraction of cases where the null hypothesis (S=0) is not contained in the interval given that the alternative hypothesis (S >0) is true. For null hypothesis = alternative hypothesis, power reduces to 1 - coverage.

Conclusions

In designing a hypothesis test or a method for confidence interval calculation, the first requirement is on the probability for a false detection (or how often is the true signal not contained in the interval. From our results, it can be seen that only the profile likelihood has the nominal coverage (nominal rate of type I error). Followed by the Feldman & Cousins method (which ignores the uncertainties in the background), and the Bayesian method. The χ^2 method, often used for establishing a signal undercovers by as much as 10 %, also here probably due to the fact that it ignores uncertainties in the background and that it should become less reliable for low statistics.

Allowing more false detections intuitively should imply larger power. The profile likelihood has worst power, the χ^2 method has largest power. However, one needs to keep in mind that using the χ^2 a detection nominally on 99 % confidence level only corresponds to between 90 and 96 % actual confidence level.

Comparing power for tests (CI calculation methods) which do not have the same coverage does not make much sense. The choice of test should be a two step process: 1) calculate the de facto coverage (false detection rate), 2) among those tests which have similar coverage choose the one with largest power.

References

[1] W. A. Rolke, A. M. Lopez and J. Conrad, ”Confidence Intervals with Frequentist Treatment of Statistical and Systematic Uncertainties”, Nucl. Instrum. Meth. A, 551, (2005) 493 [arXiv:physics/0403059].

[2] R. D. Cousins, G. J. Feldman, “A Unified Approach to the Classical Statistical Analysis of Small Signals”, Phys. Rev, D57, (1998) 3873.

[3] J. Neyman, Phil. Trans. Royal Soc. London A, 333, (1937).

[4] J. Conrad, ”Discovery and Upper Limits in Search for Exotic Physics with Neutrino Telescopes”, arXiv:astro-ph/0612082