STATISTICAL CHALLENGES IN MODERN ASTRONOMY IV
ASP Conference Series, Vol. 371, 2007
G. Jogesh Babu and Eric D. Feigelson

Discussion on "The Small-N Problem in High Energy Physics" by G. Cowan and "Bayesian Methods in Particle Physics from Small-N to Large" by H. B. Prosper

Michael Woodroofe and Bodhisattva Sen

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

1. Introduction

Though they address similar problems, the papers are different in important ways. Professor Cowan has provided an informative review of the problem, and its implications to high energy physics. It is surprising that such a simple problem is so perplexing. Professor Prosper writes as an advocate of the Bayesian approach and shows how it can be implemented. As far a foundations go, we applaud the emphasis on specific problems, believing that success stories will do more to spread the Bayesian word than foundational arguments. We also agree that both Bayesian and frequentist methods are needed. We enjoyed both papers having worked on related problems ourselves.

2. Cowan's Paper

We will center our discussion around the simplest version of the problem, setting an upper confidence limit for the expected signal when the expected background is known, hoping that a solution to the simple problem will provide some insight into more complex ones. Thus, suppose one observes a count N, combining background events and signal events. So, writing N_b and N_s for the (unobserved) number of background and signal events, one observes (only) $N = N_b + N_s$. Suppose now that N_b and N_s are independent Poisson variables with means b and s respectively, so that N has a Poisson distribution with mean $b+s$, as in Equation (1) of the paper. Suppose further that b is (at least approximately) known, leaving s as the unknown parameter. An upper confidence bound for s is desired. This seems innocuous, but is made much harder by imposing the following two litmus tests:

(L1) The solution should be equivariant under increasing transformations of s.
(L2) The solution should not depend on b when $N = 0$.

We will confess to some doubt about the importance of (L1), but the consensus among the physicists that we know is that it is necessary. To understand (L2), observe that if $N = 0$, then necessarily $N_b = 0$ and $N_s = 0$. Once that is known, the a priori expected number of background events cannot be relevant.

The textbook solution to this problem, Equations (3) and (5) in the paper, satisfies the first litmus test, but clearly violates the second. For a confidence level of 90%, the upper confidence bound so obtained is $\max[0, 2.31 - b]$, which is 0 when $b > 2.31$. A second frequentist approach adapts the method of high relative likelihood (Feldman & Cousins 1998) to the one sided case. The intervals consist of s for which either $s < \hat{s}$ or $\hat{s} \leq s$ and $L(s|n)/L(\hat{s}|n) \geq c_s$, where $\hat{s} = \max[0, N - b]$ and c_s is chosen to control the coverage probability. This method always delivers a positive upper limit, but the limit depends on b when $N = 0$ and can be arbitrarily small. It can be argued that 2.31 is the correct answer, from a frequentist perspective. It would be the correct answer if $N_s = 0$ were observed; and if $N = 0$, then $N_s = 0$.

The Bayesian solution with a uniform prior has several things to recommend it. Writing P_s and P^n to denote frequentist probability when the parameter value is s and conditional probability in the Bayesian model when $N = n$, the upper limit $s_{\mathrm{up}} = s_{\mathrm{up}}(n)$ for a level $(1-\gamma)$ credible bound is determined by the condition $P^n[s > s_{up}] = \gamma$. Here

$$P^n[s > s_{\mathrm{up}}] = \frac{\int_{s_{\mathrm{up}}}^{\infty}(s+b)^n e^{-(s+b)}ds/n!}{\int_0^{\infty}(s+b)^n e^{-(s+b)}ds/n!} = \frac{\int_{s_{\mathrm{up}}+b}^{\infty} s^n e^{-s}ds}{\int_b^{\infty} s^n e^{-s}ds},$$

which simplifies to $P^n[s > s_{\mathrm{up}}] = P_{s_{\mathrm{up}}}[N \leq n]/P_0[N \leq n] = P_{s_{\mathrm{up}}}[N \leq n | N_b \leq n]$. So, the equation for $s_{\mathrm{up}} = s_{\mathrm{up}}(n)$ may be written as

$$P_{s_{\mathrm{up}}}[N \leq n | N_b \leq n] = \gamma. \tag{1}$$

It then follows from (1) that the frequentist coverage probability of the Bayesian intervals is at least $1 - \gamma$,

$$P_s\left[s \leq s_{\mathrm{up}}(N)\right] > 1 - \gamma \tag{2}$$

for all $0 \leq s < \infty$.

As noted in the paper, the uniform prior is not invariant under transformations of s, but invariance of the interval can be obtained by simply forgetting the derivation and keeping the result. The solutions to (1) may reasonably be called *conditional frequentist* upper limits, as in Roe & Woodroofe (1999). With this definition, the conditional frequentist limits satisfy both litmus tests: the construction is equivariant under increasing transformations and does not depend on b when $N = 0$. The conditional frequentist intervals are conservative for small n when compared to the frequentist solutions. A conservative solution seems necessary in order to avoid dependence on b when $N = 0$ and the embarrassing possibility of a degenerate interval.

Moving beyond the simplest formulation, there are several issues. First, confidence intervals may be desired, instead of a confidence bound. Both the method of high relative likelihood (Feldman & Cousins 1998) and Bayesian methods (Roe & Woodroofe 2000) extend naturally to intervals, as does the connection between the Bayesian and conditional frequentist approaches. There may be uncertainty in b or other nuisance parameters. Relations like (2) do not extend to the two-sided case, although more conservative bounds can be obtained (Roe & Woodroofe 2000). Heinrich (2006) has reported some interesting simulations of Bayesian confidence limits with flat priors in the presence of nuisance parameters. These suggest that relations like (2) may hold in some cases. Sen et al. (2006) develop several examples of the use of regions of high likelihood in the presence of nuisance parameters.

3. Prosper's Paper

There are several specific problems addressed in the paper. We will center our discussion around problems 2 and especially 3, "Finding a Needle in a Haystack", but emphasizing a different aspect of the problem than Professor Prosper.

As above suppose that we observe a count of the form $N = N_b + N_s \sim \text{Poisson}(b+s)$, where N_b and N_s are (unobserved) independent Poisson variables with means b and s. Suppose also that with each event a mark X is observed and X has density f_b for a background event and f_s for a signal event. Then the overall density of X is

$$f(x) = \frac{bf_b(x) + sf_s(x)}{b+s}.$$

Suppose initially that b, f_b and f_s are known and that s has a prior distribution function Π. Suppose also that $N = n$ has been observed along with the marks $X_1 = x_1, \ldots, X_n = x_n$.

Is there really a signal? The conventional way to formulate this question is as a testing problem $H_0 : s = 0$. Letting π_0 be the prior probability that $s = 0$, the posterior odds for $s = 0$ given the data are

$$\frac{\pi_0^*}{1-\pi_0^*} = \left(\frac{\pi_0}{1-\pi_0}\right) \frac{1}{\sum_{k=0}^{n} \binom{n}{k} C_{n,k} b^{-k} \int_0^\infty s^k e^{-s} \pi(s) ds}, \tag{3}$$

where $\pi(s)$ is the conditional density of s given $s > 0$,

$$C_{n,k} = \frac{1}{\binom{n}{k}} \sum_{j_1+\cdots+j_n=k} \prod_{i=1}^{n} \left(\frac{f_s(x_i)}{f_b(x_i)}\right)^{j_i},$$

and $j_i = 0$ or 1. The optimal Bayesian decision procedure is then to decide for $s > 0$ if and only if π_0^* is sufficiently small. Thus, the procedure depends crucially on π_0. It also depends on π but, within broad limits, that dependence is much less pronounced. For example, if $\pi_0 = 0$ then $\pi_0^* = 0$ for any n and x.

Another way to formulate the question is to ask is $N_s > 0$; that is, have we seen a signal event yet? The probability of $N_s = 0$ given the data may be computed as

$$\frac{\int_0^\infty e^{-s} d\Pi(s)}{\sum_{k=0}^{n} \binom{n}{k} C_{n,k} b^{-k} \int_0^\infty s^k e^{-s} d\Pi(s)}, \tag{4}$$

and the optimal Bayesian decision procedure is again to decide that $N_s > 0$ if and only if this posterior probability is sufficiently small. This appears to depend much less crucially on π_0. For example, it is possible to have large value of (4) even if $\pi_0 = 0$.

References

Feldman, G. & Cousins, R. 1998. Unified approach to the classical analysis of small signals. Physics Reviews D 57, 3873-3889

Heinrich, J. 2006. Limits and nuisance parameters. Talk at the Samsi Workshop on Particle Physics (http://www.samsi.info/200506/astro/workinggroup/phy/)

Roe, B., & Woodroofe, M. 1999. Improved probability method for estimating signal in the presence of background. Physics Reviews D 60, 053009

Roe, B., & Woodroofe, M. 2000. Setting confidence belts. Physics Reviews D 63, 013009

Sen, B., Walker, M., & Woodroofe, M. 2006. On the unified method with nuisance parameters. Talk at the Samsi Workshop on Particle Physics (http://www.samsi.info/200506/astro/workinggroup/phy/)