Evaluation of Two Methods for Incorporating a Systematic Uncertainty into a Test of the Background-only Hypothesis for a Poisson Process

Jordan Tucker
Dept. of Physics and Astronomy, University of California, Los Angeles, California, USA

Abstract
Hypothesis tests for the presence of new sources of Poisson counts amidst background processes are frequently performed in high energy physics, gamma ray astronomy, and other branches of science. This talk briefly summarizes work in which we evaluate two classes of algorithms for dealing with uncertainty in the mean background in such tests.

This talk briefly summarizes studies, performed with Robert Cousins and described in Ref. [1], of two methods for incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process. In a situation common in both gamma-ray astronomy (GRA) and high-energy physics (HEP), \( n_{\text{on}} \) events are observed from a Poisson process with mean \( \mu_s + \mu_b \); the signal mean \( \mu_s \) is of interest, while the background mean \( \mu_b \) is a nuisance parameter. In this work, we study tests of the background-only null hypothesis \( (\mu_s = 0) \) in two prototypical problems in GRA and HEP as follows.

The “on/off” problem. In GRA, \( n_{\text{on}} \) photons are detected with a telescope pointed on-source, i.e. with some putative source in the field of view; and \( n_{\text{off}} \) photons are detected with the telescope pointed off-source. The ratio \( \tau \) of observing time \( t_{\text{off}} / t_{\text{on}} \) is assumed known exactly. In the analogous example from HEP, one counts \( n_{\text{on}} \) events in a signal region where one is looking for an excess above background. One observes \( n_{\text{off}} \) events in a background control (sideband) region where no excess is expected. The ratio \( \tau \) of sideband to signal region events under the background-only null hypothesis is again assumed known.

The “Gaussian-mean background” problem. In another scenario, there is a subsidiary measurement which determines \( \mu_b \) with normal (Gaussian) uncertainty with rms deviation \( \sigma_b \). We assume \( \sigma_b \) to be precisely known, either absolutely, or as a set fraction of \( \mu_b \).

In either problem, for a data set one can then proceed to calculate the tail probability \( (p\text{-value}) \) under the null hypothesis. In HEP, one typically quotes the significance \( S \) (known in the statistics literature as the \( Z \)-value) of the data set, namely the \( p\text{-value} \) converted to equivalent normal standard deviations.

As detailed by Linnemann [2] at PhyStat 2003, there is an approximate correspondence between the two problems. For the on/off problem, an estimate of the mean background in the signal region is

\[
\hat{\mu}_b = n_{\text{off}} / \tau,
\]

and the (rough) uncertainty on this estimate is then

\[
\sigma_b = \sqrt{n_{\text{off}} / \tau}.
\]

Combining the two equations and eliminating \( n_{\text{off}} \), we have

\[
\tau = \hat{\mu}_b / \sigma_b^2.
\]

This suggests that a recipe to estimate the significance for one of the prototypical problems can be applied to the other; then the performance of the recipe can be studied. Here, we quantify performance in terms of coverage.

There is a frequentist solution to the on/off problem, discussed by Linnemann [2] at PhyStat 2003 and by a very few references in that work. The key idea is to reformulate the null hypothesis: if the
signal mean $\mu_s$ is zero, then the ratio of Poisson means in the sideband region and the signal region is exactly $\tau$ (i.e. it is the ratio expected given background alone, and no signal). Then, one can use the standard frequentist solution for the hypothesis test for the ratio of Poisson means, expressed in terms of binomial probabilities. This recipe ($Z_{Bi}$) is then easily carried out, for example in ROOT [3]. In ROOT, one function call returns a $p$-value, and another calculates the equivalent number of standard deviations, $Z$. By the properties of the frequentist construction, $Z_{Bi}$ never under-covers, but it over-covers due to the discreteness of $n$, especially for small counts.

It is common in HEP to integrate out the nuisance parameter (here, the unknown background mean $\mu_b$) in an otherwise frequentist calculation (Cousins and Highland [4] integrated out an unknown luminosity). For the Gaussian-mean background problem, starting from the Poisson probability to obtain $n_{\text{on}}$ or more background events:

$$pp = \sum_{j=n_{\text{on}}}^{\infty} e^{-\mu_b} \frac{\mu_b^j}{j!},$$

one can calculate the weighted average over a given pdf for the background mean $\mu_b$, to obtain the $p$-value:

$$p = \int pp(\mu_b) d\mu_b.$$

Then, depending on the pdf one chooses for $\mu_b$, one has different recipes to calculate $Z$-values.

Choosing a gamma function pdf for $\mu_b$ (the result of a flat prior times the likelihood function from the Poisson sideband observation of $n_{\text{off}}$), one has the recipe $Z_{\Gamma}$. Amazingly, this yields an answer which is identical [2] to that of the frequentist-constructed $Z_{Bi}$!

Letting the pdf for $\mu_b$ be a Gaussian with rms deviation $\sigma_b$ as above, one obtains the recipe $Z_N$ (with the subscript denoting normal). This method was presented in a poster at PhyStat 2005 by Bityukov [5] et al. and was the recommendation out of the CMS Higgs group, adopted by CMS. But, the frequentist coverage of $Z_N$ is not guaranteed, and Cranmer [6] gave examples where it was poor.

We check the coverage of the two recipes, $Z_{Bi}$ and $Z_N$, scanning over the true background mean $\mu_b$ and the other experimental setup parameter ($\tau$ for the on/off problem, or $f = \sigma_b/\mu_b$ for the Gaussian-mean background problem). Choosing a “claimed” $Z$-value, $Z_{\text{claim}}$, from the common choices 1.28 (corresponding to a $p$-value of 0.1), 3, or 5, we calculate the frequency, in the absence of a signal, that the claimed $Z$-value is exceeded for an ensemble of experiments with the chosen $\mu_b$ and $\tau$ or $f$. This is then converted to the “true” $Z$-value, $Z_{\text{true}}$. We then plot what we call $\Delta Z = Z_{\text{true}} - Z_{\text{claim}}$; then the coverage is easily checked by looking for deviations above or below $\Delta Z = 0$, corresponding to over or under-coverage, respectively.

We present here just four sample plots showing the results of these scans for a claimed $Z$-value of 5 (i.e. a claimed $p$-value of $2.87 \times 10^{-7}$). One pair of plots applying $Z_{Bi}$ and $Z_N$ to the on/off problem is shown in Fig. 1, and another pair applying the two recipes to the Gaussian-mean background problem for absolute $\sigma_b$ is shown in Fig. 2. Plots for other combinations of problems, recipes, claimed $Z$-values, and for larger values of $\mu_b$ and $\tau$ or $f$ are in Ref. [1].

For the on/off experiments analyzed using the $Z_{Bi}$ recipe (Fig. 1 (left)), $Z_{\text{true}} \geq Z_{\text{claim}}$ everywhere, as expected. Using the $Z_N$ recipe (Fig. 1 (right)), one gets under-coverage as severe as two units of $\Delta Z$ for some regions in the plot. This agrees with the result of Cranmer [6], who (using a Monte Carlo coverage calculation method) finds that for the ensemble of experiments with $\mu_b = 100$ and $\tau = 1$ using $Z_N$ for the on/off problem under-covers for $Z_{\text{claim}} = 5$, obtaining $Z_{\text{true}} = 4.2$.

There is significant over-coverage for small values of $n$, as seen in the lower left corners of the plots; these the discreteness issues come into play as mentioned above. We choose to leave blank those regions in the plot where a $p$-value less than $\sim 10^{-15}$ is obtained and the precise calculation of $Z$ breaks down due to numerical precision limitations.
Fig. 1: For the on/off problem analyzed using the $Z_{Bi}$ (left) and $Z_N$ (right) recipes, for each fixed value of $\tau$ and $\mu_b$, the plot indicates the calculated $Z_{true} - Z_{claim}$ for the ensemble of experiments quoting $Z_{claim} \geq 5$. The lower left corner is devoid of entries due to machine round-off, as described in Ref. [1].

Fig. 2: For the Gaussian-mean background problem with exactly known $\sigma_b$, analyzed using the $Z_{Bi}$ (left) and $Z_N$ (right) recipes, for each fixed value of $f = \sigma_b/\mu_b$ and $\mu_b$, the plot indicates the calculated $Z_{true} - Z_{claim}$ for the ensemble of experiments quoting $Z_{claim} \geq 5$. The upper left corner of the left plot is again devoid of entries due to machine round-off, as described in Ref. [1].
For the Gaussian-mean background experiments with exactly known $\sigma_b$, analyzed with $Z_{Bi}$ (Fig. 2 (left)), there is over-coverage everywhere, and by a large amount for increasing values of $f = \sigma_b/\mu_b$. Using $Z_N$, one runs into under-coverage for increasing $f$ and $\mu_b$. This under-coverage can be even more severe for other choices of $Z_{claim}$ and values of $\mu_b$ and $f$, as seen in the full set of plots in Ref. [1].

For small values of $f$ and larger values of $\mu_b$, using the correspondence Eqn. 3 to approximate the Gaussian-mean background problem as Poisson for calculating $Z_{Bi}$ leads to numerical difficulties as explained in Appendix B of Ref. [1]; therefore we leave the upper left region of the left plot in Fig. 2 blank.

**Recommendations.** For the on/off problem $Z_{Bi} = Z_I$ avoids under-coverage by construction, but can be quite conservative for small numbers of events; we recommend $Z_{Bi}$ for general use in this problem. One may wish to use less conservative tests, either ones constructed directly for the ratio of Poisson means and never under-cover, or (as our referee Nancy Reid suggested) less conservative approximate methods for the binomial problem, such as mid p-values.

For the Gaussian-mean background problem, $Z_{Bi}$ works as well as or better than $Z_N$ in much of the space, but in this implementation there are numerical issues for very small uncertainties on a large mean background. Since neither $Z_{Bi}$ nor $Z_N$ has coverage built in by construction for the Gaussian-mean background problem, one should check the coverage where used.

Of course, it is of interest to extend the studies to other recipes and more complex problems, as previously begun by Tegenfeldt and Conrad [7], and by Rolke, Lopez, and Conrad [8]. For example, we have not yet considered the uncertainty on $\tau$.

We thank Kyle Cranmer and James Linnemann for numerous enlightening discussions. This work was partially supported by the U.S. Department of Energy.

**References**


