Glen Cowan
RHUL Physics
28 June 2011

Exercises on discovery and exclusion significance with statistical uncertainty in the background estimate

This note describes the problem of a simple counting experiment where one observes n events in a search region assumed to follow a Poisson distribution with a mean $\mu s + b$, where s is the expected number of events from the nominal signal model (here taken as known), b is the total expected number of background events and μ is rate parameter that we want to test. The background is treated as a sum of components each constrained by a control measurement. The procedure below for obtain the discovery and exclusion signficance values is based on the large-sample formulae described in Ref. [1].

Suppose n events are selected in a search region where both signal and background could be present. The expectation value of n can be written

$$E[n] = \mu s + b_{\rm tot} \ , \tag{1}$$

where s is the expected number from signal and $b_{\rm tot}$ is the expected total background (i.e., from all sources). Here μ is a strength parameter defined such that $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ gives the nominal signal rate plus background.

Suppose that $b_{\rm tot}$ consists of N components, i.e.,

$$b_{\rm tot} = \sum_{i=1}^{N} b_i \ . \tag{2}$$

To estimate the expected number of events from background component i, we construct N control regions, where we make subsidiary measurement m_i modeled as following a Poisson distribution with expectation value

$$E[m_i] = \tau_i b_i \ . \tag{3}$$

Here τ_i is a scale factor that relates the mean number of events that contribute to n (the primary measurement), to that of the ith subsidiary measurement. That is, τ_i is effectively the ratio of the sizes of the control to signal regions. Here we will assume that the τ_i can be determined with negligible uncertainty.

The likelihood function for the parameters μ and $\mathbf{b} = (b_1, \ldots, b_N)$ is the product of Poisson probabilities:

$$L(\mu, \mathbf{b}) = \frac{(\mu s + b_{\rm tot})^n}{n!} e^{-(\mu s + b_{\rm tot})} \prod_{i=1}^{N} \frac{(\tau_i b_i)^{m_i}}{m_i!} e^{-\tau_i b_i} \ . \tag{4}$$

Here μ is the parameter of interest; the components of $\mathbf{b}$ are nuisance parameters.

To test a hypothesized value of μ, one computes the profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\mathbf{b}}})}{L(\hat{\mu}, \hat{\mathbf{b}})} \tag{5}$$

where the double-hat notation refers to the conditional maximum-likelihood estimators (MLEs) for the given value of μ, and the single hats denote the unconditional MLEs.

For purposes of establishing a one-sided upper limit, we define the test statistic

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \ , \\ 0 & \hat{\mu} > \mu \ , \end{cases} \tag{6}$$

and for purposes of testing the background-only hypothesis for discovery, we define

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} > 0 \ , \\ 0 & \hat{\mu} \leq 0 \ , \end{cases} \tag{7}$$

The ratio $\lambda(\mu)$ is expected to be close to unity (i.e., q_μ is near zero) if the data are in good agreement with the hypothesized value of μ.

Suppose the data results in a value of $q_\mu = q_{\rm obs}$. The level of agreement between the data and hypothesized μ is given by the p-value,

$$p = \int_{q_{\rm obs}}^{\infty} f(q_\mu|\mu)\, dq_\mu \ , \tag{8}$$

where $f(q_\mu|\mu)$ is the sampling distribution of q_μ under the assumption of μ.

One can define the significance corresponding to a given p-value as the number of standard deviations Z at which a Gaussian random variable of zero mean would give a one-sided tail area equal to p. That is, the significance Z is related to the p-value by

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z) \ , \tag{9}$$

where Φ is the cumulative distribution for the standard (zero mean, unit variance) Gaussian. Equivalently one has

$$Z = \Phi^{-1}(1-p) \ , \tag{10}$$

where Φ^{-1} is the quantile of the standard Gaussian (inverse of the cumulative distribution).

As shown in Ref. [1], in the large-sample limit, the significance with which one would reject a value of μ approaches the asymptotic form

$$Z_\mu = \sqrt{q_\mu} \ , \tag{11}$$

and similarly the significance with which one would reject $\mu = 0$ is

$$Z_0 = \sqrt{q_0} \ . \tag{12}$$

To calculate the value of q_μ that one would obtain from a set of data values n and $m_1, \ldots, m_N$, one needs the unconditional estimators $\hat{\mu}$ and $\hat{\mathbf{b}}$, and the conditional MLEs $\hat{\hat{\mathbf{b}}}$, i.e., the values of $\mathbf{b}$ that maximize the likelihood for the specified value of μ. The value of n itself is obtained from the real data. In cases with more than one background component, it is easiest to solve for the required quantities numerically.

As an example consider a search with two background components as illustrated in Table 1. Using these numbers $n = 23$ for the number of events found gives $Z_0 = 3.00$.

Table 1: Number of expected background events m_i observed in the control measurements and the corresponding scale factors τ_i.

m_i	τ_i
8	2.0
3	1.0
1	0.5
0	1.0

As suggested exercises, one may use the program `SigCalc`, available on the course website, to investigate how the discovery significance would change if the background measurements were different, or if the scale factors τ_i were changed. In the default version of the program, he values needed are entered in a text file, and example of which called `inputFile.txt` can be found together with the `SigCalc` code. The format should be clear from looking at the file, and one may then change the values or the number of background components.

Recall that the estimated background value $\hat{b}_i = m_i/\tau_i$, so one could, for example investigate what happens if the first background component were characterized by $m_1 = 40$, $\tau_1 = 10$, versus, say, $m_1 = 1$, $\tau_1 = 0.25$. In both cases, as well as with the numbers given in the table above, the number of background events for the first background component is predicted to be 4.0. But with large τ_1, the statistical precisiion with which this background is estimated is greater, and thus one expects to have a more sensitive analysis.

Similarly, for the fourth background component in the table above, there were zero events observed in the control measurement. Suppose, however, that the corresponding τ_4 value had been equal to 0.1 (poorly constrained background) or 10.0 (well constrained background). One may investigate what these and similar changes have on the discovery significance.

One may extend the program SigCalc in any number of ways. For example, instead of finding the significance for observed values of n and $m_1, \ldots, m_M$, one may generate these values from a Monte Carlo model under assumption of specified values of μ and b_i. For each simulated experiment one can then find the significance, and by simulating the experiment a number of times one can find the median (or mean) of this distribution, which is then the expected significance under assumption of the model used to generate the data.

The program also provides the p-value of a hypothesized value of μ. If this value is found less than 5%, one would say that the value is excluded at 95% confidence level. One could therefore include a loop in the program to find the p-value for a range of hypothesized values of μ, and thus to find that value of μ for which $p_\mu = 0.05$. As the p-value given in the program is based on a one-sided test, this represents the upper limit at 95% confidence level.

As a more ambitions extention, one may also insert histograms into the program to record the exact distribution of the test statistic being used, and use these, rather than the asymptotic formulae, to determine the p-values of the hypothesis being tested.

References

[1] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C 71 (2011) 1-19; arXiv:1007.1727.