Glen Cowan
Eilam Gross
24 September, 2008

Comparison of significance from profile and integrated likelihoods

A recent Statistics Forum presentation by Dan Tovey [1] studied the problem of estimating the discovery significance based on a number of observed events n sampled from a Poisson distribution, with Gaussian uncertainty in the background. In Section 1 we treat the problem using the profile likelihood, and in Section 2 we use an integrated likelihood.

1 Statement of the problem and solution with profile likelihood

Using a slightly different notation than [1], suppose n is Poisson distributed with mean value $\mu s + b$. Here s is the expected number of signal events, b is the expected number of background events, and μ is a strength parameter defined such that $\mu = 0$ is the background-only hypothesis and $\mu = 1$ is the hypothesis of background plus signal. Here suppose s is known with no uncertainty.

Suppose our only information about b comes from a measured value m assumed to be Gaussian distributed about b with standard deviation σ. Here σ is known and b is treated as a free parameter.

The unknown parameters are thus μ and b and the measurements are n and m. The likelihood function is

$$L(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \, \frac{1}{\sqrt{2\pi}\sigma} e^{-(m-b)^2/2\sigma^2} \ , \tag{1}$$

or equivalently the log-likelihood is

$$\ln L(\mu, b) = n \ln(\mu s + b) - (\mu s + b) - \frac{(m-b)^2}{2\sigma^2} + C \ , \tag{2}$$

where C is a constant that can be dropped.

To test a hypothesized value of the strength parameter μ we form the *profile likelihood ratio*,

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{b}})}{L(\hat{\mu}, \hat{b})} \ , \tag{3}$$

where $\hat{\mu}$ and $\hat{b}$ are the (unconditional) maximum-likelihood estimators (MLEs) and $\hat{\hat{b}}$ is the condition MLE for b for a given μ. Equivalently we use the logarithmic variable

$$q_\mu = -2 \ln \lambda(\mu) \ , \tag{4}$$

which is defined such that larger value of q_μ correspond to increasing incompatibility between the data and the hypothesized μ.

Here for the model to make sense physically we consider only non-negative values of μ and b. Constraining all of the estimators for μ and b to be non-negative one finds

$$\hat{\mu} = \begin{cases} \frac{n-b}{s} & n \geq m \;, \\ 0 & \text{otherwise} \;, \end{cases} \tag{5}$$

and

$$\hat{b} = \begin{cases} m & n \geq m \;, \\ \frac{1}{2}\left(m - \sigma^2 + \sqrt{(m-\sigma^2)^2 + 4n\sigma^2}\right) & \text{otherwise} \;, \end{cases} \tag{6}$$

As we will consider only the discovery significance we are testing the hypothesis $\mu = 0$, and we need therefore $\hat{\hat{b}}$ given $\mu = 0$, which is

$$\hat{\hat{b}}(\mu = 0) = \frac{1}{2}\left(m - \sigma^2 + \sqrt{(m-\sigma^2)^2 + 4n\sigma^2}\right) \;. \tag{7}$$

To quantify the level of discrepancy between observations n and m and a hypothesized value of μ, we compute the probability, assuming μ, to find q_μ greater than or equal to the value found with the real data. This probability can be computed by Monte Carlo, or alternatively one can exploit the fact that for sufficiently large n, the sampling distribution of q_μ under the assumption of μ approaches a limiting form related to the chi-square distribution. Specifically, it is a superposition of a delta function at zero and a chi-square distribution for one degree of freedom, with each term carrying a weight of one half (see, e.g., [2]),

$$f(q_\mu|\mu) = \tfrac{1}{2}\delta(q_\mu) + \tfrac{1}{2} f_{\chi_1^2}(q_\mu) \;. \tag{8}$$

The p-value is the probability to observe a value of q_μ greater than or equal to that found with the data,

$$p = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu)\, dq_\mu \;. \tag{9}$$

The significance Z is then given by (see, e.g., [2]),

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

where Φ^{-1} is the quantile (inverse of the cumulative distribution) of the standard Gaussian. For the sampling distribution given by (8), one can show that the significance is given by the simple formula

$$Z = \sqrt{q_{\mu,\text{obs}}} \;. \tag{11}$$

Following [1] we have worked out an example with $b = 3.1$ and $\sigma = 0.5$, for a test of the hypothesis $\mu = 0$. By repeating 10^9 experiments in a toy Monte Carlo program, the sampling distribution of q_0 was obtained, and is shown in Fig. 1(a). One minus the corresponding

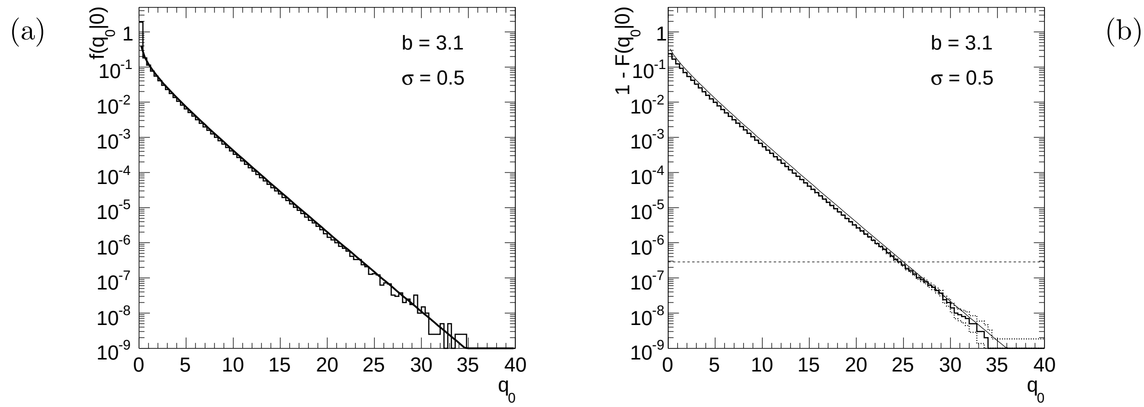

Figure 1: (a) The distribution of the test statistic q_0 under the $\mu = 0$ hypothesis; (b) one minus the corresponding cumulative distribution. The curve shows the asymptotic distribution (8). The dotted curves indicate a 68.3% central confidence interval. The horizontal line is at 2.87×10^{-7} (the 5σ discovery threshold).

cumulative distribution is shown in Fig. 1(b). One can see that the asymptotic approximation (8) is an excellent approximation event out to the level of a 5σ discovery, corresponding to $q_0 = 25$.

We can estimate the median significance given data values n and m simply by setting them equal to their expectation values, i.e., $n = s$ and $m = b$, which results in an observed value of $q_0 = 26.41$. From Monte Carlo the probability to find q_0 greater than or equal to this value gives a $p = 1.27 \times 10^{-7}$, corresponding to a significance $Z = 5.15$. Alternatively one can use the asymptotic formula (11) which gives very close to the same value: $Z = 5.14$.

2 Solution with integrated likelihood

An alternative model for the problem above is to say that the background parameter b has an uncertainty characterized by a prior probability density $\pi(b)$, which we can take to be a Gaussian of standard deviation σ, centred about a true (unknown) value b_0. The probability to find n events is taken to be a Poisson distribution with a mean of $\mu s + b$ where b is sampled from $\pi(b)$, i.e.,

$$P(n|\mu, s, b_0) = \int \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \pi(b) \, db \,. \tag{12}$$

One can use the probability (12) to define a ratio of integrated likelihoods, which could then be used in a manner similar to the profile likelihood ratio above. Alternatively we can find the p-value of the $\mu = 0$ hypothesis simply by computing the probability, under the assumption of $\mu = 0$, of finding a number of events greater than or equal to the number found in the data, i.e.,

$$p = \sum_{n=n_{\rm obs}}^{\infty} P(n|\mu, s, b_0) \,. \tag{13}$$

The p-value can be computed using Monte Carlo by sampling b from $\pi(b)$ and then generating n from the a Poisson distribution with mean $\mu s + b$ (i.e., using a mean b if one is testing the hypothesis $\mu = 0$).

If we suppose as before a data value $n_{\text{obs}} = 17$, and a mean and standard deviation for the Gaussian prior $\pi(b)$ of $b_0 = 3.1$ and $\sigma = 0.5$, then we find $p = 1.98 \times 10^{-7}$ corresponding to a significance of $Z = 5.07$.

In the case of the p-value computed from equation (13), one computes the probability $P(n \geq n_{\text{obs}})$. Because n is discrete, this means that a generated experiment is always counted as having equal compatibility with the hypothesis when $n = n_{\text{obs}}$ (and thus it counts towards the probability of the p-value). This is in contrast to the profile-likelihood method, where the level of compatibility was based on both the discrete n and the continuous m. In that case, depending on the fluctuations of m, half of the events with $n = n_{\text{obs}}$ would be counted as having equal or greater compatibility and half with less.

One can check this using the integrated likelihood by computing the number of events with equal or lesser compatibility in the following way. If $n > n_{\text{obs}}$, the event counts as one event, and if $n = n_{\text{obs}}$ it counts as one half. The fraction of events with equal or lesser compatibility determined in this way gives a p-value of $p = 1.20 \times 10^{-7}$ ($Z = 5.16$), almost exactly the same as when using the profile likelihood.

References

[1] D. Tovey, *Profile likelihood: p-values when background is well-understood*, ATLAS Statistics Forum Meeting, 24 September, 2008.

[2] The ATLAS Collaboration, *Statistical combination of several important Standard Model Higgs boson search channels*, in Expected Performance of the ATLAS Experiment, Detector, Trigger, and Physics, CERN-OPEN-2008-020, Geneva, 2008, to appear.