Bayesian Approach to Poisson Statistics of Small Numbers
SRL internal report #102

Richard Selesnick, 2/13/92

This report is intended to illustrate the use of Bayesian statistics and to point out a discrepancy with the results obtained by M. Israel (SRL internal report #4, 1968), and elaborated on by N. Gehrels (Ap. J., 303, 336, 1986), for confidence limits in the Poisson statistics of small numbers. I have discussed the discrepancy with Neil Gehrels and we concluded that it is the result of using different theoretical definitions of the confidence limits. The approach taken by Israel and Gehrels was to define the 1 sigma confidence limits such that, if the experiment was repeated many times, at least 68% of the results would be within these limits. Then most situations lead to more than 68% of the results being within these limits. Their approach is therefore useful if one wants to place conservative error bars on a data point. If instead one is using the data point in, for example, a least squares fit to an energy spectrum, then the conservative error bars are inappropriate because they can mislead one into thinking that the model is "within the error bars" when it actually may not be. That is, the value of χ^2 would be too small. The Bayesian approach is more useful in this situation, as it should give a more accurate value of χ^2.

Note that the controversy arises only because we are applying concepts appropriate to Gaussian probability distributions. For example, we often do a least-squares fit by minimizing the Gaussian χ^2 function

$$\chi^2 = \sum_i \left(\frac{\tilde{n}_i - n_i}{\sigma_i} \right)^2 \tag{1}$$

where $\tilde{n}_i$ is the model prediction of the ith data point n_i with standard error σ_i. In fact, it may not be much more difficult to minimize the Poisson likelihood χ^2 function, derived directly from the Poisson distribution (Baker and Cousins, Nucl. Instrum. Methods Phys. Res., 221, 437, 1984),

$$\chi_P^2 = 2\sum_i \left[\tilde{n}_i - n_i + n_i \ln\left(\frac{n_i}{\tilde{n}_i} \right) \right] \tag{2}$$

This makes the proper statistical weighting and therefore does not contain σ_i. However, the Gaussian least-squares approach is often the the most convenient, so it is useful to derive the error bars for that case, and they are given in equation 15 below.

I will first derive the confidence limits associated with estimating the expected number of counts from a Poisson distribution when n counts are observed. The Poisson distribution function is

$$f(p,d) = \frac{e^{-p} p^d}{d!} \tag{3}$$

where the data, d, is the observed number of counts, and the parameter, p, is the mean of the distribution. The Bayesian formula for the probability density function of the parameter as a result of the measurement is (for a general discussion see Tarantola, *Inverse Problem Theory*, Elsevier, 1987)

$$\sigma(p) = \int \rho(d)\theta(d|p)\mathrm{d}d \tag{4}$$

I assume that the experiment counts the number n without any error, so that the *a priori* density function for the data is

$$\rho(d) = \delta(d-n) \tag{5}$$

The conditional density function for measuring d given p is just

$$\theta(d \mid p) = f(p, d) \tag{6}$$

Then evaluating the integral in (2)

$$\sigma(p) = f(p, n) \tag{7}$$

This result may be obvious, but the method becomes very useful in more complicated situations. In any case, we can now derive any properties of the density function for p. For example, the maximum likelihood point (the mode of the distribution)

$$p_{ML} = n \tag{8}$$

the mean

$$\bar{p} = n+1 \tag{9}$$

and the standard deviation

$$\sigma_p = (n+1)^{1/2} \tag{10}$$

Note that the mean, $n+1$, is generally a better estimate of p than the maximum likelihood point, n, because of the asymmetry of the distribution function.

The 1 sigma upper (lower) confidence limit is defined, by analogy with the Gaussian distribution, as the value of p such that the probability of a smaller (larger) value is 84.13%. In general, an upper confidence limit of CL_u is given by $p = p_u$ if

$$1-CL_u = \int_{p_u}^{\infty} \sigma(p)\mathrm{d}p \tag{11}$$

and a lower confidence limit of CL_l is given by $p = p_l$ if

$$CL_l = \int_{p_l}^{\infty} \sigma(p)\mathrm{d}p \tag{12}$$

Substituting $\sigma = f$ from (7) and (3) and looking up the integrals in a table it is easy to show that

$$CL_u = 1-\sum_{i=0}^{n} f(p_u, i) \tag{13}$$

$$CL_l = \sum_{i=0}^{n} f(p_l, i) \tag{14}$$

Only the equation for CL_l differs from Israel's, where the summation extends only to $n-1$. By setting $CL_u = CL_l = 0.5$, $p_u = p_l$, and adding together (13) and (14), it is easy to see that the upper and lower 50% confidence limits are equal. Israel's results do not have this property. The derivation of approximate formulae for small n follows Israel's and the results are

$$p = \bar{p} \pm (n + 3/4)^{1/2} \tag{15}$$

where $\bar{p} = n+1$ is given by (9). I would recommend this as the best estimator for p and its 1 sigma confidence region. It is reasonably accurate even down to $n = 0$, although the exact equations (13) and (14) can be solved analytically in that case to give $p = 1$, $p_u = -\ln(1-CL_u) = 1.841$ and $p_l = -\ln(CL_l) = 0.173$ for $CL_u = CL_l = 0.8413$ (it may be preferable to use only the upper limit for $n = 0$).

As a second and slightly more complicated example I consider the case of the n counts being observed from a combination of a source and a background. Both obey Poisson statistics but the source again has an unknown mean p that we wish to determine, while the background distribution has a known mean, b. The only difference from the above derivation is that (6) must be replaced by

$$\theta(d \mid p) = \sum_{i+j=d} f(p,i) f(b,j) \tag{16}$$

where the summation extends over all combinations of source and background counts, i and j, that add up to the data, d. Plugging into (4) leads to

$$\sigma(p) = e^{-(p+b)} \sum_{j=0}^{n} \frac{p^{n-j} b^j}{(n-j)! j!} = f(p+b,n) \tag{17}$$

where the last step follows from the binomial expansion. Again, the result seems obvious. Clearly, in the large n limit the estimate of p is $p = n-b \pm n^{1/2}$. The derivation of a small n approximation becomes much more complicated than before. However, if you allow the possibility of negative p then the previous derivation holds and equation 15 can be used with $\bar{p} = n-b+1$.