-ln$\mathcal{L}$ for Poisson

Harry Nelson

November 12, 1998

The Poisson distribution for to see r events given a mean of μ is (see 1996 RPP Table 27.1 p. 157):

$$f(r;\mu) = \frac{\mu^r e^{-\mu}}{r!}$$

When doing a binned fit, the mean number expected from the fit in bin i is sometimes labelled $N_i^{\rm th}$, as on the 1996 RPP Eqn. 28.9 p. 160, but we will simplify the notation for $N_i^{\rm th}$ to μ; similarly, the number of observed events in bin i is sometimes labelled $N_i^{\rm obs}$, as on the 1996 RPP Eqn. 28.9 p. 160, but we will simplify the notation for $N_i^{\rm obs}$ to r.

The increment in $-\ln\mathcal{L}$ from bin i, where r events are observed, and for which a mean of μ events are expected, is:

$$-\ln f_i(r;\mu) = [\mu - r\ln\mu] \ + \ \ln r! \tag{1}$$

One quick result follows from this equation: the most probable value of r is μ. We want to make the value of the $-\ln\mathcal{L}$ zero at the most probable r, and to do, we can subtract off $-\ln f_i(\mu;\mu)$, to get an expression for $\Delta\chi^2$ from bin i:

$$\Delta\chi^2 = -2\ln\left[\frac{f_i(r;\mu)}{f_i(\mu;\mu)}\right] = \ln\left[\frac{\mu^\mu}{\mu^r}\right] + \ln\left[\frac{r!}{\mu!}\right] \tag{2}$$

In principle, Eqn. 1 can be used to perform a fit, and Eqn. 2 can be used to get the goodness of fit. However, the computation of $\ln r!$ exactly can become onerous, if r gets larger than a few.

Use of an approximation for $\ln r!$, of which the most famous is Stirling's approximation (Eqn. 3), leads to:

$$\begin{aligned}
\ln r! &\simeq r[\ln r - 1] \qquad\qquad \text{so,} && (3)\\
-\ln f_i(r;\mu) &\simeq [\mu - r] + r\ln\frac{r}{\mu} &&\\
&= \frac{1}{2}\Delta\chi^2 &&\\
\text{or,}\qquad \Delta\chi^2 &\simeq 2[\mu - r] + 2r\ln\frac{r}{\mu} && (4)
\end{aligned}$$

The relationship Eqn. 4 appears as Eqn. 28.9 on p. 160 of the 1996 RPP, as an increment in χ^2 that is improved over the more approximate $\Delta\chi^2 = (r-\mu)^2/\mu$.

However, Eqn. 4 is rather inaccurate for r and $\mu < 10$, a situation that is common.

To address this inaccuracy, we first note the further terms in the expansion of Stirling:

$$\ln r! \simeq \frac{1}{2}\ln 2\pi - r + \left[r + \frac{1}{2}\right]\ln r + \frac{1}{12r}\left[1 - \frac{1}{30r^2}\left[1 - \frac{2}{7r^2}\right]\right] \tag{5}$$

The worst that Eqn. 5 does is for $r = 1$, where it's error is about 7 parts per million. Dropping the last term raises the error to about 29 parts per million, and keeping only through $1/12n$ raises the error to 0.5 parts per thousand; omission of the $1/12n$ raises the error to 6%. Keeping through $1/12n$ seems reasonable.

So, naively,

$$-\ln \overline{f}_i(r;\mu) \simeq \frac{1}{2}\ln 2\pi + [\mu - r] + r\ln\frac{r}{\mu} + \frac{1}{2}\ln r + \frac{1}{12r} \tag{6}$$

However, the value of Eqn. 6, evaluated at the most probably $r = \mu$, is non-zero. A remedy is to subtract off the value of Eqn. 6, with r evaluated at most probable, μ, yielding:

$$\begin{aligned} -\ln f_i(r;\mu) &\simeq [\mu - r] + \left[r + \frac{1}{2}\right]\ln\frac{r}{\mu} + \frac{1}{12}\left[\frac{1}{r} - \frac{1}{\mu}\right] \\ \Delta\chi^2 &\simeq 2[\mu - r] + 2\left[r + \frac{1}{2}\right]\ln\frac{r}{\mu} + \frac{1}{6}\left[\frac{1}{r} - \frac{1}{\mu}\right] \end{aligned} \tag{7}$$

The expression 7 differs from Eq. 4 by two new factors, a $\ln(r/\mu)$ and $(1/6)(1/r - 1/\mu)$; the first order difference, in $(r-\mu)$ vanishes, when one averages over many samples. The coefficient of the second order term, $(r-\mu)^2$, contributed by the two new factors is $(-1/2 + 1/(6\mu))$, which will be negative, unless μ is smaller than $1/3$. Thus, typically, inclusion of the new terms should *lower* χ^2.

Numerical studies show that this approximation is a significant improvement on Eqn. 4 for r and $\mu < 1$. However, when $\mu << 1$, the approximation has problems. An effective remedy is to never allow the μ used in Eqn. 7 to fall below 0.05, and to never allow $\Delta\chi^2$ to be negative.

The numerical studies also reveal another deficiency in Eqn. 7: when $r = 0$. A reasonable prescription appears to be to, for $r=0$, to set:

$$\Delta\chi^2 = 2\mu - \min(1.8379 + \ln\mu + \frac{1}{6\mu}, 2\mu). \tag{8}$$

Here, $1.8379 = \ln(2\pi)$, and $\ln(2\pi) + \ln\mu + \frac{1}{6\mu}$ is twice the part of Eq. 6 that is non-zero at most probable. Then, the leading term 2μ is $-2\ln f_1(0,\mu)$, from Eq. 1. When μ gets small, $\ln(2\pi) + \ln\mu + \frac{1}{6\mu}$ gets unreasonably large; it should never be allowed to exceed 2μ, which is what the above equation does.