Maximum Likelihood Example

J. G. Heinrich

July 19, 1995

Suppose that an experiment observes n_1 events of type 1 and n_2 events of type 2. We will denote the underlying (actual) rates by μ_1 and μ_2, so that the probability of what happened is

$$P(n_1, n_2) = \frac{\mu_1^{n_1} e^{-\mu_1}}{n_1!} \frac{\mu_2^{n_2} e^{-\mu_2}}{n_2!}$$

where we have assumed independent Poisson probabilities. The rates μ_1 and μ_2 are unknown; the method of maximum likelihood can be used to provide estimates of their values and also the uncertainties of those estimates.

It may be that we are not interested in the estimating μ_1 and μ_2 separately, but rather (for example) $\mu_1/(\mu_1 + \mu_2)$. We will define $p = \mu_1/(\mu_1 + \mu_2)$ and $\mu = \mu_1 + \mu_2$; substituting the inverted relations $\mu_1 = p\mu$ and $\mu_2 = (1-p)\mu$ into the expression for $P(n_1, n_2)$ results in

$$P(n_1, n_2) = \frac{\mu^{n_1+n_2} e^{-\mu}}{(n_1+n_2)!} \frac{(n_1+n_2)!\, p^{n_1}(1-p)^{n_2}}{n_1!\, n_2!}$$

which is a Poisson probability times a Binomial probability.

To estimate μ and p using the method of maximum likelihood we first define the likelihood $\mathcal{L}$:

$$\mathcal{L}(\mu, p) = \frac{1}{c} \mu^{n_1+n_2} e^{-\mu} p^{n_1} (1-p)^{n_2}$$

where the normalization factor c is given by:

$$c = \int_0^\infty \mu^{n_1+n_2} e^{-\mu} d\mu \int_0^1 p^{n_1}(1-p)^{n_2} dp$$

Note that $\mathcal{L}$ is a function of μ and p; n_1 and n_2 are constants and c is also a constant. The negative log likelihood is then given by

$$\ell(\mu, p) = \mu - (n_1+n_2)\log\mu - n_1 \log p - n_2 \log(1-p)$$

where we have dropped the $\log c$ term—an overall additive constant term in $\ell(\mu, p)$ has no effect because we will only be interested in derivatives (or differences) of $\ell(\mu, p)$.

The first derivatives are:

$$\frac{\partial \ell}{\partial \mu} = 1 - \frac{n_1+n_2}{\mu} \qquad \frac{\partial \ell}{\partial p} = \frac{n_2}{1-p} - \frac{n_1}{p}$$

We want to find the values of μ and p for which $\ell(\mu, p)$ is a minimum; these values (we will denote them as μ_0 and p_0) are the estimates of μ and p provided by the maximum likelihood method. The derivatives are zero when $\mu_0 = n_1 + n_2$ and

$p_0 = n_1/(n_1+n_2)$, so these are the best estimates of μ and p (we will see that the second derivatives are positive in the next step, *i.e.* this solution is really a minimum and not a maximum).

The uncertainties of our estimates $\mu_0 = n_1+n_2$ and $p_0 = n_1/(n_1+n_2)$ are related to the second derivatives:

$$\frac{\partial^2 \ell}{\partial \mu^2} = \frac{n_1+n_2}{\mu^2} \qquad \frac{\partial \ell^2}{\partial p^2} = \frac{n_1}{p^2} + \frac{n_2}{(1-p)^2} \qquad \frac{\partial^2 \ell}{\partial \mu \partial p} = 0$$

We must evaluate the second derivatives at the minimum:

$$\left.\frac{\partial^2 \ell}{\partial \mu^2}\right|_{\substack{\mu=\mu_0\\ p=p_0}} = \frac{1}{n_1+n_2} = \frac{1}{\mu_0} \qquad \left.\frac{\partial \ell^2}{\partial p^2}\right|_{\substack{\mu=\mu_0\\ p=p_0}} = \frac{(n_1+n_2)^3}{n_1 n_2} = \frac{\mu_0}{p_0(1-p_0)} \qquad \left.\frac{\partial^2 \ell}{\partial \mu \partial p}\right|_{\substack{\mu=\mu_0\\ p=p_0}} = 0$$

The uncertainties are then given by

$$\sigma_{\mu_0} = \left(\left.\frac{\partial^2 \ell}{\partial \mu^2}\right|_{\substack{\mu=\mu_0\\ p=p_0}} \right)^{-\frac{1}{2}} = \sqrt{n_1+n_2} = \sqrt{\mu_0}$$

$$\sigma_{p_0} = \left(\left.\frac{\partial \ell^2}{\partial p^2}\right|_{\substack{\mu=\mu_0\\ p=p_0}} \right)^{-\frac{1}{2}} = \frac{\sqrt{n_1 n_2}}{(n_1+n_2)^{3/2}} = \sqrt{\frac{p_0(1-p_0)}{\mu_0}}$$

The fact that the mixed partial derivative is zero means there is no correlation between μ_0 and p_0 to complicate matters.

Next, suppose we are really not interested in $p = \mu_1/(\mu_1+\mu_2)$, but rather in $r = \mu_1/\mu_2$. We simply use the replacement $p = r/(1+r)$ to get

$$\mathcal{L}(\mu, r) = \frac{1}{c'} \mu^{n_1+n_2} e^{-\mu} r^{n_1} (1+r)^{-(n_1+n_2)}$$

where the normalization factor c' is given by:

$$c' = \int_0^\infty \mu^{n_1+n_2} e^{-\mu} d\mu \int_0^\infty r^{n_1} (1+r)^{-(n_1+n_2)} dr$$

Then the negative log likelihood is given by

$$\ell(\mu, r) = \mu - (n_1+n_2)\log\mu + (n_1+n_2)\log(1+r) - n_1 \log r$$

where the additive constant is dropped. The derivatives are

$$\frac{\partial \ell}{\partial \mu} = 1 - \frac{n_1+n_2}{\mu} \qquad \frac{\partial \ell}{\partial r} = \frac{n_1+n_2}{1+r} - \frac{n_1}{r}$$

which leads to $\mu_0 = n_1+n_2$ and $r_0 = n_1/n_2$. The second derivatives are

$$\frac{\partial^2 \ell}{\partial \mu^2} = \frac{n_1+n_2}{\mu^2} \qquad \frac{\partial \ell^2}{\partial r^2} = \frac{n_1}{r^2} - \frac{n_1+n_2}{(1+r)^2} \qquad \frac{\partial^2 \ell}{\partial \mu \partial r} = 0$$

which leads to the uncertainties

$$\sigma_{\mu_0} = \left(\left. \frac{\partial^2 \ell}{\partial \mu^2} \right|_{\substack{\mu=\mu_0 \\ r=r_0}} \right)^{-\frac{1}{2}} = \sqrt{n_1 + n_2} = \sqrt{\mu_0}$$

$$\sigma_{r_0} = \left(\left. \frac{\partial \ell^2}{\partial r^2} \right|_{\substack{\mu=\mu_0 \\ r=r_0}} \right)^{-\frac{1}{2}} = \frac{\sqrt{n_1(n_1+n_2)}}{n_2^{3/2}} = \sqrt{\frac{r_0(1+r_0)^2}{\mu_0}}$$

There was an alternate method of going from p_0 and σ_{p_0} to r_0 and σ_{r_0} using the relation $r(p) = p/(1-p)$. Because $\ell(\mu, p) = \ell(\mu, r(p))$ we have $r_0 = r(p_0) = n_1/n_2$. The second derivatives are related by

$$\frac{\partial \ell^2}{\partial p^2} = \frac{\partial}{\partial p}\left(\frac{\partial \ell}{\partial r} \frac{\partial r}{\partial p} \right) = \frac{\partial \ell^2}{\partial r^2}\left(\frac{\partial r}{\partial p} \right)^2 + \frac{\partial \ell}{\partial r} \frac{\partial^2 r}{\partial p^2}$$

When $r = r_0$ we have by definition $\partial \ell / \partial r = 0$, so

$$\left. \frac{\partial \ell^2}{\partial p^2} \right|_{\substack{\mu=\mu_0 \\ r=r_0}} = \left. \frac{\partial \ell^2}{\partial r^2} \right|_{\substack{\mu=\mu_0 \\ r=r_0}} \left(\left. \frac{\partial r}{\partial p} \right|_{\substack{\mu=\mu_0 \\ r=r_0}} \right)^2$$

This means that

$$\sigma_{r_0} = \sigma_{p_0} \left| \frac{\partial r}{\partial p} \right|_{\substack{\mu=\mu_0 \\ r=r_0}}$$

It is interesting to note that the expression given above for σ_{r_0} is identical to the expression one calculates in the usual way as the ratio of two gaussians with uncertainties $\sqrt{n_1}$ and $\sqrt{n_2}$. There is, however, more information available from the maximum likelihood method. The errors we calculated above are the parabolic errors; we can also calculate the (asymmetric) Minos errors $\sigma_{r_0}^+$ and $\sigma_{r_0}^-$ which are given by the solutions of

$$\ell(\mu_0, r_0 + \sigma_{r_0}^+) = \ell(\mu_0, r_0 - \sigma_{r_0}^-) = \ell(\mu_0, r_0) + \frac{1}{2}$$

Substituting, we have

$$(n_1 + n_2) \log \frac{n_1 + n_2 \pm n_2 \sigma_{r_0}^\pm}{n_1 + n_2} - n_1 \log \frac{n_1 \pm n_2 \sigma_{r_0}^\pm}{n_1} = \frac{1}{2}$$

which can be solved numerically. For example, if $n_1 = 9$ and $n_2 = 16$ we get $r_0 = 9/16 \simeq 0.563$, $\sigma_{r_0} = 15/64 \simeq 0.234$, $\sigma_{r_0}^+ \simeq 0.285$ and $\sigma_{r_0}^- \simeq 0.196$. Since it is more accurate to use the Minos errors, the best answer for the ratio in this case is $0.563^{+0.285}_{-0.196}$. On the other hand, for the reciprocal case of $n_1 = 16$ and $n_2 = 9$, we get an answer for the ratio of $1.78^{+0.945}_{-0.598}$.