Foundations of Medical Statistics

Statistical Inference 10: Inference with multiple parameters (2)

Aims

The aim of this session is to introduce profile and approximate profile log-likelihoods.

Objectives

The objective is that, at the end of the session, you should undestand what a profile log-likelihood is and how to calculate it. You should further understand that the maximum likelihood estimate of the difference between two parameters, each associated with a model for its own independent data set, is the difference of the maximum likelihood estimate for each data set alone, and be able to apply this result to construct an approximate profile log-likelihood.

10.1 Introduction

We begin with an example: tackling our conditional likelihood problem from Session 9 using profile log-likelihood.

Recall there are two groups — an exposed group and an unexposed group. In the unexposed group k_0 events are observed in p_0 person-years of follow-up, and in the exposed group the equivalent figures are k_1 and p_1.

The model is defined in terms of two random variables, K_0 and K_1, the numbers of events in groups 0 and 1. Each of K_0 and K_1 are assumed to have a Poisson distribution, with λ_0 defined to be the rate parameter in the unexposed group, and λ_1 to be the rate parameter in the exposed group.

This is a situation where there are two parameters, **but** there is only one parameter of real interest, which is the rate ratio, $\theta = \lambda_1/\lambda_0$.

The joint log-likelihood is

$$l(\lambda_0, \lambda_1) = k_0 \log \lambda_0 - \lambda_0 p_0 + k_1 \log \lambda_1 - \lambda_1 p_1$$

By writing $\lambda_1 = \lambda_0 \theta$, l can be rewritten as a function of λ_0 and θ:

$$\begin{aligned} l(\lambda_0, \theta) &= k_0 \log \lambda_0 - \lambda_0 p_0 + k_1 \log(\lambda_0 \theta) - \lambda_0 \theta p_1 \\ &= (k_0 + k_1) \log \lambda_0 + k_1 \log \theta - \lambda_0 p_0 - \lambda_0 \theta p_1 \\ &= k \log \lambda_0 + k_1 \log \theta - \lambda_0 (p_0 + \theta p_1) \end{aligned}$$

where $k = k_0 + k_1$. The value of θ which maximises l now depends on λ_0, so in this context, λ_0 is a **nuisance parameter**.

Deriving a conditional log likelihood depends on removing a nuisance parameter at the stage of determining a probability model – by conditioning the probability (reducing the relevant sample space). The profile likelihood strategy does not condition the probability: it removes the nuisance parameter after the full log-likelihood has been derived.

This strategy removes the nuisance parameter by estimating its value in terms of the parameter of interest. For each value of θ, the value of λ_0 which maximises the log-likelihood is calculated using differentiation. This will be a function of θ and we'll call it $\hat{\lambda}_0(\theta)$. Then the **profile log-likelihood** for θ is

$$l_p(\theta) = l(\hat{\lambda}_0(\theta), \theta)$$

Profile log-likelihoods can (usually) be used in exactly the same way as log-likelihoods (i.e. to find estimates, calculate confidence limits and evaluate hypothesis tests). It is not always possible to calculate an algebraic expression for the profile log-likelihood, but in the case of the rate ratio it is. Treating θ as fixed,

$$\frac{\partial l}{\partial \lambda_0} = \frac{k}{\lambda_0} - (p_0 + \theta p_1)$$

Setting the first derivative equal to 0 gives

$$\hat{\lambda}_0(\theta) = \frac{k}{p_0 + \theta p_1}$$

Substituting this value into the expression above for the log-likelihood gives the profile log-likelihood for θ:

$$l_p(\theta) = k \log k - k \log(p_0 + \theta p_1) + k_1 \log \theta - k$$

so ignoring terms not involving θ:

$$l_p(\theta) = k_1 \log \theta - k \log(p_0 + \theta p_1)$$

which is identical to the conditional likelihood obtained in 9.5.

EXERCISE 10.1.1 *Profile log-likelihood*
Use the profile log-likelihood above to show the following:

1. The MLE of the rate ratio θ, $\hat{\theta}$, is $\frac{k_1/p_1}{k_0/p_0}$, confirming the invariance of the MLE.
2. The standard error of the MLE of $\log \theta$ is $\sqrt{1/k_0 + 1/k_1}$.

10.2 General procedure for profile log-likelihoods and distribution of profile log-likelihood ratio

From the discussion above we can describe a general procedure for obtaining a profile log-likelihood.

1. Divide parameters into those of interest ($\underline{\psi}$) and nuisance parameters ($\underline{\lambda}$).
2. Find the MLE of $\underline{\lambda}$ at each $\underline{\psi}$: denote the function of $\underline{\psi}$ that gives us this by $\hat{\underline{\lambda}}\left(\underline{\psi}\right)$
3. In $l\left(\underline{\psi}, \underline{\lambda}\right)$ replace $\underline{\lambda}$ by $\hat{\underline{\lambda}}\left(\underline{\psi}\right)$ to give l_p, now a function of $\underline{\psi}$ alone.

Suppose we wish to construct confidence intervals and test hypotheses using the distribution of the profile log-likelihood ratio ($pllr$) statistic. What is its distribution?

As for log-likelihood ratio, we can show that, under $H_0: \underline{\psi} = \underline{\psi}_0$

$$-2\, pllr\left(\underline{\psi}_0\right) = -2\left\{l_p\left(\underline{\psi}_0\right) - l_p\left(\hat{\underline{\psi}}\right)\right\}$$

is approximately χ^2, provided the sample size is sufficiently large and the regularity conditions hold (i.e. in particular that the MLE is not at a discontinuity or boundary of the likelihood function). The degrees of freedom of the χ^2 distribution are equal to the number of parameters whose value(s) are specified under H_0. Confusingly, these are sometimes known 'free' parameters, as they are 'free' from being profiled away. So the degrees of freedom is the total number of parameters minus the number we are profiling out.

For example, if the dimension of $\underline{\psi}$ is p and the dimension of $\underline{\lambda}$ is r, and we profile out $\underline{\lambda}$, and our null hypothesis $H_0: \underline{\psi} = \underline{\psi}_0$ only specifies the values of $\underline{\psi}$, then the degrees of freedom is $(p + r) - r = p$.

In the Poisson example above, we have 2 parameters in total (θ and λ_0) and profile out λ_0, so, under $H_0: \theta = \theta^*$:

$$-2\, pllr(\theta^*) \dot{\sim} \chi_1^2$$

10.3 Approximate profile log-likelihoods

For the Poisson rate ratio an algebraic expression for the profile log-likelihood can be derived, but this is not possible in general. However, just as quadratic approximations to log-likelihoods are often useful (Session 5), so are quadratic approximations to profile log-likelihoods.

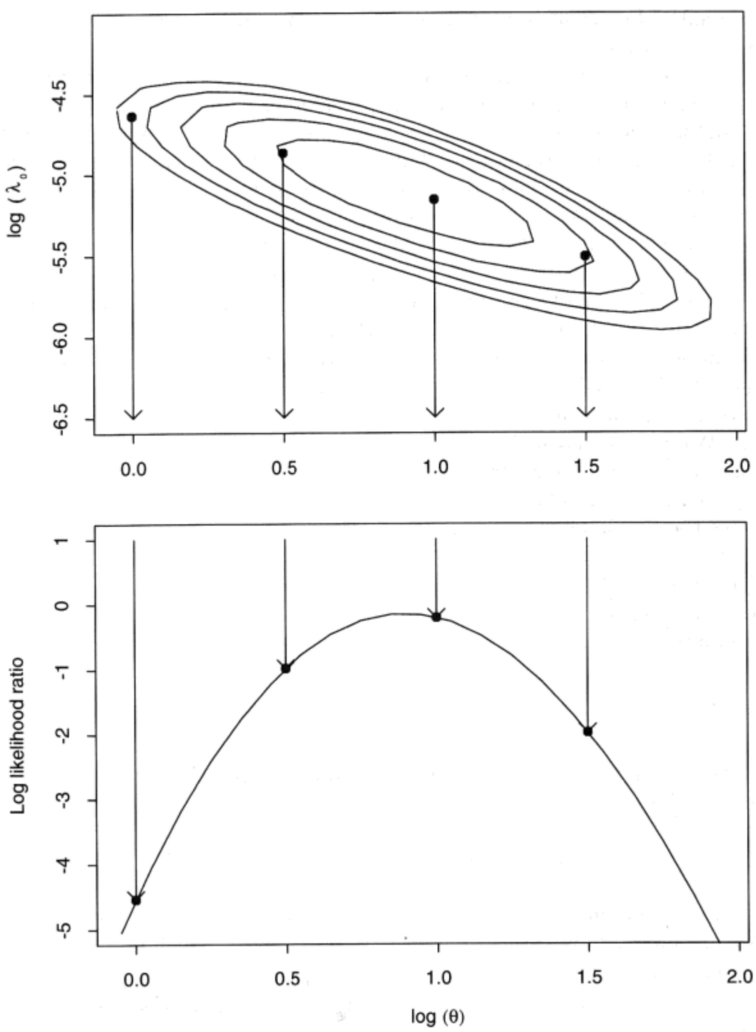

Figure 10.1: Schematic illustration of the construction of a profile log-likelihood from a log-likelihood for two parameters (Clayton & Hills, p. 126).

Example of a general problem

An important general problem is the estimation of the difference between two parameters β_0 and β_1, when these are estimated from two independent sets of data. If the log-likelihood for β_0 has a quadratic approximation defined by the most likely value $\hat{\beta}_0$ and standard error S_0, and the approximation to the log-likelihood for β_1 is defined by $\hat{\beta}_1$ and S_1, then the quadratic approximation to the log-likelihood for $\gamma = \beta_1 - \beta_0$ is defined by $\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}_0$ and $S = \sqrt{S_1^2 + S_0^2}$.

We can therefore calculate Wald tests based on the approximate profile log-likelihood.

In fact, we can go further and derive the following result:

THEOREM 10.3.1 *Suppose we form a function of parameters of the form*

$$\gamma = W_0\beta_0 + W_1\beta_1 + \cdots$$

where the W's are arbitrary constants. If the MLEs of $\beta_0, \beta_1, \ldots$, from independent data sets, are $\hat{\beta}_0, \hat{\beta}_1, \ldots$, with standard errors $S_0, S_1, \ldots$ then the MLE of γ is

$$\hat{\gamma} = W_0\hat{\beta}_0 + W_1\hat{\beta}_1 + \cdots$$

with standard error

$$S = \sqrt{(W_0S_0)^2 + (W_1S_1)^2 + \cdots}$$

For a justification of this see Appendix A.

Example 10.3.1 *Rate ratio $\theta = \lambda_1/\lambda_0$*

Let $\beta_1 = \log\lambda_1$, $\beta_0 = \log\lambda_0$, and $\gamma = \log\theta$. Then $\gamma = \beta_1 - \beta_0$.

Since $\hat{\beta}_0 = \log(k_0/p_0)$ and $\hat{\beta}_1 = \log(k_1/p_1)$ (Section 5.3),

$$\hat{\gamma} = \log\left(\frac{k_1}{p_1}\right) - \log\left(\frac{k_0}{p_0}\right) = \log\left(\frac{k_1/p_1}{k_0/p_0}\right)$$

Or $\hat{\theta} = \frac{k_1/p_1}{k_0/p_0}$ by the invariance property.

Since $S_0 = 1/\sqrt{k_0}$ and $S_1 = 1/\sqrt{k_1}$ (Section 5.3), it follows that

$$S = \sqrt{1/k_0 + 1/k_1}$$

Note: S here is the standard error of the log rate ratio, and it is only obtainable readily as the square root of the sum of two variances because the log rate ratio can be expressed as a sum (of two independent log rates). [Of course we're here considering what is actually a subtraction, as a 'sum'.]

Example 10.3.2 *Rate difference $\theta = \lambda_1 - \lambda_0$*

Here an algebraic expression for the profile log-likelihood cannot be written down. However, using quadratic approximations, $\hat{\delta} = \hat{\lambda}_1 - \hat{\lambda}_0 = k_1/p_1 - k_0/p_0$.

Also, since $S_0 = \sqrt{k_0}/p_0$ and $S_1 = \sqrt{k_1}/p_1$ (Section 5.2), it follows that:

$$S = \sqrt{k_0/p_0^2 + k_1/p_1^2}$$

Note that of course the rate difference is already in the form of a ‘sum’, unlike the situation in the previous example.