13
Likelihoods for the rate ratio

In previous chapters we have introduced the main ideas of probability models in epidemiology and discussed the use of likelihood to provide an estimate, confidence interval or p-value for the parameter of a probability model. Although we have used the joint log likelihood for several parameters our discussion of confidence intervals and p-values has been based on probability models with only a single parameter. We now consider probability models with two or more parameters.

13.1 Comparing rates using the rate ratio

A simple and important problem which involves two parameters is the comparison of two rates, for example for a cohort which was exposed to some environmental factor and an unexposed cohort. The probability model which underlies such a comparison has parameters corresponding to the rates of failure in the two cohorts. We shall use a subscript notation to denote exposure groups and write λ_1 for the rate parameter conditional on exposure, and λ_0 for the rate parameter conditional on non-exposure.

Table 13.1 shows a preliminary tabulation of some data which will be analysed in detail in this and the following chapter.* The data relate subsequent incidence of ischaemic heart disease (IHD) to dietary energy intake. The study cohort consisted of 337 men whose energy intake was assessed by a seven-day weighed dietary survey. The subsequent follow-up was for an average of 13.7 years and yielded 45 new cases of IHD. The table divides this cohort into an exposed group consisting of men whose energy intake was less than 2750 kcals per day, the remaining men being regarded as unexposed. Although it might seem odd to denote the low energy intake group as exposed, this is because low energy intake is a surrogate measure for physical inactivity. Table 13.1 also introduces some algebraic notation: D_0, D_1 for the number of disease events observed in the unexposed and exposed cohorts respectively, and Y_0, Y_1 for the corresponding person-years observation.

*Unpublished data. The study is described by Morris, J.N. *et al.* (1977) *British Medical Journal*, 19 November 1977, **2**, 1307–1314.

Table 13.1. Incidence of ischaemic heart disease by energy intake

	Energy intake	
	$<$ 2750 kcals (exposed)	$\geq$ 2750 kcals (unexposed)
Person years	1857.5 (Y_1)	2768.9 (Y_0)
New cases	28 (D_1)	17 (D_0)
Estimated rate	15.1	6.1
90% interval	(11.1 $\rightarrow$ 20.6)	(4.1 $\rightarrow$ 9.1)

The data from the unexposed group leads to

$$D_0 \log(\lambda_0) - \lambda_0 Y_0 = 17 \log(\lambda_0) - 2\,768.9\lambda_0$$

as the log likelihood for λ_0. The most likely value of λ_0 is the observed incidence rate, $17/2768.9 = 6.1$ per 1000 person-years. The fact that this estimate is based on only 17 observed cases is reflected in the rather wide 90% confidence interval for λ_0 stretching from 4.1 to 9.1 per 1000 person-years. Similarly, the data from the exposed group leads to

$$D_1 \log(\lambda_1) - \lambda_1 Y_1 = 28 \log(\lambda_1) - 1857.5\lambda_1$$

as the log likelihood for λ_1. The most likely value of λ_1 is $28/1857.5 = 15.1$ per 1000 person-years, and the 90% confidence interval stretches from 11.1 to 20.6 per 1000 person-years. The two groups provide independent sets of data, so that the two log likelihoods are added to yield the joint log likelihood

$$17 \log(\lambda_0) - 2\,768.9\lambda_0 + 28 \log(\lambda_1) - 1\,857.5\lambda_1.$$

This is the likelihood for any specified pair of values for the two parameters λ_0 and λ_1. Its maximum value is achieved when these parameters take values equal to the corresponding observed rates — 6.1 and 15.1 per 1000 person-years respectively.

The 90% confidence intervals for the two rates do not overlap and it might seem that the data support the proposition that the two rates are different. In general, however, the degree of overlap of confidence intervals is a poor criterion for comparing rates. If the interval in the high intake group had stretched from, say, 3.0 to 12.0 then it could be argued that, since values of the rate parameter in the range from 11.1 to 12.0 are included in both intervals, the data do not support the idea that the rates are different. The flaw in this argument is that this range is at the extreme of both ranges; the support for the proposition that the rates are similar requires two rather poorly supported propositions to hold simultaneously.

The way to approach such problems is to reparametrize the model in such a way that one of the new parameters makes a comparison. The usual comparison parameter for two rates is the *rate ratio*, which we shall denote by the Greek letter θ. Since $\theta = \lambda_1/\lambda_0$, the rate in the exposed cohort may be written as $\theta\lambda_0$ instead of λ_1 and our model can be written in terms of the parameters (θ, λ_0) instead of (λ_1, λ_0).

The log likelihood for λ_0 and λ_1 in terms of D_0, D_1, Y_0, Y_1 is

$$D_0 \log(\lambda_0) - \lambda_0 Y_0 + D_1 \log(\lambda_1) - \lambda_1 Y_1.$$

To express the log likelihood in terms of the new parameter system, we substitute $\theta\lambda_0$ for λ_1, to get

$$D_0 \log(\lambda_0) - \lambda_0 Y_0 + D_1 \log(\theta\lambda_0) - \theta\lambda_0 Y_1,$$

which reduces to

$$D \log(\lambda_0) + D_1 \log(\theta) - \lambda_0 Y_0 - \theta\lambda_0 Y_1,$$

where $D = D_0 + D_1$ is the total number of observed disease events. For the example in Table 13.1, the log likelihood is

$$45 \log(\lambda_0) + 28 \log(\theta) - 2768.9\lambda_0 - 1857.5\theta\lambda_0$$

The purpose of this choice of new parameters for the model is to concentrate the comparison of the rates into the parameter θ, but unfortunately, the log likelihood for these new parameters cannot be divided into a sum of separate parts, one for each parameter. The appearance of the term $1857.5\theta\lambda_0$ means that the shape of the log likelihood with respect to θ depends on the value of λ_0, and this is unknown. When assessing the support for different values of θ, not knowing λ_0 is somewhat of a problem and in this context λ_0 is called a *nuisance parameter.*

There are two ways of dealing with a nuisance parameter when constructing a likelihood for the parameter of interest. These will be described in the next two sections.

13.2 Profile likelihood

The obvious way to deal with a nuisance parameter is to *estimate* its value. For each value of the rate ratio θ, the value of λ_0 which maximizes the likelihood can be determined and substituted into the joint log likelihood. The resulting maximized log likelihood can then used as a measure of support for this value of θ.

This idea is illustrated in Fig. 13.1. The top graph shows the log likelihood ratio for $\log(\lambda_0)$ and $\log(\theta)$ as a contour map. The contour lines,

corresponding to parameter values which have equal log likelihood, are approximately elliptical (this has been aided by the choice of log scales for both parameters, so that they are not bounded). The contours shown correspond to log likelihood ratios of -1, -2, -3, -4, and -5 relative to the maximum value.

The vertical arrows denote specified values of $\log(\theta)$ for which we require to measure the support. For each fixed value of $\log(\theta)$, we find the value of $\log(\lambda_0)$ which maximizes the log likelihood and plot this maximized log likelihood on the lower graph. This is then used to measure the relative support lent by the data to different values of $\log(\theta)$. By analogy with physical maps, this curve is called a *profile* log likelihood. A profile log likelihood is not a true log likelihood since it cannot be directly obtained by taking the log of the probability of the data. However, in most situations it behaves in exactly the same way as a log likelihood. It can be seen from Fig. 13.1 that the value of θ which gives the largest value of the profile log likelihood is also the value corresponding to the maximum of the total log likelihood. The curvature of the profile log likelihood at this maximum point can be used to calculate approximate confidence intervals and Wald tests, and score tests for null values of θ can be carried out using the gradient and curvature of the profile log likelihood at the null value. Similarly, a log likelihood ratio test can be carried out by calculating minus twice the profile log likelihood ratio at the null value of θ.

In the case of the the rate ratio, this process is simplified since the derivation of the profile log likelihood can be carried out algebraically, leading to a mathematical equation for the curve. The value of λ_0 which maximizes the log likelihood for any given value of θ may be shown to be

$$\frac{D}{Y_0 + \theta Y_1}$$

and substituting this for λ_0 in the log likelihood expression gives the profile log likelihood:

$$D_1 \log(\theta) - D \log(Y_0 + \theta Y_1) + D \log(D) - D.$$

Since the last two terms do not depend upon θ, they are irrelevant and may be omitted. We are also at liberty to *add* terms which do not involve θ, and addition of

$$D_1 \log(Y_1) + D_0 \log(Y_0)$$

yields, after some rearrangement, the expression:

$$D_1 \log\left(\frac{\theta Y_1}{Y_0}\right) - D \log\left(1 + \frac{\theta Y_1}{Y_0}\right).$$

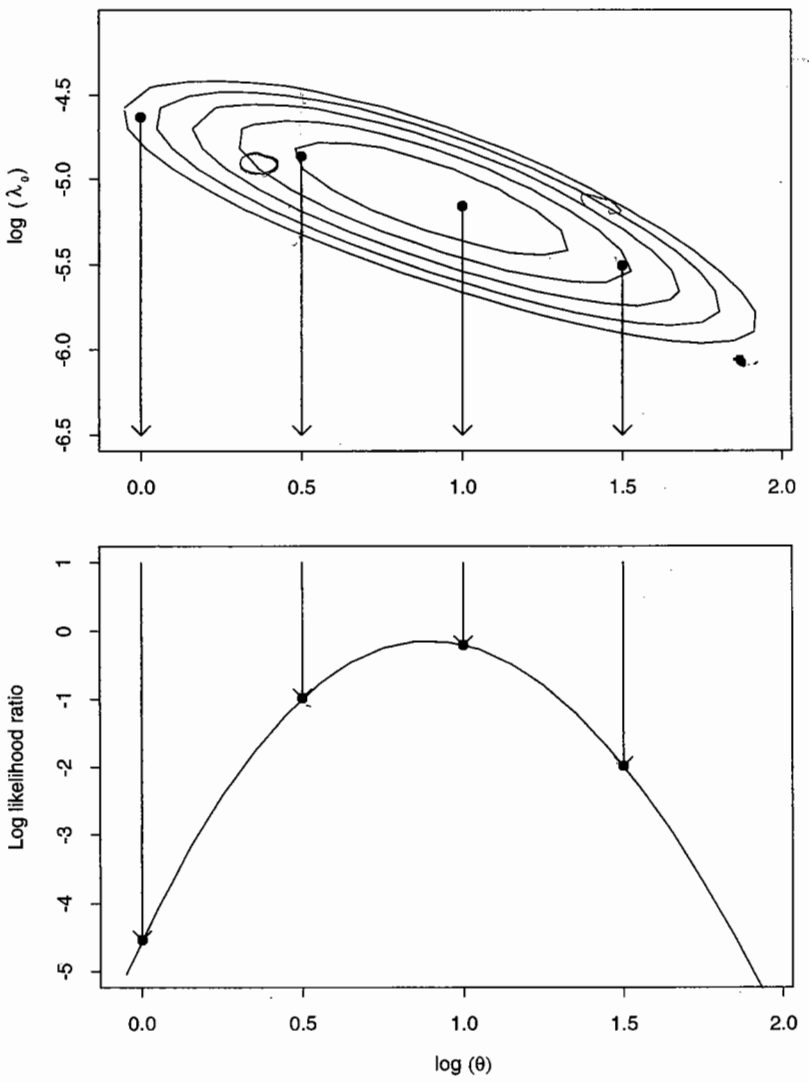

Fig. 13.1. Log likelihood surface for θ and λ (above) and profile log likelihood for θ (below).

This is exactly the same as a Bernoulli likelihood for the odds parameter

$$\Omega = \frac{\theta Y_1}{Y_0}$$

based on a split of D cases as D_1 exposed and $D_0 = D - D_1$ unexposed. It follows that estimation of θ using the profile log likelihood is equivalent to estimating the odds, Ω, in the binary model; the two estimates differ only by the known multiplier, Y_1/Y_0.

From the Bernoulli likelihood, the most likely value of Ω is D_1/D_0 and the standard deviation of $\log(\Omega)$ is

$$\sqrt{\frac{1}{D_0} + \frac{1}{D_1}}.$$

It follows that the most likely value of θ is

$$\frac{D_1/D_0}{Y_1/Y_0} = \frac{D_1/Y_1}{D_0/Y_0}$$

which is the ratio of the most likely values of the two rates and since $\log(\theta)$ differs from $\log(\Omega)$ only by a known constant, the shape of the log likelihoods are identical, and the standard deviation of $\log(\theta)$ is also

$$\sqrt{\frac{1}{D_0} + \frac{1}{D_1}}.$$

Exercise 13.1. Calculate the maximum likelihood estimate of the rate ratio for the data of Table 13.1 and give 90% confidence limits.

For the calculation of p-values, the null hypothesis generally of interest is that the two rates are equal, so that $\theta_\varnothing = 1$ and $\Omega_\varnothing = Y_1/Y_0$. In terms of the corresponding risk parameter the null hypothesis is that

$$\pi_\varnothing = \frac{\Omega_\varnothing}{1 + \Omega_\varnothing} = \frac{Y_1}{Y_0 + Y_1}.$$

The score is

$$U = D_1 - D\pi_\varnothing,$$

which can be written as

$$U = D_1 - E_1$$

where $E_1 = D\pi_\varnothing$ is the expected number of exposed cases under the null hypothesis. The score variance is

$$V = D\pi_\varnothing(1 - \pi_\varnothing).$$

Exercise 13.2. Test the significance of the effect of low energy intake in the data of Table 13.1.

13.3 Conditional likelihood

The approach outlined above starts from the question: what is the probability that, during follow-up, D_0 events occur in the unexposed cohort and D_1 in the exposed cohort? The resulting likelihood involves not only the rate ratio θ (the parameter of interest), but also a nuisance parameter, λ_0. Replacing the unknown nuisance parameter by its most likely value leads to the profile log likelihood for θ. This argument is appealing in that it closely follows the way in which cohort studies are designed and executed — we decide in advance upon the cohort to be followed and the duration of follow-up and wait to see how many disease events occur in different subgroups. However, it is not essential that the likelihood argument should correspond so closely with the study design. In particular, if some aspect of the result contains little or no information about the parameter of interest, then we are free to treat it as if it were fixed by the study design. The aim of such an argument, which is called a *conditional argument*, is to obtain a new probability model for the data which does not involve the nuisance parameter.

In this case the total number of cases tells us nothing about the effect of exposure, which depends on the split among cases between exposed and not exposed. We therefore take the total number of events as fixed, corresponding to a study in which the follow-up continues for just long enough for D events to be observed. The analysis of the study then concentrates on the split of cases between the exposed and unexposed sections of the cohort, and starts from the question: given that D failures occurred, what is the probability that D_0 of them occurred in the unexposed group and D_1 in the exposed group?

The split of the D failures between exposed and unexposed groups may be described using the binary probability model. This is illustrated in Fig. 13.2. The left-hand tree shows the observed split of the failures and the right-hand tree shows the expected split of cases. If Y_1 and Y_0 can be regarded as fixed, the odds that a case was exposed is

$$\Omega = \frac{\lambda_1 Y_1}{\lambda_0 Y_0} = \frac{\theta Y_1}{Y_0},$$

and the log likelihood for θ is

$$D_1 \log\left(\frac{\theta Y_1}{Y_0}\right) - D \log\left(1 + \frac{\theta Y_1}{Y_0}\right).$$

Thus regarding the number of cases as fixed leads directly to a *conditional* log likelihood which depends only upon θ. The log likelihood is conditional

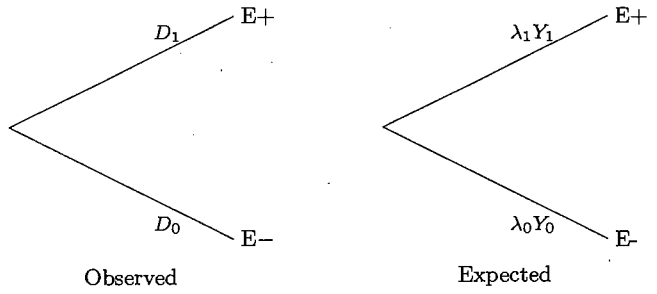

Fig. 13.2. The conditional argument.

in the sense that it takes as fixed an aspect of the data (the total number of events, D) that was, in reality, an unpredictable outcome of the study. In this case the profile and conditional likelihood approaches have led to the same log likelihood and, therefore, to identical estimates and confidence intervals, but in general this will not be the case.

The conditional approach always yields a true log likelihood, being based upon a probability (albeit a *conditional* probability) of observed data. Also, because this probability depends only on the parameter of interest, it can be used to calculate exact p-values and confidence intervals. In our current example, the probabilities for different splits of cases between exposed and unexposed groups, given θ, can be obtained from the binomial distribution. However, the conditional approach is not an automatic method, but relies on our ingenuity in recognizing a suitable conditional argument. Such arguments are not always possible. For example, it has not proved possible to find an argument which leads to a conditional likelihood for the *rate difference*, $\lambda_1 - \lambda_0$.

In contrast, the profile method has the considerable virtue that it can *always* be employed. Even if it is impossible to use an algebraic method to obtain an explicit formula for the profile log likelihood curve, the derivation of the curve numerically by the procedure illustrated in Fig. 13.1 can always be carried out by computer. The difficulty with this approach is that the profile curve is not necessarily a true log likelihood. However, in most situations it does approximately possess the properties of a true log likelihood. These properties can safely be assumed when the number of nuisance parameters is small in comparison with the total quantity of data.

We should note that our current use the conditional approach requires $\lambda_0 Y_0$ and $\lambda_1 Y_1$, the expected numbers of cases in the two groups, to be constants not influenced by the study outcome. Although this is approximately true for the rare events usually studied by epidemiologists (see section 6.3), it may not be an acceptable argument when the probabilities of failure are high. In these cases, the likelihood derived in this chapter

can only be regarded as a profile likelihood and exact tests and confidence intervals are not available.

⋆ 13.4 Approximating profile log likelihoods

For the rate ratio it is possible to derive a *mathematical expression* for the profile log likelihood and hence find a Gaussian approximation from which approximate p-values and confidence intervals can be calculated in the usual way. This is not possible in general. The profile likelihood can always be computed by going through the steps indicated in Fig. 13.1, but the resulting curve usually cannot be represented by a simple algebraic expression. Fortunately some simple rules, derived from calculus, allow us to calculate Gaussian approximations to such profile log likelihoods, and hence algebraic expressions for M, S, U, and V, which we can go on to use in the usual way. These rules and their derivation are explained in Appendix C. Here we briefly summarize the most important rules.

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 bodies of data. If the log likelihood for β_0 has a Gaussian approximation defined by the most likely value M_0 and standard deviation S_0 and the approximation to the log likelihood for β_1 is defined by M_1 and S_1, then the Gaussian approximation of the log likelihood for $\beta_1 - \beta_0$ has

$$\begin{aligned} M &= M_1 - M_0, \\ S &= \sqrt{(S_1)^2 + (S_0)^2}. \end{aligned}$$

The rate ratio is a special case of this more general problem since its logarithm may be written

$$\log\left(\frac{\lambda_1}{\lambda_0}\right) = \log(\lambda_1) - \log(\lambda_0)$$

and in Appendix C it is shown that these rules lead to the same Gaussian log likelihood approximation as we obtained earlier. Here we use them to approximate the profile log likelihood for the rate difference. The most likely value is the difference between the most likely values of the rates,

$$M = \frac{D_1}{Y_1} - \frac{D_0}{Y_0},$$

and, from Chapter 9, $S_1 = \sqrt{D_1}/Y_1$ and $S_0 = \sqrt{D_0}/Y_0$ so the value of S for the rate difference is

$$\sqrt{\frac{D_1}{(Y_1)^2} + \frac{D_0}{(Y_0)^2}}.$$

Exercise 13.3. Calculate an approximate 90% confidence interval for the difference between the rates using the data of Table 13.1.

A still more general problem concerns a weighted sum of parameters, of the form

$$W_1\beta_1 + W_2\beta_2 + W_3\beta_3 + \cdots \quad ,$$

each β parameter again being estimated from independent bodies of data. The Gaussian approximation to the profile log likelihood for the weighted sum has

$$\begin{aligned} M &= W_1M_1 + W_2M_2 + W_3M_3 + \cdots \\ S &= \sqrt{(W_1S_1)^2 + (W_2S_2)^2 + (W_3S_3)^2 + \cdots} \end{aligned} \quad ,$$

where $M_1, S_1, \ldots$ etc. are the most likely values and standard deviations for $\beta_1, \ldots$ etc.. An example is the profile log likelihood for the cumulative failure rate. In Chapter 5 we defined the cumulative rate by

$$\lambda^1T^1 + \lambda^2T^2 + \cdots$$

where $\lambda^1, \lambda^2, \ldots$ are probability rates operating for time periods $T^1, T^2, \ldots$. The cumulative rate is, therefore, a weighted sum of the form discussed in this section.

Exercise 13.4. Using the Gaussian approximation given in Chapter 9 for the log likelihoods for rate parameters, derive an expression for the Gaussian approximation to the profile log likelihood for the cumulative rate.

Solutions to the exercises

13.1 The most likely value of θ is

$$\frac{D_1/Y_1}{D_0/Y_0} = \frac{28/1857.5}{17/2768.9} = 2.48.$$

The standard deviation of the estimate of $\log(\theta)$, is

$$S = \sqrt{1/28 + 1/17} = 0.3075,$$

so that the 90% error factor for θ is

$$\exp(1.645 \times 0.3075) = 1.66.$$

The 90% confidence limits for the rate ratio are $2.48/1.66 = 1.49$ (lower limit) and $2.48 \times 1.66 = 4.12$ (upper limit).

13.2 The observed number of events in the low energy intake group is 28. There were 45 events in total and, under the null hypothesis, the probability of having been exposed is $\pi_0 = 1857.5/4626.4 = 0.402$. The score is

$$U = 28 - 45 \times 0.402 = 9.93,$$

and the score variance is

$$V = 45 \times 0.402 \times (1 - 0.402) = 10.81.$$

The score test is $(U)^2/V = 9.12$, giving $p \approx 0.003$.

13.3

$$M = \frac{28}{1857.5} - \frac{17}{2768.9} = 0.00893 \text{ (8.93 per 1000 person-years).}$$

$$S = \sqrt{\frac{28}{(1857.5)^2} + \frac{17}{(2768.9)^2}} = 0.00321 \text{ (3.21 per 1000 person-years).}$$

The 90% confidence interval is

$$M \pm 1.645S = 3.65 \text{ to } 14.2 \text{ per 1000 person-years.}$$

13.4 The log likelihood for λ^1 is approximated by a Gaussian curve with

$$M^1 = \frac{D^1}{Y^1}, \qquad S^1 = \frac{\sqrt{D^1}}{Y^1}.$$

Similarly for $\lambda^2, \lambda^3, \ldots$ etc. The weights are the durations of observation, $T^1, T^2, \ldots$, so that the profile log likelihood for the cumulative rate has its maximum at

$$M = \frac{D^1}{Y^1}T^1 + \frac{D^2}{Y^2}T^2 + \cdots$$

and the standard deviation of the Gaussian approximation is

$$S = \sqrt{D^1\left(\frac{T_1}{Y^1}\right)^2 + D^2\left(\frac{T^2}{Y^2}\right)^2 + \cdots}.$$

Note that, as we narrow the time bands to clicks, the ratio T/Y approaches $1/N$, where N is the number of subjects under observation during the click. In these circumstances, M is the Aalen–Nelson estimate of the cumulative rate and S may be used to calculate an approximate confidence interval.