9
Approximate likelihoods

Because the Gaussian log likelihood for the mean parameter, μ, takes the simple form

$$-\frac{1}{2}\left(\frac{M-\mu}{S}\right)^2$$

the supported range for μ also takes a simple form, namely

$$M \pm 1.645S.$$

For log likelihoods such as the Bernouilli and Poisson there is no simple algebraic expression for the supported range, and the values of the parameters at which the log likelihood is exactly -1.353 must be found by systematic trial and error. However, the shapes of these log likelihoods are *approximately* quadratic, and this fact can be used to derive simple formulae for approximate supported ranges. Methods based on quadratic approximation of the log likelihood are particularly important because the quadratic approximation becomes closer to the true log likelihood as the amount of data increases.

9.1 Approximating the log likelihood

Consider a general likelihood for the parameter, θ, of a probability model and let M be the most likely value of θ. Since the quadratic expression

$$-\frac{1}{2}\left(\frac{M-\theta}{S}\right)^2$$

has a maximum value of zero when $\theta = M$ it can be used to to approximate the true log likelihood ratio, after an appropriate value of S has been chosen. Small values of S give quadratic curves with sharp peaks and large values of S give quadratic curves with broad peaks. We shall refer to S as the standard deviation of the estimate of θ. Alternatively, it is sometimes called the *standard error* of the estimate.

Once M has been found and S chosen, an approximate supported range for θ is found by solving the equation

$$-\frac{1}{2}\left(\frac{M-\theta}{S}\right)^2 = -1.353,$$

to give

$$\theta = M \pm 1.645S.$$

Full details of how S is chosen are given later in the chapter, but for the moment we shall give formulae for S, without justification, and concentrate on how to use these in practice.

THE RISK PARAMETER

The log likelihood for π, the probability of failure, based on D failures and $N-D$ survivors is

$$D\log(\pi) + (N-D)\log(1-\pi).$$

The most likely value of π is D/N. To link with tradition we shall also refer to the most likely value of π as P (for proportion). The value of S which gives the best approximation to the log likelihood ratio is

$$S = \sqrt{\frac{P(1-P)}{N}}.$$

For the example we worked through in Chapter 3, $D = 4$ and $N = 10$ so that the value of P is 0.4 and

$$S = \sqrt{\frac{0.4 \times 0.6}{10}} = 0.1549.$$

An approximate supported range for π is given by

$$0.4 \pm 1.645 \times 0.1549$$

which is from 0.15 to 0.65, while the supported range obtained from the true curve lies from 0.17 to 0.65. The true and approximate log likelihood curves are shown in Fig. 9.1. The curve shown as a solid line is the true log likelihood ratio curve, while the broken line indicates the Gaussian approximation.

THE RATE PARAMETER

The log likelihood for a rate λ based on D cases and Y person years is

$$D\log(\lambda) - \lambda Y.$$

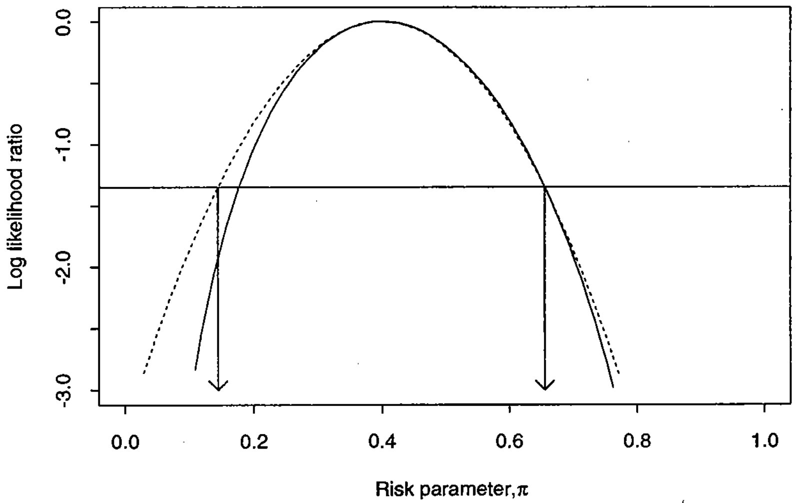

Fig. 9.1. True and approximate Bernouilli log likelihoods.

The most likely value of λ is D/Y and the value of S which gives the best approximation to the log likelihood ratio is

$$S = \frac{\sqrt{D}}{Y}.$$

For the example in Chapter 5, $D = 7$ and $Y = 500$. The most likely value of λ is 0.014 and

$$S = \sqrt{7}/500 = 0.00529.$$

An approximate supported range for λ is therefore

$$0.014 \pm 1.645 \times 0.00529$$

which is from 5.3/1000 to 22.7/1000. The true (solid line) and approximate (broken line) log likelihood ratio curves are shown in Fig. 9.2. The range of support obtained from the true curve spans from 7.0 to 24.6 per 1000.

Exercise 9.1. Find the approximate supported range for π, the probability of failure, based 7 failures and 93 survivors. Find also the approximate supported range for λ, the rate of failure, based on 30 failures over 1018 person-years.

9.2 Transforming the parameter

The Gaussian log likelihood curve for μ is symmetric about M and extends *indefinitely* to either side. However, the parameters of some probability

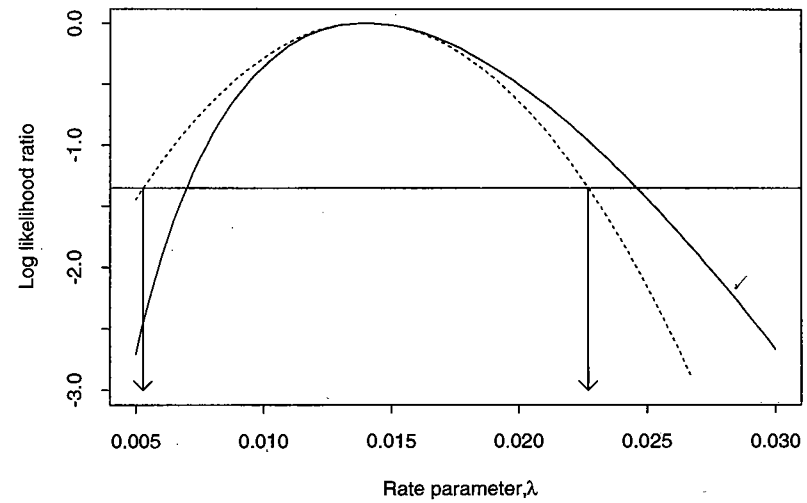

Fig. 9.2. True and approximate Poisson log likelihoods.

models are not free to vary in this manner. For example, the rate parameter λ can take only positive values, and the risk parameter must lie between 0 and 1. Approximate supported ranges for such parameters calculated from the Gaussian approximation can, therefore, include impossible values.

The solution to this problem is to find some function (or *transformation*) of the parameter which is unrestricted and to first find an approximate supported range for the transformed parameter.

THE LOG RATE PARAMETER

The rate parameter λ can take only positive values, but its logarithm is unrestricted. To calculate an approximate supported range for λ it is better, therefore, to first calculate a range for $\log(\lambda)$, and then to convert this back to a range for λ. Note that the range for $\log(\lambda)$ will always convert back to positive values for λ. To find the approximate range for $\log(\lambda)$ we need a new value of S — that which gives the best Gaussian approximation to the log likelihood ratio curve when plotted against $\log(\lambda)$. When a rate λ is estimated from D failures over Y person-years, this value of S is given by

$$S = \sqrt{1/D}.$$

Fig. 9.3 illustrates this new approximation for our example in which $D = 7$ and $Y = 500$ person-years. Here,

$$S = \sqrt{1/7} = 0.3780,$$

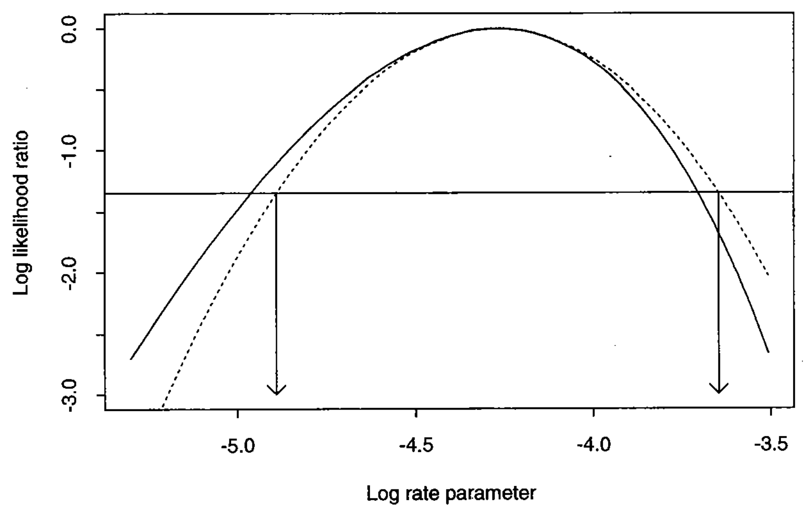

Fig. 9.3. Approximating the log likelihood for $\log(\lambda)$.

and an approximate supported range for $\log(\lambda)$ is

$$\log(7/500) \pm 1.645 \times \sqrt{1/7},$$

which is from −4.890 to −3.647. The range for λ is therefore from $\exp(-4.890)$ to $\exp(-3.647)$ which spans from 7.5/1000 to 26.1/1000.

A more convenient way of carrying out this calculation is suggested by noting that the limits of the range for λ are given by

$$\frac{7}{500} \overset{\times}{\div} \exp\left(1.645\sqrt{\frac{1}{7}}\right) = 0.014 \overset{\times}{\div} 1.862.$$

The range is then from $0.014/1.862 = 7.5/1000$ to $0.014 \times 1.862 = 26.1/1000$, as before. We shall refer to the quantity

$$\exp(1.645S)$$

as an *error factor*.

THE LOG ODDS PARAMETER

The same thing can be done when calculating a supported range for the risk parameter π based on D failures in N subjects. The value of π is restricted on both sides, by 0 on the left and by 1 on the right. The value of $\log(\pi)$ is

still restricted on the right by zero because $\log(1) = 0$, but $\log(\Omega)$, where Ω is the odds corresponding to π, is not restricted at all. Hence we first find a range for $\log(\Omega)$ and then convert this back to a range for π. The most likely value of $\log(\Omega)$ is

$$M = \log\left(\frac{D}{N-D}\right)$$

and the value of S for approximating the log likelihood for $\log(\Omega)$ is

$$S = \sqrt{\frac{1}{D} + \frac{1}{N-D}}.$$

For the example where $D = 4$ and $N - D = 6$,

$$S = \sqrt{\frac{1}{4} + \frac{1}{6}} = 0.6455,$$

and an approximate supported range for $\log(\Omega)$ is given by

$$\log\left(\frac{4}{6}\right) \pm 1.645 \times 0.6455,$$

that is, from −1.4673 to 0.6564. This is a range for $\log(\Omega)$ and it is equivalent to a range for Ω from $\exp(-1.4673) = 0.231$ to $\exp(0.6564) = 1.928$. This can be calculated more easily by first calculating the error factor

$$\exp(1.645 \times 0.6455) = 2.892.$$

The most likely value of Ω is $4/6 = 0.667$, so that the supported range for Ω is

$$0.667 \overset{\times}{\div} 2.892$$

that is, from 0.231 to 1.928 as before. Finally, remembering that $\pi = \Omega/(1+\Omega)$, the range for π is given by

$$\frac{0.231}{1.231} \text{ to } \frac{1.928}{2.928}$$

which is from 0.19 to 0.66.

Some of the more commonly used values of S obtained by approximating the log likelihood are gathered together in Table 9.1.

Exercise 9.2. Repeat Exercise 9.1 by first finding 90% intervals for $\log(\Omega)$ and $\log(\lambda)$ respectively, and then converting these to intervals for π and λ.

Exercise 9.3. Repeat the above exercise using error factors.

Table 9.1. Some important Gaussian approximations

Parameter	M	S
π	$D/N = P$	$\sqrt{P(1-P)/N}$
λ	D/Y	$\sqrt{D}/Y$
$\log(\Omega)$	$\log[D/(N-D)]$	$\sqrt{1/D + 1/(N-D)}$
$\log(\lambda)$	$\log(D/Y)$	$\sqrt{1/D}$

★ 9.3 Finding the best quadratic approximation

We now return to the problem of how to determine the values for M and S. To do this we need some elementary ideas of calculus summarized in Appendix B. In particular, we need to be able to find the *gradient* (or *slope*) of the log-likelihood curve together with its *curvature*, which is defined as the rate of change of the gradient. The mathematical terms for these quantities are the first and second *derivatives* of the log likelihood function.

The value of M can be found by a direct search for that value of of θ which maximizes the log likelihood, but it is often easier to find the value of θ for which the gradient of the log likelihood is zero; this occurs when $\theta = M$.

The value of S is chosen to make the curvature of the quadratic approximation equal to that of the true log likelihood curve at M, thus ensuring that the true and approximate log likelihoods are very close to each other near $\theta = M$. The quadratic approximation to the log likelihood ratio is

$$-\frac{1}{2}\left(\frac{M-\theta}{S}\right)^2,$$

and the rules summarized in Appendix B show that the curvature of this is constant and takes the value

$$-\frac{1}{(S)^2}.$$

We therefore choose the value of S to make $-1/(S)^2$ equal to the curvature of the true log likelihood curve at its peak.

THE RATE PARAMETER

The log likelihood for a rate λ is

$$D\log(\lambda) - \lambda Y.$$

Using the rules of calculus given in Appendix B the gradient of $\log(\lambda)$ is $1/\lambda$ and the gradient of λ is 1. Hence the gradient of the log likelihood is

$$\frac{D}{\lambda} - Y.$$

The maximum value of the log likelihood occurs when the gradient is zero, that is, when $\lambda = D/Y$, so the most likely value of λ is D/Y. The curvature of a graph at a point is defined as the rate of change of the gradient of the curve at that point. The rules of calculus show this to be

$$-\frac{D}{(\lambda)^2}.$$

The peak of the log likelihood occurs at $\lambda = D/Y$ so the curvature at the peak is found by replacing λ by D/Y in this expression to obtain

$$-\frac{(Y)^2}{D}.$$

Setting this equal to $-1/(S)^2$ gives

$$S = \sqrt{D}/Y,$$

which is the formula quoted earlier.

THE RISK PARAMETER

The log likelihood for the probability π based on D positive subjects out of a total of N is

$$D\log(\pi) + (N-D)\log(1-\pi).$$

The gradient of the log likelihood is

$$\frac{D}{\pi} - \frac{N-D}{1-\pi}$$

which is zero at $\pi = D/N$, also referred to as P. The gradient of the gradient is

$$-\frac{D}{(\pi)^2} - \frac{N-D}{(1-\pi)^2},$$

so the curvature at $\pi = P$ is

$$-\frac{D}{(P)^2} - \frac{N-D}{(1-P)^2}.$$

Replacing D by NP and $N-D$ by $N(1-P)$, this reduces to

$$-\frac{N}{P(1-P)}$$

so

$$S=\sqrt{\frac{P(1-P)}{N}}.$$

☆ 9.4 Approximate likelihoods for transformed parameters

When the log likelihood for a parameter is plotted against the log of the parameter rather than the parameter itself, the curvature at the peak will be different. For example, the log likelihood for a rate parameter λ is

$$D\log(\lambda)-\lambda Y.$$

Plotting this against $\log(\lambda)$ is the same as expressing the log likelihood as a function of $\log(\lambda)$. To do this we introduce a new symbol β to stand for $\log(\lambda)$, so

$$\beta=\log(\lambda),\quad \lambda=\exp(\beta).$$

In terms of β the log likelihood is

$$D\beta-Y\exp(\beta).$$

The gradient of this with respect to β is

$$D-Y\exp(\beta)$$

and the curvature is

$$-Y\exp(\beta).$$

The most likely value of $\exp(\beta)$ (which equals λ) is D/Y, so the curvature at the peak is

$$-Y\times(D/Y)=-D.$$

It follows that

$$S=\sqrt{1/D}.$$

In general, derivations such as that above can be simplified considerably by using some further elementary calculus which provides a general rule for the relationship between the values of S on the two scales. In the case of the log transformation, this rule states that multiplying the value of S on the scale of λ by the gradient of $\log(\lambda)$ at $\lambda=M$ gives the value of S on the scale of $\log(\lambda)$. The rules of calculus tell us that, at $\lambda=M$, the gradient

of the graph of $\log(\lambda)$ against λ is $1/M$. Since, on the λ scale, $M=D/Y$ and $S=\sqrt{D}/Y$, the rule tells us that the value of S for $\log(\lambda)$ is

$$\frac{\sqrt{D}}{Y}\times\frac{Y}{D}=\sqrt{\frac{1}{D}}.$$

This agrees with the expression obtained by the longer method.

A similar calculation shows that the curvature of the Bernouilli log likelihood, when plotted against $\log(\Omega)$, the log odds, is given by

$$S=\sqrt{\frac{1}{D}+\frac{1}{N-D}}.$$

Solutions to the exercises

9.1 An approximate supported range for π is given by

$$0.07\pm1.645S$$

where $S=\sqrt{0.07\times0.93/100}$. This gives a range from 0.028 to 0.112.
An approximate supported range for λ is given by

$$30/1018\pm1.645S$$

where $S=\sqrt{30}/1018$. This gives a range from 21/1000 to 38/1000.

9.2 The approximate supported range for $\log(\Omega)$ is given by

$$\log(7/93)\pm1.645S$$

where

$$S=\sqrt{\frac{1}{7}+\frac{1}{93}}=0.3919.$$

This gives a range from −3.231 to −1.942. The range for Ω is from 0.040 to 0.143, and the range for π is from 0.038 to 0.125.
The approximate supported range for $\log(\lambda)$ is given by

$$\log(30/1018)\pm1.645S$$

where

$$S=\sqrt{1/30}=0.1826.$$

This gives a range from −3.825 to −3.224. The range for λ is from 22/1000 to 40/1000.

9.3 The error factor for Ω is

$$\exp(1.645 \times 0.3919) = 1.905.$$

The most likely value for Ω is $7/93 = 0.075$ and the range for Ω is from $0.075/1.905 = 0.040$ to $0.075 \times 1.905 = 0.143$. The range for π is from 0.038 to 0.125.
The error factor for the rate is

$$\exp(1.645 \times 0.1826) = 1.350.$$

The most likely value of the rate is $29/1000$ with range from $29/1.350 = 22$ per 1000 to $29 \times 1.350 = 40$ per 1000.