Foundations of Medical Statistics

Statistical Inference 5: Approximate log-likelihood ratios

Aims

The aim of this session is to introduce (i) the concept of approximating a log-likelihood ratio by a normal log-likelihood ratio and (ii) to show how can be used to construct confidence intervals.

Objectives

At the end of this session you should:

- understand how normal log-likelihood ratios can be used to approximate other log-likelihood ratios;
- be able to construct a normal log-likelihood ratio to approximate another log-likelihood ratio;
- be able to get a confidence interval from this.

5.1 Motivation

For likelihoods such as the binomial or Poisson, there is no simple expression for the end points of the likelihood ratio confidence interval. For example, in the binomial case, there is no closed form algebraic solution to the equation

$$k \log(\pi) + (n - k) \log(1-\pi) = \text{constant}.$$

However, we have already seen that log-likelihoods are approximately quadratic; we have further seen an example of a quadratic log-likelihood – the log-likelihood for the mean of a normal distribution when the variance is known. Thus we know it is easy to find the endpoints of the likelihood ratio confidence interval from a quadratic log-likelihood.

Therefore, in order to derive approximate likelihood ratio confidence intervals, it makes sense to form a quadratic approximation to the log-likelihood ratio. Further, when the likelihood is for a normal mean with known variance, our approximation should be exact.

5.2 Normal approximation to the log-likelihood

We know that if $\bar{X} \sim N(\mu,\sigma^2/n)$

$$llr(\mu | \bar{X}) = \ell(\mu | \bar{X}) = -\frac{1}{2}\left(\frac{\mu - \bar{X}}{\sigma/\sqrt{n}}\right)^2$$

where $\bar{X}$ is the maximum likelihood estimator of the Normal parameter μ with known variance parameter, and $\sigma/\sqrt{n}$ is the standard error of $\bar{X}$.

This leads us to propose the following approximation. Suppose θ is our parameter of interest, then consider the following quadratic function of θ:

$$f(\theta|\text{data}) = -\frac{1}{2}\left(\frac{\theta - M}{S}\right)^2 \tag{5.1}$$

This has the form of a Normal quadratic log-likelihood (ratio), with maximum likelihood estimator *M* having standard error *S*. This function will approximate the true likelihood $\ell\,(\theta|\text{data})$ if we choose *M* and *S* appropriately.

M we can choose to achieve agreement with the behaviour of the Normal log-likelihood ratio, so we choose $M = \hat{\Theta}$, the maximum likelihood estimator for θ.

However, to choose *S* we need to satisfy two constraints:

- S should be the standard error of estimator $\hat{\Theta}$;

- We need to choose S so that the quadratic function f approximates the true log-likelihood ratio as closely as possible *at the maximum*: therefore at the maximum it should have the same curvature as the true $llr(\theta)$. Note a) that *M* is irrelevant to the curvature of the quadratic, which is only a function of *S*; and b) that since the log-likelihood ratio and the log-likelihood are the same shape, their curvatures are identical.

Now, the curvature of a function is its second derivative; so the curvature of the true log-likelihood ratio at the maximum is

$$\frac{d^2}{d\theta^2}\ell(\theta)\Big|_{\theta=\hat{\theta}} = \ell''(\hat{\theta})$$

and it turns out that we can satisfy both constraints if we set

$$S^2 = -\frac{1}{\ell''(\theta)}\Big|_{\theta=\hat{\theta}}$$

This is because theoretical results (see Appendix A, non-examinable, at back of notes) tell us that

$$\text{Var}(\hat{\Theta}) \approx -\frac{1}{\ell''(\theta)}\Big|_{\theta=\hat{\theta}}$$

with equality as $n \to \infty$, or if the true log-likelihood is quadratic, and the curvature of (5.1) is

$$f''(\theta) = -\frac{1}{S^2}$$

So if we obtain $\ell''(\hat{\theta})$ and set

$$\ell''(\theta) = -\frac{1}{S^2} \Rightarrow S^2 = -\frac{1}{\ell''(\hat{\theta})}$$

then the approximate log-likelihood ratio $f(\theta)$ will have the same curvature at the maximum as the true log-likelihood ratio $llr(\theta)$.

5.2.1 Obtaining approximate log-likelihood ratio confidence intervals

Once the log-likelihood ratio is approximated by the Normal quadratic, the log-likelihood ratio confidence interval follows directly as seen in Inference 4, so that a 100(1-α)% confidence interval is given by θ such that

$$-2f(\theta) < \chi^2_{1,(1-\alpha)}$$

so the confidence limits will be $M \pm \sqrt{\chi^2_{1,(1-\alpha)}}S$. For 95% intervals, $\chi^2_{1,0.95} = 3.84$, so we have the familiar $M \pm 1.96S$.

Example 5.2.1 *Approximate likelihood for Poisson distribution*

Recall from Section 3.8 that the Poisson log-likelihood with p person years at risk, d events and a rate parameter λ is $\ell(\lambda|d) = d\log(\lambda) - \lambda p$ (ignoring terms not involving λ). Thus $\ell''(\lambda) = -d/\lambda^2$.

The MLE $\hat{\lambda} = d/p$ (see 3.6), so equate $-1/S^2$ to $-d/\hat{\lambda}^2 = -p^2/d$, so $S = \sqrt{d}/p$.
So the approximate 95% likelihood ratio confidence interval for λ is

$$d/p \pm 1.96\sqrt{d}/p$$

EXERCISE 5.2.1
Suppose $K \sim Bin(n, \pi)$. Show that S = $\sqrt{p(1-p)/n}$ where $p = k/n$, and derive an approximate supported range for π.

Note: You have seen this result in the context of the Normal approximation to the binomial. Such 'Normal approximation' results encountered elsewhere in the course are here derived in terms of likelihood.

5.3 Parameter transformations

In many situations, a better quadratic approximation to the true log-likelihood ratio curve may be achieved by transforming the parameter.

If θ is a parameter which is transformed to $g(\theta)$, then the MLE of $g(\theta)$ will be $g(\hat{\theta})$, where $\hat{\theta}$ is the MLE of θ. Similarly, if θ_1 to θ_2 is the likelihood ratio confidence interval for θ, then $g(\theta_1)$ to $g(\theta_2)$ will be the likelihood ratio confidence interval for $g(\theta)$.

Example 5.3.1 *Transformation of the Poisson rate*

The rate parameter λ must be positive, while log(λ) is unrestricted. For rates based on small numbers of observed events, a better quadratic approximation is obtained working with log(λ). Write β = log(λ). Then we can substitute into the Poisson log-likelihood:

$$\ell(\lambda|d) = d\log(\lambda) - \lambda p$$
$$\Rightarrow \quad \ell(\beta|d) = d\beta - pe^{\beta}.$$

By the invariance property of MLEs, $\hat{\beta} = \log(\hat{\lambda}) = \log(d/p)$. To obtain S for the approximate log-likelihood ratio we take the second derivative:

$$\ell''(\beta) = -pe^{\beta}$$
$$\Rightarrow \quad \ell''(\hat{\beta}) = -pd/p = -d = -1/S^2$$
$$\Rightarrow \quad S = 1/\sqrt{d}\ .$$

EXERCISE 5.3.1 *Transformation of binomial risk*

Several transformations of a risk parameter π may be considered. Transformations 3 and 4 below are termed the **odds** and **log odds** parameters respectively. What ranges of values can the following parameters take?

1. π 2. log(π) 3. $\dfrac{\pi}{1-\pi}$ 4. $\log\left(\dfrac{\pi}{1-\pi}\right)$

EXERCISE 5.3.2 *Binomial example continued*

Show that based on observing k events out of n subjects, and a binomial likelihood, the standard error of the MLE of the log risk parameter log(π) is given by $S = \sqrt{1/k - 1/n}$.

Note: These results for SEs of MLEs on a transformed scale are exactly analogous to those obtained using the variance formula for data transformation: $\mathrm{Var}[g(x)] \approx [g'(\mathrm{E}[x])]^2.\mathrm{Var}[x]$, also known as the 'delta' method.

It is very important to appreciate the implication of this: whereas the MLE of a transformed estimator is directly obtainable using the invariance property of MLEs, the standard error is not invariant to transformation: the standard error of a function of an estimator *cannot* be obtained simply as the function of the standard error of the estimator.