Foundations of Medical Statistics

Statistical Inference 4: Log-likelihood ratio

Aims

The aim of this Chapter is to introduce the concept of log-likelihood ratios and show how their distribution can be used to construct confidence intervals.

Objectives

At the end of this session you should:

- understand what a log-likelihood ratio is;
- know what its distribution is;
- be able to calculate log-likelihood ratios and find log-likelihood ratio confidence intervals in simple situations.

4.1 Log-likelihood ratio

It is helpful to rescale the log-likelihood so that its maximum value is 0. This is achieved by constructing the **log-likelihood ratio**:

$$llr(\theta) = \ell(\theta \mid \text{data}) - \ell(\hat{\theta} \mid \text{data})$$

Recall that $\ell(\theta)$ has a maximum value of $\ell(\hat{\theta})$, so $llr(\theta)$ has a maximum of 0 at $\theta = \hat{\theta}$. $llr(\theta)$ is called the log-likelihood ratio because of course it is the log of the ratio of likelihoods $LR(\theta)=L(\theta)/L(\hat{\theta})$, a ratio with a maximum value of 1.

We saw in Inference 3 that only terms involving the parameter are important to the log-likelihood; and in fact terms not involving the parameter would in any case cancel out of the log-likelihood ratio.

Example 4.1.1

For the binomial example in Section 3.2, based on observing 4 events out of 10 subjects:

$$L(\pi|X=4) = \pi^4 (1-\pi)^{(10-4)} \binom{10}{4} \Rightarrow \ell(\pi) = \log[\pi^4 (1-\pi)^6]\text{, ignoring terms not in } \pi.$$

$$\Rightarrow \quad llr(\pi) = \ell(\pi) - \ell(\hat{\pi}) = \log[\pi^4 (1-\pi)^6] - \log[0.4^4 (1-0.4)^6]$$

It is important to note that constructing $LR(\theta)$ or $llr(\theta)$ merely corresponds to a rescaling of the y-axes of the plots against θ: there is no change in shape of the two functions. Plots for the example above, of $LR(\pi)$ and $llr(\pi)$, are shown in Figure 4.1, and show this rescaling.

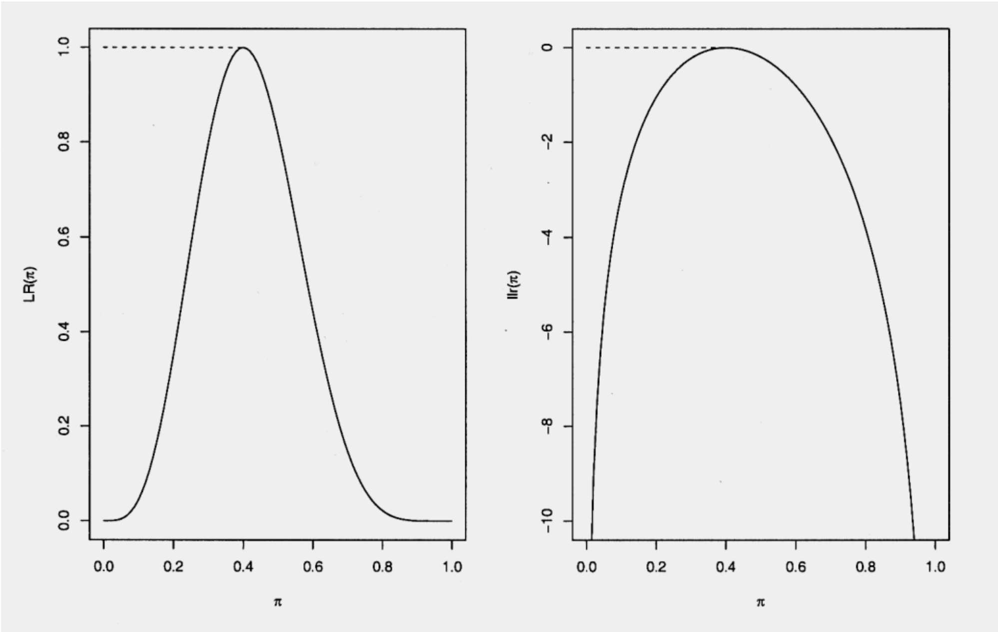

Figure 4.1: left panel, binomial likelihood ratio; right panel: binomial log-likelihood ratio, when n=10 and 4 events observed.

EXERCISE 4.1.1

Suppose y is a single observation from random variable $Y \sim N(\mu, \tau^2)$, where τ^2 is **known**. Show that the log-likelihood for μ is $-\frac{1}{2}\left(\frac{y-\mu}{\tau}\right)^2$, calculate the maximum likelihood estimate $\hat{\mu}$, and hence write down $llr(\mu)$.

4.2 Log-likelihood ratio for *n* independent Normally distributed random variables

Suppose random variables $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where we again assume σ^2 is **known**. We denote observed data $x_1, ..., x_n$ as $\underline{x}$. Then

$$L(\mu|\underline{x}) = \prod_{i=1}^{n} f(x_i \mid \mu)$$

$$\ell(\mu|\underline{x}) = \sum_{i=1}^{n} \log f(x_i \mid \mu) = \sum_{i=1}^{n} -\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2 = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

ignoring terms not involving μ. Now, we know from Inference 2 of the following partition:

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\mu)^2 ;$$

Thus, again ignoring terms not involving μ,

$$\ell(\mu|\underline{x}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\bar{x}-\mu)^2 = -\frac{n}{2\sigma^2}(\bar{x}-\mu)^2 = -\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 \qquad (1)$$

As in the case of a single Normal random variable, $\ell(\hat{\mu})$=0, so that in this Normal case again $llr(\mu) = \ell(\mu)$. However, note that when we refer to the 'Normal case' we are generally referring to the Normal mean parameter: for the Normal variance parameter, it is *not* the case that $llr(\sigma^2)$ is the same as $\ell(\sigma^2)$.

Alternatively, it is important to note that we could have argued directly:

$$X_1,..., X_n \overset{iid}{\sim} N(\mu,\sigma^2) \Rightarrow \bar{X} \sim N(\mu,\sigma^2/n).$$

Now we have reduced the context to a single random variable (whose realisation we observe as $\bar{x}$), so that $\ell(\mu|\underline{x}) = \ell(\mu|\bar{x})$; we can therefore apply directly the result shown in Exercise 4.1.1 to give the result above in (1).

Note: If we can manage with just the single random variable $\bar{X}$, why do we need to analyse these situations with the n separate random variables from which the sample mean is calculated? After all, in binomial and Poisson contexts we just use the single random variable counting up the events: we could for these two models have used separate variables for each subject in the sample, indicating X = 0 or 1 depending on whether the subject had the event (or was in the appropriate binomial category); we then would have obtained exactly the same results as we have obtained by adding the separate 0/1 variables and using just the resulting sum. However, for the Normal model in most practical applications we do not only have to estimate the population mean – for which $\bar{X}$ is sufficient (in both the colloquial and technical sense, since it is a *sufficient* statistic for estimating the mean): we also have to estimate the population variance from the sample, in order then in terms of this to estimate the variance of the estimator $\bar{X}$. Then, to estimate the population variance, we need to think in terms of the n separate random variables. In binomial and Poisson models the variance of the relevant estimator, eg the sample proportion, does not depend on any other quantity we need to estimate, so analyzing in terms, eg, of the total number of events is quite sufficient.

4.3 Distribution of log-likelihood ratio for *n* independent Normally distributed random variables

We have been considering curves of the form $\ell(\mu|\underline{x})$ ranging over varying values of the parameter μ. Consider, however, a single value of the parameter, the true value. To emphasise this we now denote this value μ_0, to distinguish it from other possible values which generate the likelihood curve; by fixing μ at μ_0, $\ell(\mu_0|\underline{x})$ is then a single value on the curve.

It is now important to note that likelihood, log-likelihood and log-likelihood ratio curves will vary over repeated sampling, since the data we condition on are realisations of random variables: the single value $\ell(\mu_0|\underline{x})$ is a realisation of the random variable $\ell(\mu_0|\underline{X})$. We will now obtain the distribution of this random variable in the Normal case.

Consider the Normal example with a single random variable with the true mean parameter value emphasised: $Y \sim N(\mu_0, \tau^2)$, and with known τ^2, and consider the log-likelihood ratio for the true parameter value:

$$llr(\mu_0|Y) = \ell(\mu_0) - \ell(\hat{\mu}) = -\frac{1}{2}\left(\frac{Y-\mu_0}{\tau}\right)^2$$

We know, however, that

$$\left(\frac{Y-\mu_0}{\tau}\right) \sim N(0,1)$$

$$\Rightarrow \quad \left(\frac{Y-\mu_0}{\tau}\right)^2 \sim \chi_1^2$$

$$\Rightarrow \quad -2\, llr(\mu_0|Y) \sim \chi_1^2$$

And therefore in the case where $X_1,..., X_n \overset{iid}{\sim} N(\mu_0,\sigma^2)$, and σ^2 is known,

$$-2\, llr(\mu_0|\bar{X}) \sim \chi_1^2$$

The distribution of the statistic $\bar{X}$ on both sides around the true mean, familiar as a Normal distribution with variance σ^2/n, has been transposed into the chi-squared distribution of the log-likelihood ratio value at the true mean, conditional on the statistic. This distribution can be visualised as the distribution of the varying gap, with repeated sampling, between $\ell(\mu_0|\bar{X})$ and $\ell(\hat{\mu}|\bar{X})$, the latter always being zero in the Normal case.

4.4 Distribution of log-likelihood ratio for *n* independent random variables

Given the ramifications of the central limit theorem, and the result for the Normal case, the following result is perhaps not surprising:

THEOREM 4.4.1 *Suppose* $X_1,..., X_n \overset{iid}{\sim} f(x|\theta)$. *Under certain 'regularity conditions', in repeated sampling from a population where the true parameter value is* θ_0 *the distribution of the statistic*

$$-2llr(\theta_0) = -2\{\ell(\theta_0) - \ell(\hat{\theta})\} \xrightarrow[n\to\infty]{} \sim \chi_1^2$$

Note: in practice the regularity condition to observe is that the maximum likelihood estimate does not occur at either a discontinuity in the likelihood, or the boundary of the parameter space.

The proof of this theorem is given in Appendix A (*non-examinable*) at the end of the course notes.

4.5 Likelihood ratio confidence intervals

The implication of Theorem 4.4.1 is that in repeated sampling, if the sample size is large enough — and in practice n>30 is often adequate — we can assume that

$$-2llr(\theta_0) = -2\{\ell(\theta_0) - \ell(\hat{\theta})\} \sim \chi_1^2 .$$

Therefore

$$\text{Prob}(-2llr(\theta_0) \leq \chi^2_{1,0.95}=3.84) = 0.95$$

$$\Rightarrow \quad \text{Prob}(llr(\theta_0) \geq -3.84/2 \;=\; -3.84/2 = -1.92) = 0.95$$

so $\{\theta$ such that $\ell(\theta) - \ell(\hat{\theta}) \geq -1.92\}$ forms a 95% likelihood ratio confidence interval for the true value θ_0. The 95% likelihood ratio confidence limits are the two values that solve

$$-2\, llr(\theta) = \chi^2_{1,0.95}=3.84 \;\Rightarrow llr(\theta) = -1.92$$

We shall see below that *likelihood ratio* based confidence intervals coincide with the familiar *Normal theory* confidence intervals for Normally distributed populations.

4.5.1 Binomial example

Returning to our binomial example where $X \sim Bin(10, \pi)$ and we observe X=4, we saw in Example 4.1.1 that

$$llr(\pi) = \log[\pi^4\,(1-\pi)^6] - \log(0.4^4\, 0.6^6)$$

Equating this with -1.92 cannot be solved algebraically. However, if we graph this log-likelihood ratio and draw a line across at -3.84/2 = -1.92, we can read off the range of values of π for which $llr(\pi) \geq -1.92$, and thus obtain the confidence interval. Figure 4.5 shows such a plot; the 95% likelihood ratio confidence interval is the range (0.15,0.7).

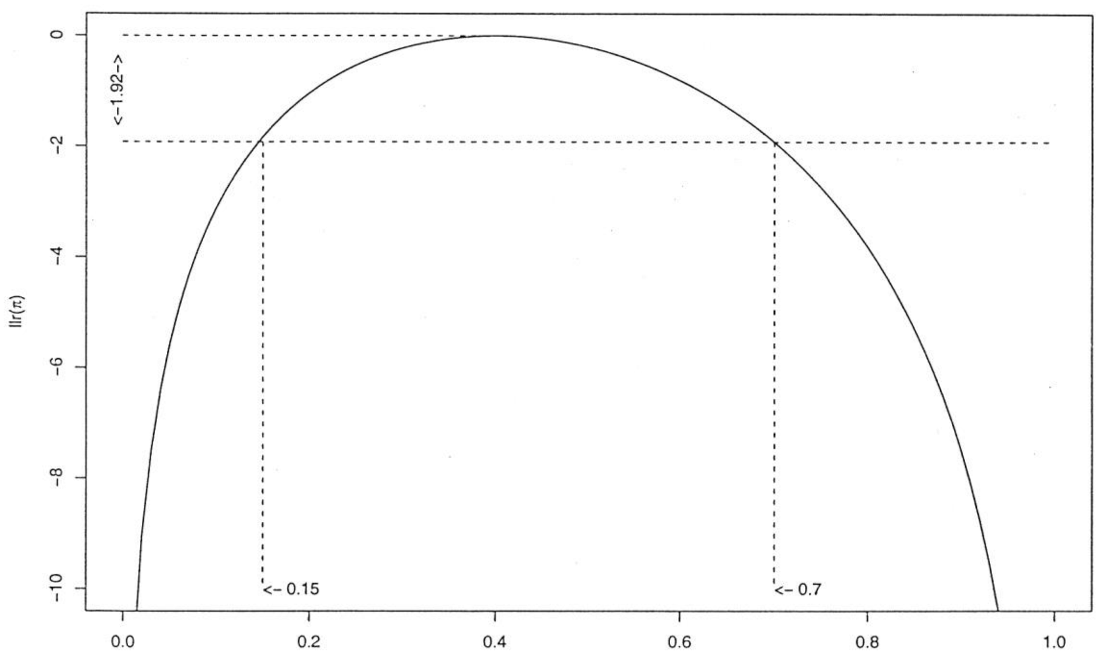

Figure 4.5: Log-likelihood ratio for binomial example, with 95% likelihood ratio confidence interval shown.

4.5.2 Normal example

We saw in 4.2 that in the example $X_1,..., X_n \overset{iid}{\sim} N(\mu,\sigma^2)$, with σ^2 known,

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

This is a **quadratic** function of μ, with maximum 0 at the MLE $\hat{\mu} = \bar{x}$. The 95% likelihood ratio confidence interval has as confidence limits the two values that solve

$$-2 \times -\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 = 3.84$$

i.e. $\bar{x} - \sqrt{3.84}\,\sigma/\sqrt{n}$ and $\bar{x} + \sqrt{3.84}\,\sigma/\sqrt{n}$

This is the familiar Normal theory confidence interval, when we note that $3.84 = 1.96^2$; and, indeed, when the log-likelihood is a quadratic function, the Normal theory confidence interval of the form [mean ± $z_{\alpha/2}$ standard error] corresponds exactly to the likelihood ratio confidence interval.

Although not all log-likelihood functions are exactly quadratic, if the regularity conditions hold, all become approximately quadratic (and therefore symmetric) as the number of independent observations increases. However, excepting the Normal case, note that likelihood ratio confidence intervals are not in general symmetric about the mle; neither are they equitailed (if a 95% confidence interval is equitailed then the probability the true parameter value is above or below it is 5/2=2.5%).

Coverage probability: This is the probability of a correct estimate of the unknown parameter; so the coverage probability of a confidence interval is the probability that it contains the unknown parameter. Thus for a Normal mean model, the coverage probability of a 95% confidence interval is exactly 0.95. [See Pawitan p129 for non-normal model coverage.]