10
Likelihood, probability, and confidence

The supported range for a parameter has so far been defined in terms of the cut-point -1.353 for the log likelihood ratio. Some have argued that the scientific community should accept the use of the log likelihood ratio to measure support as *axiomatic*, and that supported ranges should be reported as 1.353 unit supported ranges, or 2 unit supported ranges, with the choice of how many units of support left to the investigator. This notion has not met with widespread acceptance because of the lack of any intuitive feeling for the log likelihood ratio scale — it seems hard to justify the suggestion that a log likelihood ratio of -1 indicates that a value is supported while a log likelihood ratio of -2 indicates lack of support. Instead it is more generally felt that the reported plausible range of parameter values should be associated in some way with a *probability*. In this chapter we shall attempt to do this, and in the process we shall finally show why -1.353 was chosen as the cut-point in terms of the log likelihood ratio.

There are two radically different approaches to associating a probability with a range of parameter values, reflecting a deep philosophical division amongst mathematicians and scientists about the nature of probability. We shall start with the more orthodox view within biomedical science.

10.1 Coverage probability and confidence intervals

Our first argument is based on the frequentist interpretation of probability in terms of relative frequency of different outcomes in a very large number of repeated "experiments". With this viewpoint the statement that there is a probability of 0.9 that the parameter lies in a stated range does not make sense; there can only be one correct value of the parameter and it will either lie within the stated range or not, as the case my be. To associate a probability with the supported range we must imagine a very large number of repetitions of the study, and assume that the scientist would calculate the supported range in exactly the same way each time. Some of these ranges will include the true parameter value and some will not. The relative frequency with which the ranges include the true value is called the *coverage probability* for the range, although strictly speaking it is the coverage probability for the method of choosing the range.

We shall start with Gaussian probability model and consider the estimation of the mean μ, from a single observation x, when the standard deviation, σ, is known. The log likelihood ratio for μ is

$$-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2.$$

We saw in Chapter 8 that the range of values for μ with log likelihood ratios above the cut-point of -1.353 is

$$x \pm 1.645\sigma.$$

We shall now show that the coverage probability of this range is 0.90 by imagining an endless series of repetitions of the study with the value of μ remaining unchanged at the true value. Each study will yield a different observation, X, and hence a different range (see Fig. 10.1). The range for any particular repetition will contain the true value of μ provided the true value is judged to be supported by the data X — in other words, provided that

$$-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2 > -1.353,$$

where μ now refers to the true value. Writing

$$z = \left(\frac{X-\mu}{\sigma}\right)$$

this condition is equivalent to $(z)^2$ being less than 2.706, and since $(z)^2$ has a chi-squared distribution this occurs with probability 0.90. Hence the coverage probability is 0.90.

Exercise 10.1. In a computer simulation of repetitions of a study in which a single observation is made from a Gaussian distribution with $\mu = 100$ and $\sigma = 10$, the first four repetitions produced the observations 104, 115, 82, and 92. Calculate the log likelihood ratio for $\mu = 100$ for each of these four observations. In which repetitions would the true value of μ have been supported?

The idea of coverage probability has allowed us to attach a frequentist probability, such as 0.90, to a range of parameter values, but we cannot say that the probability of the true value lying within the stated range is 0.90, because the stated range either does or does not include the true value. To avoid having to say precisely what is meant every time the probability for a range is reported, statisticians took refuge in an alternative word and professed themselves 90% *confident* that the true value lies in the reported interval. Not surprisingly the distinction between probability and confidence is rarely appreciated by scientists.

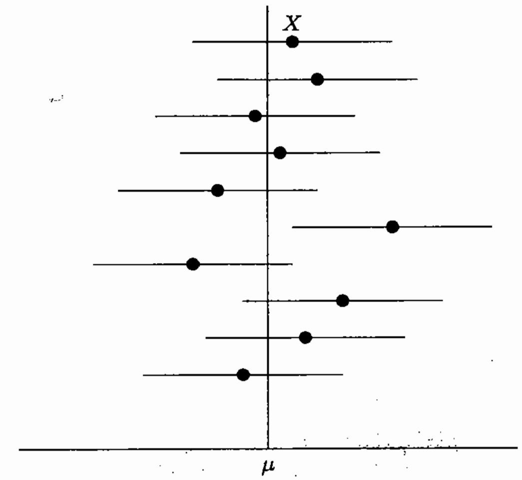

Fig. 10.1. Repeated studies and their supported ranges.

Exercise 10.2. Use tables of the chi-squared distribution to work out the cut-point for the log likelihood ratio which leads to a 95% coverage probability for the corresponding supported range, and give the formula for this range.

We have demonstrated the correspondence between the -1.353 cut-point for the log likelihood ratio and 90% coverage, but only for the case of the Gaussian log likelihood where the standard deviation is known. Fortunately the relationship also holds approximately for other log likelihoods such as the Bernoulli and Poisson. With increasing amounts of data these log likelihoods approach the quadratic shape of the Gaussian log likelihood and the coverage probability for the supported range based on the -1.353 cut-point is approximately 90%. In other words, if M is the most likely value of a parameter and S is the standard deviation of the Gaussian approximation to the likelihood, then the supported range

$$M \pm 1.645S$$

is also, at least approximately, a 90% confidence interval.

This raises the question of how much data is needed to use this approximate theory. For the Bernoulli likelihood, a reasonable guide is that the approximations are good if both D and $N - D$ are larger than 10, but can be misleading if either count is less than 5. In the Poisson case the observed number of events, D, should be larger than 10; there is no restriction on the number of person-years since this is irrelevant to the shape of the log likelihood curve. In Chapter 12 we discuss what to do when there are too few data to use the approximate theory.

The only likelihood for which the relationship between the supported range and the 90% confidence interval holds *exactly* is Gaussian likelihood, and even here we have made the assumption that the parameter σ is known. In the early years of this century it was shown that the practice of *estimating* the standard deviation using the data and thereafter pretending that this estimate is the true value, leads to intervals with *approximately* the correct coverage probability, providing N is large enough (more than 15).

The intervals we have chosen to present correspond to 90% confidence intervals but 95% intervals are more usually reported in the scientific literature. The routine use of 90% intervals in the epidemiological literature has recently been proposed on the grounds that they give a better impression of the range of plausible values. If you prefer 95% intervals these can be obtained by replacing 1.645 by 1.960 in the calculations.

★ 10.2 Subjective probability

The second approach to the problem of assigning a probability to a range of values for a parameter is based on the philosophical position that probability is a subjective measure of ignorance. The investigator uses probability as a measure of subjective *degree of belief* in the different values which the parameter might take. With this view it is perfectly logical to say that there is a probability of 0.9 that the parameter lies within a stated range.

Before observing the data, the investigator will have certain beliefs about the parameter value and these can be measured by *a priori* probabilities. Because they are subjective every scientist would be permitted to give different probabilities to different parameter values. However, the idea of scientific objectivity is not completely rejected. In this approach objectivity lies in the rule used to modify the *a priori* probabilities in the light of the data from the study. This is Bayes' rule and statisticians who take this philosophical position call themselves Bayesians.

Bayes' rule was described in Chapter 2, where it was used to calculate the probabilities of exposure given outcome from the probabilities of outcome given exposure. Once we are prepared to assign probabilities to parameter values, Bayes' rule can be used to calculate the probability of each value of a parameter (θ) given the data, from the probability of the data given the value of the parameter.

The argument is illustrated by two tree diagrams. Fig. 10.2 illustrates the direction in which probabilities are specified in the statistical model — given the choice of the value of the parameter, θ, the model tells us the probability of the data. The probability of any particular combination of data and parameter value is then the product of the probability of the parameter value and the probability of data given the parameter value. In

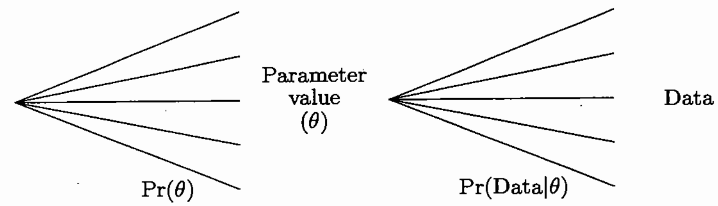

Fig. 10.2. From parameter value to data.

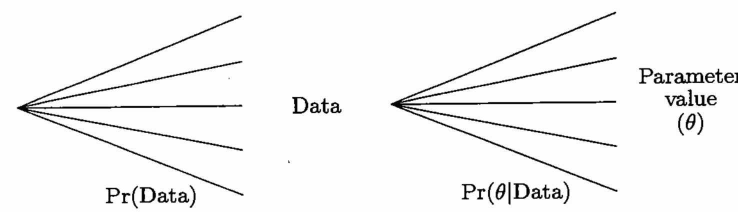

Fig. 10.3. From data to parameter value.

this product, the first term, $\Pr(\theta)$, represents the *a priori* degree of belief for the value of θ and the second term, $\Pr(\text{Data}|\theta)$, is the likelihood. Fig. 10.3 reverses the conditioning argument, and expresses the joint probability as the product of the overall probability of the data multiplied by the probability of the parameter given the data. This latter term, $\Pr(\theta|\text{Data})$, represents the *posterior* degree of belief in the parameter value once the data have been observed. Since the joint probability of data and parameter value is the same no matter which way we argue,

$$\Pr(\theta) \times \Pr(\text{Data}|\theta) = \Pr(\text{Data}) \times \Pr(\theta|\text{Data}),$$

so that

$$\Pr(\theta|\text{Data}) = \frac{\Pr(\theta) \times \Pr(\text{Data}|\theta)}{\Pr(\text{Data})}.$$

Thus elementary probability theory tells us how prior beliefs about the value of a parameter should be modified after the observation of data.

We shall now apply this idea to the problem of estimating the Gaussian mean, μ, given a single observation x. The likelihood for μ is

$$\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

If prior to observing x we believe that no value of μ is any more probable than any other, then the prior probability density does not vary with μ and the posterior probability density is proportional to the likelihood. Writing the likelihood as

$$\exp\left[-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2\right].$$

we see that after choosing the constant of proportionality to make the total probability for μ equal to 1, the posterior distribution for μ is a Gaussian distribution which has mean x and standard deviation σ. The 5 and 95 percentiles of the standard Gaussian distribution are -1.645 and 1.645 respectively so there is a 90% probability that μ lies in the range $x \pm 1.645\sigma$. This range is called a 90% credible interval.

When the quadratic approximation

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

is used for likelihoods such as the Bernoulli and Poisson, a similar argument shows that, provided the prior probability density for θ does not vary with θ, then the posterior distribution for θ is approximately Gaussian with mean M and standard deviation S. It follows that there is a 90% probability that θ lies in the range $M \pm 1.645S$.

It appears from this discussion that the frequentists and the Bayesians end up making very similar statements, differing only in their use of the words *confidence* and *probability*. But to achieve this agreement we have had to make the rather extreme assumption that *a priori* no one value of the parameter is more probable than any other. This is taking open mindedness too far and Bayesians would generally advocate the use of more realistic priors. When there is a large amount of data the posterior is more influenced by the likelihood than by the prior, and both approaches lead to similar answers regardless of the choice of prior. However, when the data are sparse, there can be serious differences between the two approaches. We shall return to this in Chapter 12.

Solutions to the exercises

10.1 When $x = 104$, the log likelihood ratio for $\mu = 100$ is

$$-\frac{1}{2}\left(\frac{104-100}{10}\right)^2 = -0.08.$$

For $x = 115, 82, 92$ the log likelihood ratio turns out to be -1.125, -1.62, and -0.32 respectively. Thus only for $x = 82$ is the support for the true value of μ less than the cut-off value of -1.353. In all other repetitions $\mu = 100$ is supported.

10.2 From tables of chi-squared, the value 3.841 is exceeded with probability 0.05, so

$$\left(\frac{x-\mu}{\sigma}\right)^2 > 3.841$$

with probability 0.05. The log likelihood ratio, which is minus one half of this quantity, is therefore less than

$$-0.5 \times 3.841 = -1.921$$

with probability 0.05. Thus the cut-point for the log likelihood ratio is -1.921.