12 ★ Small studies

In small studies the shape of the log likelihood for a parameter can be appreciably different from the quadratic shape of the Gaussian log likelihood and p-values and confidence intervals based on Gaussian approximations can then be misleading. It is conventional in such situations to report *exact* p-values and confidence intervals. In this chapter we will explain how these are conventionally calculated, while drawing attention to some serious difficulties.

12.1 Exact p-values based on the binomial distribution

Consider again the example in Chapter 11 concerning genetic linkage between a gene which renders a subject susceptible to a disease, and a marker gene. The test for linkage was based on the 16 sib pairs with two haplotypes in common and the 3 pairs with no haplotypes in common, so the log likelihood for Ω, the odds of having two haplotypes in common, is

$$16\log(\Omega) - 19\log(1+\Omega).$$

The most likely value of Ω is $16/3 = 5.33$ and the log likelihood takes its maximum value of -8.29 at this value of Ω. The value $\Omega = 1$ corresponds to no linkage and the log likelihood ratio for $\Omega = 1$ is therefore

$$16\log(1) - 19\log(1+1) - (-8.29) = -4.88.$$

The corresponding p-value is defined as the probability of obtaining a log likelihood ratio, less than -4.88, during many repetitions of the study in which $\Omega = 1$. In the last chapter this probability was obtained approximately from the chi-squared distribution; the problem now is to find its exact value.

Each new repetition of the study will give rise to a log likelihood ratio for $\Omega = 1$. To calculate this it is necessary to go through the same steps as for the split of 16:3. For example, a repetition in which the split was 10:9 gives a log likelihood for Ω of

$$10\log(\Omega) - 19\log(1+\Omega).$$

Table 12.1. A computer simulation and the binomial distribution

	Log likelihood ratio		Simulated	Binomial
Split	Two-sided	One-sided	frequency	probability
0:19	−13.17	0	0	0.000002
1:18	−9.25	0	1	0.000036
2:17	−6.78	0	17	0.000326
3:16	−4.88	0	112	0.001848
4:15	−3.39	0	512	0.007393
5:14	−2.22	0	1777	0.022179
6:13	−1.32	0	4519	0.051750
7:12	−0.67	0	9238	0.096107
8:11	−0.24	0	14523	0.144161
9:10	−0.03	0	18160	0.176197
10:9	−0.03	−0.03	18035	0.176197
11:8	−0.24	−0.24	14857	0.144161
12:7	−0.67	−0.67	9675	0.096107
13:6	−1.32	−1.32	5278	0.051750
14:5	−2.22	−2.22	2306	0.022179
15:4	−3.39	−3.39	750	0.007393
16:3	−4.88	−4.88	194	0.001848
17:2	−6.78	−6.78	38	0.000326
18:1	−9.25	−9.25	7	0.000036
19:0	−13.17	−13.17	1	0.000002

The most likely value for Ω is $10/9 = 1.11$ and the maximum value of the log likelihood is

$$10\log(1.11) - 19\log(1+1.11) = -13.14.$$

The log likelihood for $\Omega = 1$ based on this split is therefore

$$10\log(1) - 19\log(1+1) - (-13.14) = -0.03.$$

Exercise 12.1. Calculate the log likelihood ratio for $\Omega = 1$ when the split between the two outcomes is 15:4.

For a split such as 4:15, the log likelihood ratio depends on whether we regard the model as allowing values of Ω less than one. If not, then the best supported value of Ω given such a split is 1, and the log likelihood ratio is zero. In this case a one-sided p-value is appropriate.

The way the log likelihood ratio for $\Omega = 1$ depends on the observed split is shown in full in Table 12.1, for both two-sided and one-sided views

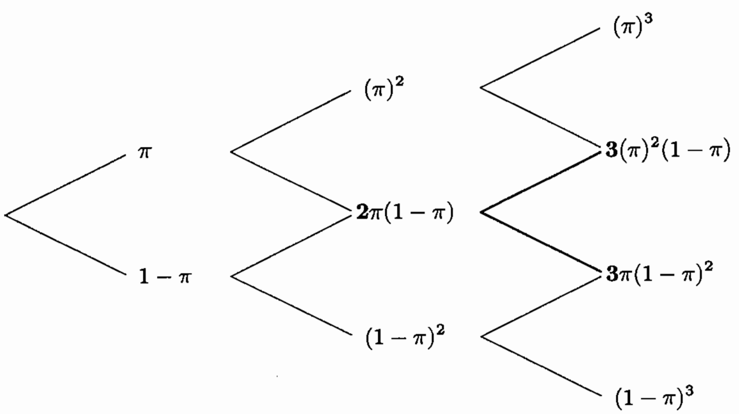

Fig. 12.1. Generating the binomial distribution.

of the problem.* In the two-sided case, the splits 2:17, 1:18, 0:19 and 17:2, 18:1, 19:0 all produce log likelihood ratios which are less than −4.88, and the splits 3:16 and 16:3 produce log likelihood ratios equal to −4.88. In the one-sided case, the splits 17:2, 18:1, and 19.0 give log likelihood ratios less than −4.88 and the split 16:3 gives a log likelihood ratio equal to −4.88. To find the p-values exactly we need to find the probabilities of the different splits when $\Omega = 1$.

One way of calculating these p-values is to use a Monte Carlo approach similar to that described in Chapter 11. A computer program is written which splits the 19 sib pairs between the two outcomes with odds 1, and repeats the process (say) 100 000 times. The result of doing this is shown in the third column of Table 12.1. Out of 100 000 repetitions of the study, none produced the split 0:19, one produced the split 1:18, 17 produced the split 2:17, and so on. The probabilities of the different splits are therefore estimated by the computer to be 0.00000, 0.00001, 0.00017, and so on.

As in the case of the Gaussian mean, the probabilities can also be worked out theoretically, in this case using the *binomial distribution*. Fig. 12.1 illustrates the derivation of the binomial distribution. The first level of branching represents the possible outcomes of the first observation, the upper branch indicating failure (with probability π) and the lower branch indicating survival (with probability $1-\pi$). The second level of branching represents the outcome of the second observation. The probability that both subjects fail is $(\pi)^2$ and the probability that both survive is $(1-\pi)^2$; the remaining two possibilities both have one failure and one survivor and, since we do not need to differentiate between these, the branches are allowed to merge, with a total probability of $2\pi(1-\pi)$. The diagram continues with the inclusion of a third observation. The probability that all three observations are failures is now $(\pi)^3$ and that all three are survivors is $(1-\pi)^3$. The remaining probabilities correspond to 2:1 and 1:2 splits of failures to survivors and have probabilities $3(\pi)^2(1-\pi)$ and $3\pi(1-\pi)^2$ respectively, the multiplier 3 arising because each of these points represents the merging of 3 paths through the tree.

*When calculating these log likelihood ratios when the splits are 0:19 or 19:0, note that the expression $0\log(0)$ takes the value 0.

Exercise 12.2. Continue the diagram to generate the probabilities for all possible splits of $N = 4$ observations and also for $N = 5$.

When this process is continued it leads to the general result that the probability that N observations split as D failures and $N - D$ survivors is

$$\mathrm{C}(D,N)(\pi)^D(1-\pi)^{N-D}.$$

where $\mathrm{C}(D,N)$, the number of ways of selecting D objects from N, is 1 when $D = 0$ or $D = N$ and

$$\frac{N \times (N-1) \times \cdots \times (N-D+1)}{D \times (D-1) \cdots \times 2 \times 1}$$

otherwise. Binomial probabilities may easily be calculated by computer, and tables are available for values of N and D up to about 20.

The binomial distribution with $N = 19$ and $\pi = 0.5$ is shown in the fourth column of Table 12.1. A comparison between the third and fourth columns of this table shows that the values estimated by the Monte Carlo method are quite close to the correct values, particularly in the centre of the distribution.

One of the areas of dispute when defining an exact p-value is whether to define this as the probability of obtaining a log likelihood ratio less than −4.88 or less than *or equal* to −4.88. This difficulty does not arise with the Gaussian log likelihood because the probability of any one *precise* outcome is zero, but it does arise here; in the two-sided case the splits 3:16 and 16:3 both give rise to the observed log likelihood ratio of −4.88 and have probabilities 0.001848. If these splits are excluded, the two-sided p-value is

$$0.000002 + 0.000036 + 0.000326 + 0.000002 + 0.000036 + 0.000326$$

which adds up to 0.000728. If these splits are included, two further contributions of 0.001848 must be included and the two-sided p-value is 0.004424. Conventionally, splits giving rise to the observed log likelihood ratio are included, but there are arguments in favour of including only one half of the probability for these splits. This course of action gives the *mid-p* value. In our example the mid-p value is 0.002576.

Table 12.2. Log likelihood ratios and probabilities ($N = 27$, $\pi = 0.25$)

Split	LLR	Probability	Split	LLR	Probability
0:27	-7.767	0.000423	14:13	-4.452	0.001775
1:26	-4.589	0.003810	15:12	-5.699	0.000513
2:25	-2.835	0.016509	16:11	-7.096	0.000128
3:24	-1.645	0.045858	17:10	-8.647	0.000028
4:23	-0.836	0.091716	18:9	-10.357	0.000005
5:22	-0.323	0.140632	19:8	-12.233	0.000001
6:21	-0.057	0.171883	20:7	-14.288	
7:20	-0.006	0.171883	21:6	-16.536	
8:19	-0.149	0.143236	22:5	-18.999	
9:18	-0.469	0.100796	23:4	-21.709	
10:17	-0.956	0.060477	24:3	-24.716	
11:16	-1.603	0.031155	25:2	-28.103	
12:15	-2.403	0.013847	26:1	-32.054	
13:14	-3.353	0.005326	27:0	-37.430	

If these arguments are repeated for one-sided p-values it can be seen that, whichever convention is adopted, the one-sided p-value is half of the two-sided value. This is not generally true and is only the case here because of the symmetry of the binomial distribution in this case. This in turn derives from the fact that the null value of Ω is 1, corresponding to $\pi = 0.5$. For a test of the null value $\pi = 0.25$, the relationship between one- and two-sided p-values is not as simple.

Exercise 12.3. In the genetic linkage example, one of the tests for linkage compares the observed split of the 27 sib pairs into 16 with two haplotypes in common and 11 with one or zero in common with the probabilities 0.25 and 0.75 under the hypothesis of no linkage. The log likelihood ratios and probabilities corresponding to the different possible splits are shown in Table 12.2 (probabilities less than 0.000001 are omitted). Find the exact two-sided p-value for the hypothesis of no linkage.

In this exercise the probability distribution for the different splits is not symmetric and the one-side p-value cannot be obtained by halving the two-sided value. In such situations there is no general agreement about how two-sided p-values should be calculated, because there is no general agreement about how to compare extremeness of splits at opposite ends of the distribution. We have chosen to measure extremeness in terms of the log likelihood ratio, but other criteria are also used and lead to different two-sided p-values.

Table 12.3. Log likelihood ratios and probabilities ($\eta = 0.25$)

Cases	LLR	Probability
0	-0.25	0.778801
1	-0.64	0.194700
2	-2.41	0.024338
3	-4.70	0.002028
4	-7.34	0.000127
5	-10.23	0.000006
6	-13.32	0.000000
etc.		

12.2 The Poisson distribution

When the population at risk, N, is very large and the probability of failure, π, is very small, the binomial distribution takes on a very simple form, called the Poisson distribution:

$$\frac{1}{D!}(\eta)^D \exp(-\eta)$$

where $D!$ denotes D *factorial*

$$D \times (D-1) \cdots \times 2 \times 1$$

and $\eta = N\pi$. The same is approximately true of the number of failures in a cohort subject to rate λ and with Y person-years of observation. Providing we can regard Y, at least approximately, as a fixed constant then the probability of D failures is given by the Poisson distribution with $\eta = \lambda Y$.

The main use of the Poisson distribution is to calculate the p-value corresponding to the null hypothesis which states that the rate in the study cohort is the same as a reference rate, λ_{R}. The null value of η is $E = \lambda_{\mathrm{R}} Y$, the expected number of cases. Given $\eta = E$, the Poisson distribution tells us the probability for any value of D. The idea extends to the case where the expected number of cases is calculated taking account of variation of rates with time.

To illustrate the use of the Poisson distribution, we return to our example of leukaemia surrounding a nuclear reprocessing plant (Exercise 11.8). In that case the expected number of failures was 0.25 and the Poisson probabilities for each possible value of D are shown in Table 12.3. The table also lists the corresponding values of the log likelihood ratio for the null hypothesis, which we showed in Chapter 11 to be given by the expression

$$-D \log\left(\frac{D}{E}\right) + (D - E)\,.$$

Table 12.4. Definition of the exact confidence interval

Cases	Probability	
	$(\eta = 1.3663)$	$(\eta = 9.1535)$
0	0.25505	0.00011
1	0.34847	0.00097
2	0.23806	0.00443
3	0.10842	0.01353
4	0.03703	0.03096
5	0.01012	0.05668
6	0.00230	0.08647
7	0.00045	0.11307
8	0.00001	0.12938
etc.		

The observed number of cases of leukaemia was 4 and the corresponding log likelihood ratio -7.34. To find the p-value we add the probabilities of all values of D with log likelihood ratio less than or equal to -7.34 :

$$0.000127 + 0.000006 + 0.000000 = 0.000133.$$

Note that, in this case, there is no difference between the one- and two-sided p-values.

12.3 Exact confidence intervals

An *exact confidence interval* for a parameter is defined in terms of exact p-values. The lower limit of the 90% interval for a parameter θ is found by searching for the null value, $\theta_{\oslash}$, whose p-value is exactly 0.05. Here, the *one-sided* p-value which assumes that $\theta \geq \theta_{\oslash}$ is used. The upper limit is defined similarly, save for the fact that the reverse one-sided p-value is used, that is the p-value under the assumption $\theta \leq \theta_{\oslash}$. The search for these values must be carried out by computer and is laborious, although computational methods have been considerably improved in recent years.

Table 12.4 illustrates the idea of exact confidence intervals using the leukaemia data discussed above. Poisson distributions are shown for two values of $\eta = \theta E$. Both values give one-sided p-values of approximately 0.05 when the observed number of cases is 4, since

$$0.03703 + 0.01012 + 0.00230 + 0.00045 + 0.00001 = 0.04991$$

and

$$0.00011 + 0.00097 + 0.00443 + 0.01353 + 0.03096 = 0.05000.$$

Thus values of θE smaller than 1.3663 and values larger than 9.1535 have one-sided p-values smaller than 0.05. Since $E = 0.25$, the exact confidence interval for θ lies between $1.3663/0.25 = 5.465$ and $9.1535/0.25 = 36.614$.

Exact confidence intervals are only exact in the sense that they are derived from exact p-values. They do *not* necessarily have coverage probabilities exactly equal to 0.90. For the Gaussian mean, μ, when the standard deviation is known, an exact 90% confidence interval does have a coverage probability of exactly 0.90, but for parameters of other models this is often not the case. This is because, in these cases, the coverage probability depends on the unknown true value of the parameter. Thus, exact confidence intervals are not exact in any *scientifically useful* sense.

This observation, taken together with the fact that there are several different ways in which exact p-values may be defined, lead us to doubt the practical usefulness of exact methods. Instead we would argue that, since it is the *log likelihood* which measures the support for different values of the parameter, scientific papers should aim to communicate the log likelihood accurately and concisely. For large studies Gaussian approximations allow us to communicate the log likelihood curve using only M and S, the most likely value and a standard deviation. For small studies it might be necessary to report the log likelihood in greater detail.

12.4 A Bayesian approach

The Bayesian approach goes further and uses the likelihood to update a prior distribution for the parameter into a posterior distribution, using Bayes' rule as described in Chapter 10. No new difficulties are introduced by the fact that a study is small, apart from the inevitable consequence that the information in the likelihood will also be small, so the posterior distribution will not be much different from the prior distribution. This means that conclusions depend more upon our prior beliefs about the parameter in a small study than they would in a large study.

Similar answers to those yielded by the classical exact approach can be obtained using Bayesian arguments if it is assumed *a priori* that we are completely ignorant about the value of the parameter. Such an assumption is called a *vague* prior belief and holds that no value of the parameter is any more probable than any other value, so that the prior distribution is flat. One difficulty is that a flat prior for a parameter θ is not flat with respect to $\log(\theta)$, so a flat prior for θ and a flat prior for $\log(\theta)$ lead to different posterior beliefs.

This may be illustrated by our example of leukaemia in the neighbourhood of a nuclear plant, where the observed number of cases was $D = 4$ while the expected number from national rates was $E = 0.25$. It is conventional to compare rates in the study population with reference rates by the ratio of observed to expected cases, in this case $4/0.25 = 16.0$. This

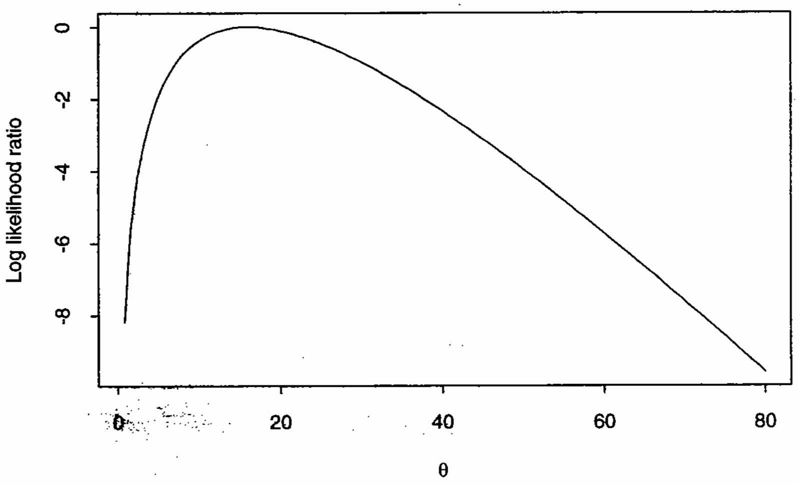

Fig. 12.2. Log likelihood for the leukaemia data ($D = 4,\ E = 0.25$).

Table 12.5. Posterior distributions for θ for three vague priors

Prior	Posterior probability distribution for θ			
(flat with		90% probability interval		Probability
respect to)	Mean	Lower limit	Upper limit	$\theta < 1.0$
$\log(\theta)$	16.0	5.5	31.0	0.000133
θ	20.0	7.9	36.6	0.000007
$\sqrt{\theta}$	18.0	6.6	33.8	0.000030

may be regarded as the most likely value of the parameter, θ, of the Poisson probability model with $\eta = \theta E$. The parameter θ may be regarded as an index of mortality in the cohort, relative to national rates.† The log likelihood for θ remains Poisson in form and is plotted in Fig. 12.2.

In Bayesian statistics we start with the prior distribution for θ and multiply it by the likelihood to obtain the posterior distribution. The posterior distribution is then used to calculate the (subjective) probability that θ lies in a given range. Table 12.5 summarizes the results of such calculations for the leukaemia data for three different prior belief distributions — each of them vague in some sense.

According to these analyses, it is almost certain that there is an effect of living near Sellafield and the magnitude of this effect, as measured by the mean of the posterior distribution, is very large. Unfortunately, these conclusions are not scientifically credible. Ratios of observed to expected cases of 5 are extremely rare in epidemiology when the numbers of cases are large. This is true even for studies of heavily exposed versus completely unexposed groups, and we would expect much smaller ratios for groups defined only in terms of area of residence. That 5.5 is the *lowest* plausible value for θ does not seem to be a reasonable conclusion.

†A fuller discussion of this model will be encountered in Chapter 15.

Table 12.6. Posterior distributions for θ for three *realistic* priors

Prior	Posterior probability distribution for θ			
belief		90% probability interval		Probability
(90% limits)	Mean	Lower limit	Upper limit	$\theta < 1.0$
0.3–2.0	2.00	0.97	3.33	0.06
0.5–1.6	1.37	0.83	2.02	0.15
0.7–1.3	1.15	0.82	1.52	0.25

The problem lies with the choice of prior distributions. Prior to seeing these data, no epidemiologist would seriously have believed that $\theta = 1000$ and $\theta = 2$ are equally probable. Bayesian analyses with more realistic prior distributions give more sensible answers. Table 12.6 shows the results of analysis for three epidemiologists with more realistic prior beliefs. All these prior distributions have mean 1.0, indicating that the epidemiologists have no prior expectation of elevated rather than reduced risk of disease, but they do differ in the *range* of values of θ, around 1.0, which they consider believable.‡

Exercise 12.4. With which of the three epidemiologists would you most closely identify yourself?

The conclusions of the three epidemiologists after seeing the data still differ substantially. All tend towards the belief that there is an elevated risk but the extent of the increase is now a lot less than before. The Bayesian approach has therefore shown that such a small study as this cannot lead to identical beliefs within the scientific community. The posterior distribution is too influenced by prior belief and too little by the data.

‡For mathematical convenience only, all three distributions have been chosen from the chi-squared family.

Solutions to the exercises

12.1 For a 15:4 split, the log likelihood is

$$15\log(\Omega) - 19\log(1+\Omega),$$

which takes its maximum value when $\Omega = 15/4 = 3.75$. The values of the log likelihood when Ω takes on values of 3.75 and 1 are, respectively

$$\begin{aligned} 15\log(3.75) - 19\log(4.75) &= -9.778, \\ 15\log(1) - 19\log(2) &= -13.170. \end{aligned}$$

The log likelihood ratio at $\Omega = 1$ is the difference between these, which is -3.392.

12.2 Fig. 12.3 shows the extension of the diagram from $N = 3$ to $N = 4$ and $N = 5$. The numbers in boldface represent the values of $C(D, N)$.

12.3 Table 12.2 shows that when the observed data are a 16:11 split, the log likelihood ratio for $\pi = 0.25$ is -7.096. The two-sided p-value is the sum of the probabilities for those outcomes leading to log likelihood ratios at least this small, that is

$$\begin{array}{c} 0.000128 + 0.000028 + 0.000005 + 0.000001 \\ +0.000423 \end{array} = 0.000585.$$

12.4 There is no solution to this exercise!

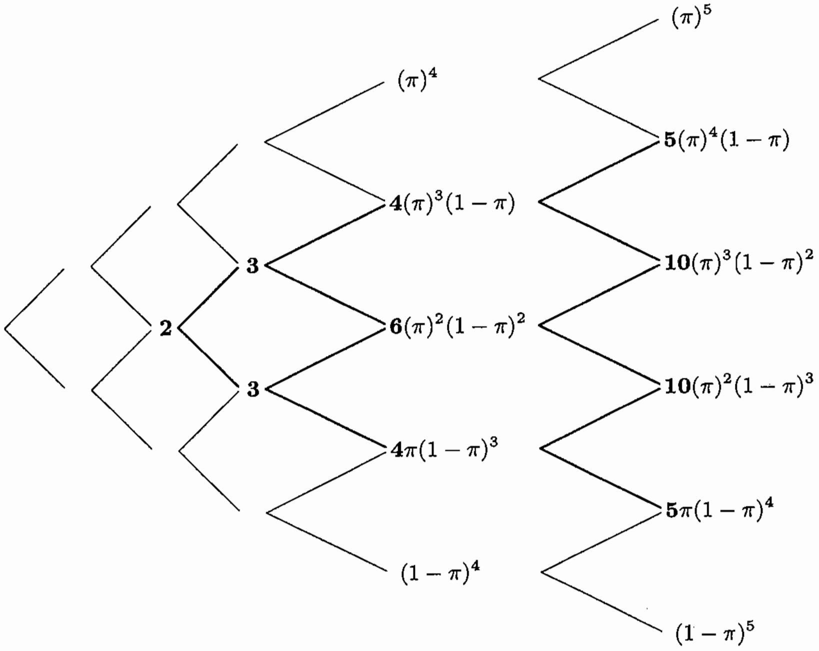
$(\pi)^5$
$(\pi)^4$
$\mathbf{5}(\pi)^4(1-\pi)$
$\mathbf{4}(\pi)^3(1-\pi)$
3
$\mathbf{10}(\pi)^3(1-\pi)^2$
2
$\mathbf{6}(\pi)^2(1-\pi)^2$
$\mathbf{10}(\pi)^2(1-\pi)^3$
3
$\mathbf{4}\pi(1-\pi)^3$
$\mathbf{5}\pi(1-\pi)^4$
$(1-\pi)^4$
$(1-\pi)^5$

Fig. 12.3. Binomial distributions with $N = 4$ and $N = 5$.