## 31. The method of maximum likelihood.

In Lectures 19 and 21, we fitted continuous distributions to samples of discrete data by minimizing discrepancy between the continuous c.d.f. of the model and the discrete c.d.f. of the sample. The result was a model of the sample distribution. Usually, however, what we really want is a model of the population from which the sample is taken. Here we explain how to obtain such a fit between model and population (as opposed to between model and sample).

A simple example will serve to introduce the method. Accordingly, let X be the winner of a contest between two male toads. If X = 1 then the smaller toad wins, if X = 2 then the larger one wins. So X is a discrete random variable with sample space {1, 2}. Contests in nature are usually won by the larger animal. Moreover, the greater the discrepancy in size between animals, the more likely it is that the larger one wins. Thus if $p = \mathrm{Prob}(X = 1)$ is the probability of victory by the smaller animal (so that $1 - p$ the probability of victory by the larger animal), we can safely assume that $p \le 1/2$ and that p is a decreasing function of size discrepancy. In a study of dominance hierarchies, Mesterton-Gibbons and Dugatkin (1995) assumed that

$$p = \frac{1}{2}e^{-\alpha z} \tag{31.1}$$

where $\alpha$ is a positive parameter and z ($\ge 0$) is the difference in size. We know from Lecture 22 that (1) makes p decrease with z. Moreover, if z is extremely large, then p is virtually zero; whereas if z is very close to zero, then p is virtually 0.5, and the smaller animal has an almost even chance of winning. Thus p has qualitatively the right dependence on z. But how do we obtain a value for $\alpha$?

To illustrate, consider 41 fights that Davies and Halliday (1977) staged between males of the species *Bufo bufo.* Thirteen fights were won by the male with smaller snout-vent length, and 28 were won by the larger male. In every case, the difference in snout-vent length was 10 mm. Thus, from (1), the probability of victory by a smaller male was $e^{-10\alpha}/2$, and the probability of victory by a larger one was $1 - e^{-10\alpha}/2$. The probability of the observed sequence of wins and losses by smaller animals is the product of the 41 individual probabilities. That is, if L denotes the probability of the observed outcome, then

$$L(\alpha) = \left(\frac{1}{2}e^{-10\alpha}\right)^{13}\left(1 - \frac{1}{2}e^{-10\alpha}\right)^{28}. \tag{31.2}$$

The function L defined by (2) is called the **likelihood** (of the observed outcome, given that the parameter had value $\alpha$).

The graph of L is plotted in Figure 1. You can see that $L(\alpha)$ is always extremely small, being largest at $\alpha = 0.0455$, where it assumes the scarcely enormous value of $0.754 \times 10^{-11}$. Yet $L(\alpha)$ is the probability of an event that actually occurred! An event that occurred cannot have had a thoroughly negligible probability, so we choose for $\alpha$ the value that makes $L(\alpha)$ least neglible, i.e., the **maximum likelihood estimate** or **MLE**, which we denote by $\hat{\alpha}$. Here, for example, $\hat{\alpha} = 0.0455$.

In practice, MLEs are often obtained not from the likelihood function itself, but instead from the **loglikelihood** function Q defined by

$$Q(\alpha) = \ln(L(\alpha)). \tag{31.3}$$

Because the logarithm is a strictly increasing function, if L increases to a maximum before decreasing again, then Q must do exactly the same. Thus the maximum of Q must always correspond to the maximum of L, as illustrated by Figure 1(b). But the maximum of Q is often easier to find than the maximum of L.

In the case of *B. bufo,* for example, from (2)-(3) we have

$$\begin{aligned} Q(\alpha) &= 13\ln\left(\tfrac{1}{2}e^{-10\alpha}\right) + 28\ln\left(1 - \tfrac{1}{2}e^{-10\alpha}\right) \\ &= -13\ln(2) - 130\alpha + 28\ln\left(1 - \tfrac{1}{2}e^{-10\alpha}\right), \end{aligned} \qquad (31.4)$$

implying

$$Q'(\alpha) = \frac{5e^{-10\alpha}}{1 - \tfrac{1}{2}e^{-10\alpha}} - 130 \qquad (31.5)$$

(Exercise 1). It is now straightforward to show that $Q'(\alpha) = 0$ where $e^{-10\alpha} = 26/41$, implying that the MLE of $\alpha$ is

$$\hat{\alpha} = \tfrac{1}{10}\ln(41/26). \qquad (31.6)$$

So far, we have introduced the maximum likelihood method only for discrete distributions, but it works for continuous ones also. Suppose that the sequence $\{x_k\} = \{x_1, x_2, \ldots, x_N\}$ is a sample of size N from the distribution of a random variable X with p.d.f. f. Now, in practice, observing $x_k$ means observing x such that

$$x_k - \frac{h}{2} \leq x \leq x_k + \frac{h}{2}, \qquad (31.7)$$

where h is small and due to measurement error. So, in practice, the probability of observing $x_k$ becomes $\mathrm{Prob}(x_k - h/2 \leq X \leq x_k + h/2) = \mathrm{Area}(f, [x_k - h/2, x_k + h/2])$. For sufficiently small h, this area approximately equals $hf(x_k)$.[1] The probability of observing $x_1, x_2, \ldots, x_N$ in turn is the product of the N individual probabilities. It is therefore approximately equal to $hf(x_1)$ multiplied by $hf(x_2)$ multiplied by $hf(x_3)$, and so on, all the way down to $hf(x_N)$, or

$$h^N f(x_1)f(x_2)f(x_3)\ldots f(x_N). \qquad (31.8)$$

It is convenient, however, to have a more compact notation for the product of all the $f(x_k)$ between $k = 1$ and $k = N$, and so we *define*

$$\prod_{k=1}^{N} f(x_k) = f(x_1)f(x_2)f(x_3)\ldots f(x_N), \qquad (31.9)$$

by analogy with $\Sigma$ for sum in Lecture 2. Then, by (8), the probability of observing $x_1, x_2, \ldots, x_N$ in turn is approximately equal to $h^N L$, where

$$L = \prod_{k=1}^{N} f(x_k) \qquad (31.10)$$

is called the likelihood, as before. The probability $h^N L$ is always extremely small, by virtue of h being so small, yet it corresponds to an event that actually happened; so the event cannot have had a thoroughly negligible probability. Hence any parameter s on which L depends is chosen to make $f(x_1)f(x_2)f(x_3)\ldots f(x_N)$ least negligible. As before, this value of s is the maximum likelihood estimate or MLE, which we denote by $\hat{s}$.

For example, suppose that we want to fit the exponential distribution defined by

---

[1] More precisely, the area is $hf(x_k) + O[h]$. Similarly, the probability in (8) more precisely equals $h^N f(x_1)f(x_2)f(x_3)\ldots f(x_N)\{1+O[h]\}$

$$f(x) = \frac{1}{s}e^{-x/s}, \tag{31.11}$$

not to Lecture 19's finite sample of prairie-dog lifetimes, but rather to the population of prairie-dog lifetimes from which the sample was drawn. Here N = 545, and so from (10)-(11) the likelihood is

$$\begin{aligned} L(s) &= \prod_{k=1}^{N} f(x_k) = \prod_{k=1}^{N} \frac{1}{s}e^{-x_k/s} \\ &= \frac{1}{s}e^{-x_1/s}\cdot\frac{1}{s}e^{-x_2/s}\cdot\frac{1}{s}e^{-x_3/s}\cdots\frac{1}{s}e^{-x_N/s} \\ &= \frac{1}{s^N}e^{-x_1/s}\cdot e^{-x_2/s}\cdot e^{-x_3/s}\cdots e^{-x_N/s} \\ &= s^{-N}e^{-\{x_1+x_2+x_3+\cdots+x_N\}/s}. \end{aligned} \tag{31.12}$$

The loglikelihood is therefore

$$\begin{aligned} Q(s) &= \ln(L(s)) = \ln\left(s^{-N}\right) + \ln\left(e^{-\{x_1+x_2+x_3+\cdots+x_N\}/s}\right) \\ &= -N\ln(s) - \frac{1}{s}\{x_1+x_2+x_3+\cdots+x_N\} \\ &= -N\ln(s) - s^{-1}\sum_{k=1}^{N} x_k \\ &= -N\left\{\ln(s) + \frac{\bar{x}}{s}\right\}, \end{aligned} \tag{31.13}$$

where

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k \tag{31.14}$$

is called the **sample mean**.

| YEAR OF DEATH | FREQUENCY |
|---|---|
| 1 | 312 |
| 2 | 101 |
| 3 | 74 |
| 4 | 35 |
| 5 | 12 |
| 6 | 8 |
| ≥ 7 | 0 |

Table 31.1 Prairie-dog lifetimes

You may recall from Lecture 19 that the individual prairie-dog lifetimes were not available. Nevertheless, from Table 1, we can estimate the sample mean by supposing that the 312 lifetimes of a year or less correspond to $x_k = 0.5$, the 101 lifetimes of between 1 year and two years to $x_k = 1.5$, and so on. We thus estimate

$$\bar{x} = \frac{1}{545}\sum_{k=1}^{545} x_k = \tfrac{1}{545}\{312\times0.5+101\times1.5+74\times2.5+35\times3.5+12\times4.5+8\times5.5\}$$

$$= \frac{713}{545} = 1.308 \tag{31.15}$$

years. Then, with N = 545, we have

$$Q(s) = -545\ln(s) - \frac{713}{s} \tag{31.16}$$

from (13). The function Q is graphed in Figure 2(a), from which you can see that the maximum occurs near s = 1.3. To obtain the MLE precisely, we use

$$Q'(s) = -N\left\{\frac{1}{s} - \frac{\bar{x}}{s^2}\right\} = \frac{N}{s^2}(\bar{x} - s) \tag{31.17}$$

(Exercise 2), from which the maximum occurs at $s = \bar{x}$, because $Q'(\bar{x}) = 0$ with $Q'(s) > 0$ if $s < \bar{x}$ and $Q'(s) < 0$ if $s > \bar{x}$. So the MLE is $\hat{s} = \bar{x} = 1.308$. Both p.d.f. defined by (11) with $s = \hat{s}$ and c.d.f. defined by $F(x) = 1 - e^{-x/\hat{s}}$ are shown solid in Figure 2, with the corresponding curves from Lecture 19 shown dotted. The results are virtually indistinguishable, in this particular case.

In general, however, we expect the distribution that best fits a sample to differ from the distribution that best fits the population which yielded the sample, because rare individuals must be accounted for in the population but are unlikely to appear in the sample. In particular, if rare individuals are larger, then we expect population size distributions to have somewhat thicker tails (hence lower peaks) than sample size distributions. To illustrate, we fit the distribution with p.d.f. defined by

$$f(x) = \frac{2x}{s^2}e^{-x^2/s^2}, \tag{31.18}$$

not to Lecture 21's finite sample of 22 rat pupil radii, but rather to the population of rat pupil radii from which the sample was drawn. Here N = 22, and so from (10) and (18) the likelihood is

$$\begin{aligned} L(s) &= \prod_{k=1}^{N} f(x_k) = \prod_{k=1}^{N} \frac{2x_k}{s^2}e^{-x_k^2/s^2} \\ &= \frac{2x_1}{s^2}e^{-x_1^2/s^2}\cdot\frac{2x_2}{s^2}e^{-x_2^2/s^2}\cdot\frac{2x_3}{s^2}e^{-x_3^2/s^2} \ldots \frac{2x_N}{s^2}e^{-x_N^2/s^2} \\ &= \frac{2^N}{s^{2N}}x_1x_2x_3 \ldots x_N e^{-x_1^2/s^2}\cdot e^{-x_2^2/s^2}\cdot e^{-x_3^2/s^2} \ldots e^{-x_N^2/s^2} \\ &= 2^N s^{-2N}\left(\prod_{k=1}^{N} x_k\right)e^{-\{x_1^2+x_2^2+x_3^2+\ldots+x_N^2\}/s^2}. \end{aligned} \tag{31.19}$$

The loglikelihood is therefore

$$\begin{aligned} Q(s) &= \ln(L(s)) = \ln\left(2^N\right) + \ln\left(s^{-2N}\right) + \ln\left(\prod_{k=1}^{N} x_k\right) + \ln\left(e^{-\{x_1^2+x_2^2+x_3^2+\ldots+x_N^2\}/s^2}\right) \\ &= N\ln(2) - 2N\ln(s) + \sum_{k=1}^{N}\ln(x_k) - \frac{1}{s^2}\left\{x_1^2+x_2^2+x_3^2+\ldots+x_N^2\right\} \\ &= N\ln(2) + \sum_{k=1}^{N}\ln(x_k) - 2N\ln(s) - s^{-2}\sum_{k=1}^{N}x_k^2. \end{aligned} \tag{31.20}$$

| RADIUS | FREQUENCY | RADIUS | FREQUENCY |
|---|---|---|---|
| 0.25 mm | 2 | 1.0 mm | 3 |
| 0.375 mm | 4 | 1.25 mm | 1 |
| 0.5 mm | 7 | 1.5 mm | 1 |
| 0.75 mm | 3 | 1.75 mm | 1 |

Table 31.2 Pupil radii from an experiment with laboratory rats

The requisite rat pupil data from Lecture 21 are recorded in Table 2, from which

$$\sum_{k=1}^{22} x_k^2 = 2\times 0.25^2 + 4\times 0.375^2 + 7\times 0.5^2 + 3\times 0.75^2 + 3\times 1.0^2 + 1\times 1.25^2 + 1\times 1.5^2 + 1\times 1.75^2 = 14. \tag{31.21}$$

Thus, from (20) with N = 22, we have

$$Q(s) = 4.02648 - 44\ln(s) - \frac{14}{s^2}. \tag{31.22}$$

This loglikelihood function is graphed in Figure 3(a), where the maximum occurs near s = 0.8. To obtain the MLE precisely, we use

$$Q'(s) = -\frac{2N}{s} + 2s^{-3}\sum_{k=1}^{N} x_k^2 = \frac{2}{s^3}\left\{\sum_{k=1}^{N} x_k^2 - Ns^2\right\}, \tag{31.23}$$

(Exercise 2), from which the maximum occurs at

$$\hat{s} = \sqrt{\frac{1}{N}\sum_{k=1}^{22} x_k^2}, \tag{31.24}$$

because $Q'(\hat{s}) = 0$ with $Q'(s) > 0$ if $s < \hat{s}$ and $Q'(s) < 0$ if $s > \hat{s}$. So the MLE is $\hat{s} = \sqrt{7/11}$ = 0.798 mm. Both p.d.f. defined by (18) with $s = \hat{s}$ and c.d.f. defined by $F(x) = 1 - e^{-x^2/\hat{s}^2}$ are shown solid in Figure 3, with the corresponding curves from Lecture 21 shown dotted.

The distributions defined by (11) and (18) are special cases of the Weibull with shape parameters c = 1 and c = 2, respectively. In the general case we have

$$f(x) = \frac{cx^{c-1}}{s^c} e^{-x^2/s^2} \tag{31.25}$$

where, in view of Lecture 22, c need not be an integer. Thus, in general, the Weibull likelihood and loglikelihood functions are not ordinary functions of s; rather, they are bivariate functions of c and s. Specifically, from (25) we have

$$\begin{aligned} L(c,s) &= \prod_{k=1}^{N} f(x_k) = \prod_{k=1}^{N} \frac{cx_k^{c-1}}{s^c} e^{-x_k^c/s^c} \\ &= \frac{cx_1^{c-1}}{s^c} e^{-x_1^c/s^c} \cdot \frac{cx_2^{c-1}}{s^c} e^{-x_2^c/s^c} \cdot \frac{cx_3^{c-1}}{s^c} e^{-x_3^c/s^c} \cdots \frac{cx_N^{c-1}}{s^c} e^{-x_N^c/s^c} \\ &= \frac{c^N}{s^{cN}} x_1^{c-1} x_2^{c-1} x_3^{c-1} \cdots x_N^{c-1} e^{-x_1^c/s^c} \cdot e^{-x_2^c/s^c} \cdot e^{-x_3^c/s^c} \cdots e^{-x_N^c/s^c} \\ &= c^N s^{-cN} \left(\prod_{k=1}^{N} x_k^{c-1}\right) e^{-\{x_1^c + x_2^c + x_3^c + \cdots + x_N^c\}/s^c} \end{aligned} \tag{31.26}$$

and

$$\begin{aligned} Q(c,s) &= \ln(L(c,s)) = \ln\left(c^N\right) + \ln\left(s^{-cN}\right) + \ln\left(\prod_{k=1}^{N} x_k^{c-1}\right) + \ln\left(e^{-\{x_1^c + x_2^c + x_3^c + \ldots + x_N^c\}/s^c}\right) \\ &= N\ln(c) - cN\ln(s) + \sum_{k=1}^{N}\ln(x_k^{c-1}) - \frac{1}{s^c}\left\{x_1^c + x_2^c + x_3^c + \ldots + x_N^c\right\} \\ &= N\ln(c) + \sum_{k=1}^{N}\ln(x_k^{c-1}) - cN\ln(s) - s^{-c}\sum_{k=1}^{N} x_k^c. \end{aligned} \tag{31.27}$$

in place of (19)-(20).

As before, any parameter on which L (and hence Q) depends is chosen to make $f(x_1)f(x_2)f(x_3)\ldots f(x_N)$ least negligible, but now there are two parameters instead of one. Let the MLEs of c and s, i.e., the values of c and s that maximize Q(c, s), be denoted by $\hat{c}$ and $\hat{s}$, respectively. In other words, $Q(\hat{c}, \hat{s})$ is the maximum loglikelihood, implying that $L(\hat{c}, \hat{s})$ is the maximum likelihood. Then, for the rat pupil data in Figure 2, the summit of the hilltop in Figure 4(a), where altitude represents loglikelihood, has coordinates $(\hat{c}, \hat{s}, Q(\hat{c}, \hat{s}))$. But it is difficult to estimate $\hat{c}$ and $\hat{s}$ from this diagram, and so Figure 4(b) shows the corresponding contour map, where the horizontal coordinates of points with the same altitude are joined by continous curves. From innermost to outermost, the three closed contours correspond to $Q = -7.98$, $Q = -8$ and $Q = -8.1$, respectively; thereafter, Q decreases by 0.1, so that, e.g., the contour that touches the right-hand edge of the plot ($c = 2.25$) is where $Q = -8.6$. For a fixed decrement, the closer together the contours, the steeper the hill; and the smaller the length of a closed contour, the nearer the summit. So the hill is quite gentle near the summit, which is represented by the point $(\hat{c}, \hat{s}) \approx (1.9, 0.79)$. The precise MLEs (correct to 4 s.f.) are $\hat{c} \approx 1.899$ and $\hat{s} \approx 0.7871$ (with $Q(\hat{c}, \hat{s}) = -7.974$).[2]

This is not, however, the only way to find the Weibull MLEs, and it will be instructive to discuss another. Suppose that c is temporarily fixed. Then (27) yields

$$\begin{aligned} \frac{\partial Q}{\partial s} &= \frac{\partial}{\partial s}\left\{N\ln(c) + \sum_{k=1}^{N}\ln(x_k^{c-1}) - cN\ln(s) - s^{-c}\sum_{k=1}^{N} x_k^c\right\} \\ &= -cN\frac{\partial}{\partial s}\{\ln(s)\} - \sum_{k=1}^{N} x_k^c\frac{\partial}{\partial s}\{s^{-c}\} \\ &= -cN\frac{1}{s} - \sum_{k=1}^{N} x_k^c\{-cs^{-c-1}\} \\ &= \frac{Nc}{s^{c+1}}\left\{\frac{1}{N}\sum_{k=1}^{N} x_k^c - s^c\right\} \end{aligned} \tag{31.28}$$

because the first two terms of Q are independent of s. If we define an ordinary function u by

$$u(c) = \left(\frac{1}{N}\sum_{k=1}^{N} x_k^c\right)^{\frac{1}{c}}, \tag{31.29}$$

then $\partial Q/\partial s = 0$ for $s = u(c)$. Moreover, $s < u(c)$ implies $\partial Q/\partial s > 0$ and $s > u(c)$ implies $\partial Q/\partial s < 0$. Thus, for any given c, Q has a maximum where $s = u(c)$, and the value of

---

[2] The exact values can be found, e.g., by the Mathematica command **FindMinimum[–Q[c,s],{c,1.9},{s, 0.8}]**.

the maximum is $Q(c, s) = Q(c, u(c))$, which defines an ordinary function of c. That is, on substituting (29) back into (27), we have

$$\begin{aligned} Q(c,u(c)) &= N\ln(c) + \sum_{k=1}^{N}\ln(x_k^{c-1}) - cN\ln(u(c)) - \frac{1}{\{u(c)\}^c}\sum_{k=1}^{N}x_k^c \\ &= N\ln(c) + \sum_{k=1}^{N}\ln(x_k^{c-1}) - N\ln\left(\frac{1}{N}\sum_{k=1}^{N}x_k^c\right) - N. \end{aligned} \tag{31.30}$$

Now we can use tried and true methods for ordinary functions, e.g., drawing an accurate graph, to find the MLE $\hat{c}$ as the value of c that maximizes $Q(c, u(c))$; and the MLE $\hat{s}$ follows directly from (29). For example, $z = Q(c, u(c))$ is graphed in Figure 5(a) for the rat pupil data from Figure 2. The global maximizer is $\hat{c} = 1.899$, and from (29) and (21) we have $\hat{s} = u(\hat{c}) = 0.63458^{1/\hat{c}} = 0.787$. The corresponding distribution, defined by (25) with $c = \hat{c}$ and $s = \hat{s}$ or

$$f(x) = 2.993x^{0.8994}e^{-1.576x^{1.899}}, \tag{31.31}$$

is shown dashed in Figure 6.

The same approach yields the MLEs of c and s for a Gamma distribution, with p.d.f. defined by

$$f(x) = \frac{x^{c-1}e^{-x/s}}{s^c\Gamma(c)} \tag{31.32}$$

and likelihood

$$\begin{aligned} L(c,s) &= \prod_{k=1}^{N}f(x_k) = \prod_{k=1}^{N}\frac{x_k^{c-1}}{s^c\Gamma(c)}e^{-x_k/s} \\ &= \frac{x_1^{c-1}}{s^c\Gamma(c)}e^{-x_1/s}\cdot\frac{x_2^{c-1}}{s^c\Gamma(c)}e^{-x_2/s}\cdot\frac{x_3^{c-1}}{s^c\Gamma(c)}e^{-x_3/s}\cdots\frac{x_N^{c-1}}{s^c\Gamma(c)}e^{-x_N/s} \\ &= \left(\frac{1}{s^c\Gamma(c)}\right)^N x_1^{c-1}x_2^{c-1}x_3^{c-1}\cdots x_N^{c-1}e^{-x_1/s}\cdot e^{-x_2/s}\cdot e^{-x_3/s}\cdots e^{-x_N/s} \\ &= s^{-cN}\{\Gamma(c)\}^{-N}\left(x_1x_2x_3\cdots x_N\right)^{c-1}e^{-\{x_1+x_2+x_3+\cdots+x_N\}/s} \\ &= \left(x_1x_2x_3\cdots x_N\right)^{c-1}\{\Gamma(c)\}^{-N}s^{-cN}e^{-N\bar{x}/s}, \end{aligned} \tag{31.33}$$

by (14). So if $q = \ln(L)$ denotes the loglikelihood, then

$$\begin{aligned} q(c,s) &= \ln(L(c,s)) = \ln\left(\{x_1x_2x_3\cdots x_N\}^{c-1}\right) + \ln\left(\{\Gamma(c)\}^{-N}\right) + \ln\left(s^{-cN}\right) + \ln\left(e^{-N\bar{x}/s}\right) \\ &= (c-1)\ln(x_1x_2x_3\cdots x_N) - N\ln(\Gamma(c)) - cN\ln(s) - N\frac{\bar{x}}{s} \\ &= (c-1)\sum_{k=1}^{N}\ln(x_k) - N\ln(\Gamma(c)) - cN\ln(s) - N\frac{\bar{x}}{s}. \end{aligned} \tag{31.34}$$

Again, suppose that c is temporarily fixed. Then (33) yields

$$\begin{aligned}\frac{\partial q}{\partial s} &= \frac{\partial}{\partial s}\left\{(c-1)\sum_{k=1}^{N}\ln(x_k) - N\ln(\Gamma(c)) - cN\ln(s) - N\frac{\bar{x}}{s}\right\}\\ &= -cN\frac{\partial}{\partial s}\{\ln(s)\} - N\bar{x}\frac{\partial}{\partial s}\{s^{-1}\}\\ &= -cN\frac{1}{s} - N\bar{x}\{-s^{-2}\}\\ &= \frac{N}{s^2}\{\bar{x}-cs\}\end{aligned} \tag{31.35}$$

because the first two terms of q are independent of s. If we define

$$w(c) = \frac{\bar{x}}{c} = \frac{1}{cN}\sum_{k=1}^{N}x_k, \tag{31.36}$$

then $\partial q/\partial s = 0$ for $s = w(c)$. Moreover, $s < w(c)$ implies $\partial q/\partial s > 0$ and $s > w(c)$ implies $\partial q/\partial s < 0$. Thus, for any given c, q has a maximum where $s = w(c)$, and the value of the maximum is $q(c, s) = q(c, w(c))$, which defines an ordinary function of c. That is, on substituting (35) back into (33), we have

$$\begin{aligned}q(c,w(c)) &= (c-1)\sum_{k=1}^{N}\ln(x_k) - N\ln(\Gamma(c)) - cN\ln(w(c)) - N\frac{\bar{x}}{w(c)}\\ &= (c-1)\sum_{k=1}^{N}\ln(x_k) - N\ln(\Gamma(c)) - cN\ln(\bar{x}/c) - Nc.\end{aligned} \tag{31.37}$$

The MLE $\hat{c}$ is now found as the value of c that maximizes $q(c, w(c))$, and $\hat{s} = w(\hat{c})$ by (35). For the rat pupil data, (33) and Table 2 imply that

$$q(c,w(c)) = 22\{0.510 - 1.510c - \ln(\Gamma(c)) - c\ln(0.693/c)\} \tag{31.38}$$

(Exercise 3). The corresponding graph, $z = q(c, w(c))$, is shown in Figure 5(b), where it appears that $\hat{c} \approx 3.6$. The precise MLEs (correct to 4 s.f.) are $\hat{c} = 3.639$ and $\hat{s} = w(3.639) = 0.1905$ (with $Q(\hat{c}, \hat{s}) = -6.789$). The corresponding distribution, defined by (31) with $c = \hat{c}$ and $s = \hat{s}$ or

$$f(x) = 107.3x^{2.639}e^{-5.249x} \tag{31.39}$$

is shown dotted in Figure 6.

| ln(RADIUS) | FREQUENCY | ln(RADIUS) | FREQUENCY |
|---|---|---|---|
| –1.3863 | 2 | 0 | 3 |
| –0.9808 | 4 | 0.2331 | 1 |
| –0.6931 | 7 | 0.4055 | 1 |
| –0.2877 | 3 | 0.5596 | 1 |

Table 31.3 Logarithms of pupil radii from an experiment with laboratory rats

Comparing the c.d.f. of each fitted distribution with that of the sample, we find on the whole that the dotted curve in Figure 6(b) is closer than the dashed curve to the data points. But an even better model of rat pupil variation appears to be the solid curve, which corresponds to a lognormal distribution. Recall from Lecture 28 that X has a lognormal distribution if $U = \ln(X)$ has a normal distribution. To apply the method of maximum likelihood, therefore, we must transform the data in Table 2 by taking logarithms. The results are shown in Table 3. Now, from Appendix 31, if U has a normal distribution with mean μ and standard deviation σ, then the MLEs of the two parameters from a sample $\{u_1, u_2, ..., u_N]$ of size N with sample mean

$$\overline{u} = \frac{1}{N}\sum_{k=1}^{N} u_k \qquad (31.40)$$

and **sample variance**

$$S^2 = \frac{1}{N}\sum_{k=1}^{N} (u_k - \overline{u})^2 \qquad (31.41)$$

are $\hat{\mu} = \overline{u}$ and $\hat{\sigma} = S$, respectively. With U = ln(X), we have

$$\overline{u} = \frac{1}{22}\sum_{k=1}^{22} u_k = \frac{1}{22}\sum_{k=1}^{22} \ln(x_k) \qquad (31.42)$$

$= \{2\times(-1.3863) + 4\times(-0.9808) + 7\times(-0.6931) + 3\times(-0.2877) + 3\times 0 + 0.2331 + 0.4055 + 0.5596\}/22 = -0.510$ from Table 3 and (40), and a similar calculation yields

$$S = \sqrt{\frac{1}{22}\sum_{k=1}^{22} (\ln(x_k) + 0.510)^2} = 0.527. \qquad (31.43)$$

(Exercise 4). Thus, from (19.59), the maximum likelihood, or ML, distribution for U is given by

$$g(u) = \frac{1}{S\sqrt{2\pi}} e^{-(u-\overline{u})^2/2S^2} = 0.757\, e^{-1.80(u+0.510)^2}, \qquad (31.44)$$

and from Exercise 5 the p.d.f. for X itself is given by

$$\begin{aligned} f(x) &= g(\ln(x))\frac{d}{dx}\{\ln(x)\} \\ &= \frac{1}{Sx\sqrt{2\pi}} e^{-(\ln(x)-\overline{u})^2/2S^2} = \frac{0.757}{x} e^{-1.80(\ln(x)+0.510)^2}. \end{aligned} \qquad (31.45)$$

This equation defines the solid curve in Figure 6(a). The solid curve in Figure 6(b) then follows from F(x) = Int(f, [0, x]), for which Exercise 6 yields an explicit expression.

Figures 7 and 8 are diagrams analogous to Figure 6 for the leaf thickness data and minnow size data in Tables 19.2 and 19.4, respectively; see Exercises 6-7. In Figure 8, the Weibull distribution appears to be the best overall model, and the lognormal appears to be the worst, so that their roles from Figure 6 are reversed. In Figure 7, by contrast, nothing suggests strongly that any of the three distributions yields a better model than the other two. Note, moreover, that even if one is clearly best, it remains possible that all are inadequate, in the sense of being too unlikely to have yielded the observed sample (or that all are adequate, in the sense of being sufficiently likely to have yielded the sample). Such matters of "goodness of fit" belong more properly in a course on statistics, however, and so we do not discuss them further here.

**References**

Davies, N. B. & T. R. Halliday (1977). Optimal mate selection in the toad *Bufo bufo.* Nature **269**, 56-58.

Mesterton-Gibbons, M. & L. A. Dugatkin (1995). Toward a theory of dominance hierarchies: Effects of assessment, group size and variation in fighting ability. Behavioral Ecology **6**, 416-423

**Exercises 31**

31.1 Verify (4)-(6). By substituting from (6) into (1), show that the MLE of p for *Bufo bufo* is $\hat{p} = 13/41$, as intuition would suggest.

31.2 Verify (13)-(17) and (19)-(23).

31.3 Verify (28)-(30) and (33)-(38) .

31.4 Verify Table 3 and (31.43).

31.5 Verify (45). Hint: Use the result you obtained in Exercise 21.1

31.6 Write a Mathematica program to find the maximum likelihood Weibull, Gamma and lognormal distributions for the sample of *D. linearifolia* leaf thicknesses in Table 19.2.

31.7 Write a Mathematica program to find the maximum likelihood Weibull, Gamma and lognormal distributions for the sample of minnow sizes (above base length) in Table 19.4.

31.8 Show that the solid curve in Figure 6(b) has equation

$$y \quad = \quad \frac{1}{2}\left\{1 + \text{erf}\left(\frac{\ln(x) - \bar{u}}{S\sqrt{2}}\right)\right\},$$

where $\bar{u}$, S and erf are defined by (42)-(43) and Appendix 28B. Hint: Don't apply F(x) = Int(f, [0, x]). The result can be deduced without integration from Appendix 28B.

**Appendix 31: Maximum likelihood mean and variance for a normal distribution**

In this appendix we calculate the MLEs from a sample of size N for the mean and variance of a normal distribution defined by

$$f(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty. \tag{31.A1}$$

On using (10), we have

$$\begin{aligned} Q(\mu,\sigma) &= \ln(L) \;=\; \ln\left(\prod_{k=1}^{N} f(x_k)\right) \;=\; \sum_{k=1}^{N} \ln\{f(x_k)\} \\ &= \sum_{k=1}^{N} \ln\left\{\frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x_k-\mu)^2/2\sigma^2}\right\} \\ &= \sum_{k=1}^{N} \left\{\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \ln\left(e^{-(x_k-\mu)^2/2\sigma^2}\right)\right\} \\ &= \sum_{k=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \sum_{k=1}^{N}\left\{-\frac{(x_k-\mu)^2}{2\sigma^2}\right\} \\ &= N\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2. \end{aligned} \tag{31.A2}$$

So

$$\begin{aligned} \frac{\partial Q}{\partial \mu} &= \frac{\partial}{\partial \mu}\left\{N\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2\right\} \\ &= \frac{\partial}{\partial \mu}\left\{N\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\right\} - \frac{1}{2\sigma^2}\frac{\partial}{\partial \mu}\left\{\sum_{k=1}^{N}(\mu-x_k)^2\right\} \\ &= 0 - \frac{1}{2\sigma^2}\sum_{k=1}^{N}2(\mu-x_k) \;=\; \frac{1}{\sigma^2}\sum_{k=1}^{N}(x_k-\mu) \\ &= \frac{1}{\sigma^2}\left\{\sum_{k=1}^{N}x_k - \sum_{k=1}^{N}\mu\right\} \;=\; \frac{1}{\sigma^2}\left\{\sum_{k=1}^{N}x_k - N\mu\right\} \\ &= \frac{N}{\sigma^2}\{\bar{x}-\mu\}, \end{aligned} \tag{31.A3}$$

where the sample mean $\bar{x}$ is defined by (14). So $\partial Q/\partial\mu = 0$ if $\mu = \bar{x}$, with $\partial Q/\partial\mu > 0$ if $\mu < \bar{x}$ and $\partial Q/\partial\mu < 0$ if $\mu > \bar{x}$: For any given $\sigma$, Q has a maximum where $\mu = \bar{x}$. If we defined the by

$$s^2 \;=\; \frac{1}{N}\sum_{k=1}^{N}(x_k-\bar{x})^2, \tag{31.A4}$$

then value of the maximum is

$$Q(\bar{x},\sigma) \;=\; N\left\{\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{s^2}{2\sigma^2}\right\}, \tag{31.A5}$$

which is an ordinary function of $\sigma$. We have

$$\begin{aligned}\frac{\partial Q}{\partial \sigma} &= N\frac{\partial}{\partial \sigma}\left\{\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{s^2}{2\sigma^2}\right\}\\ &= N\left\{\frac{\partial}{\partial \sigma}\left\{\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\right\} - \frac{s^2}{2}\frac{\partial}{\partial \sigma}\left\{\sigma^{-2}\right\}\right\},\\ &= N\left\{-\frac{1}{\sigma} - \frac{s^2}{2}\left\{-2\sigma^{-3}\right\}\right\}\\ &= \frac{N}{\sigma^3}\left\{s^2 - \sigma^2\right\}.\end{aligned} \tag{31.A6}$$

So $\partial Q/\partial\sigma = 0$ if $\sigma = s$, with $\partial Q/\partial\sigma > 0$ if $\sigma < s$ and $\partial Q/\partial\sigma < 0$ if $\sigma > s$: For any given $\mu$, Q has a maximum where $\sigma = s$. Thus $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = s$. In other words, the MLEs are

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} x_k \tag{31.A7}$$

and

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{x})^2}. \tag{31.A8}$$

**Answers and Hints for Selected Exercises**

31.6 For the Weibull distribution, $\hat{c} = 5.79457$ and $\hat{s} = 0.15344$ imply

$$f(x) = 302103.0x^{4.79457}e^{-52135.6x^{5.79457}}.$$

For the Gamma distribution, $\hat{c} = 28.0703$ and $\hat{s} = 0.00509118$ imply

$$f(x) = 1.70632\times 10^{36}x^{27.0703}e^{-196.418x}.$$

For the lognormal distribution, $\bar{u} = -1.96345$ and $S = 0.195476$ imply

$$f(x) = \frac{2.04088}{x}e^{-13.0853(\ln(x)+1.96345)^2}.$$