8

Characterization of Uncertainty

8.1 Introduction

Fitting a model to a set of data involves searching for parameter estimates that optimize the relationship between the observed data and the predictions of the model. The parameter estimates are taken to represent some aspect of the world in which we are interested. While it should be possible to find optimum parameter values in each situation, it remains the case that the data used are only a sample. A different sample of the same kind of data would very likely lead to at least slightly different parameter estimates; so the exact value of the estimated parameters is not really the issue. Rather, we want to know how repeatable such estimates are likely to be if we had the luxury of having multiple samples. That is, the parameters are just estimates, and we want to know how confident we can be in those estimates. Some aspects of uncertainty, such as highly variable data leading to parameter estimates with wide confidence intervals, are relatively familiar, and more details will be given here. However, in modelling there are many potential sources of uncertainty, and we need to be aware of them all to avoid drawing stronger conclusions than our data and models should allow.

In this chapter we will explore alternative ways available for characterizing the uncertainty inherent in any modelling situation. Bootstrapping is an excellent method for examining the levels of uncertainty stemming from the available data, but that has already been considered in Chapter 6 and we will not repeat that material. While we will introduce many sources of potential uncertainty, only some of these can be usefully investigated here. Some sources of uncertainty influence the variability of data collected, other sources can influence the type of data available. There are various ways of describing the different types of uncertainty that can influence fisheries models. These are often called sources of error, usually in the sense of residual error rather than as a mistake having been made. Unfortunately, the term *error* has the potential to lead to confusion, so it is best to use the term *uncertainty*.

Francis and Shotton (1997) listed six different types of uncertainty, while Kell et al. (2005) contract these to five; however, they can all be summarized under four headings, some with subheadings, as follows:

- Process uncertainty: Underlying natural random variation in demographic rates (such as growth, interannual recruitment, interannual natural survivorship) and other biological properties and processes. This would be a significant source of residual error between observed and predicted values, so best to call this process uncertainty rather than process error.
- Observation uncertainty: Sampling error and measurement error—samples are meant to represent a population but remain only a sample. Inadequate or nonrepresentative data collection would contribute to observational uncertainty, as would any mistakes or the deliberate misreporting of data (not unknown in the world of fisheries). This is another source of residual errors between observations and predicted values.
- Model uncertainty: Relates to the capacity of the selected model structure to describe the workings of the system dynamics. There can be more than one type of model uncertainty:
  - Different structural models may provide different answers, and uncertainty exists over which is the better representation of nature.
  - The selection of the residual error structure for a given process is a special case of model uncertainty that has already been shown (in Chapter 3) to have important implications for parameter estimates.
  - Estimation uncertainty can be thought of as a form of model uncertainty in that the model structure is not capable of capturing the process dynamics in an unbiased manner.
- Implementation uncertainty: Where the effects or extent of management actions may differ from those intended.
  - Poorly defined management options may lead to implementation uncertainty.
  - Institutional uncertainty: Inadequately defined management objectives leading to unworkable management.
- Time lags between making decisions and implementing them can lead to greater variation. Assessments are often made a year or more after fisheries data are collected, and management decisions often take another year or more to be implemented.

Model uncertainty can be both quantitative and qualitative. Thus, hierarchically structured models could be compared to one another, and the best fitting selected between such models. Such models may be considered to be related but quantitatively different. However, where models are incommensurate, as we have seen, for example, when different residual error structures are used with the same structural model, they can each generate an

optimum fit, and model selection must be based on factors other than quality of fit. Such models do not grade smoothly into each other but constitute qualitatively different descriptions of the system under study. Model uncertainty is one of the driving forces behind model selection. Even where there is only one model developed this has been implicitly selected from many possible models. Working with more than one type of model in a given situation (perhaps a surplus production model along with a fully age-structured model; see later chapters) can often lead to insights that using one model alone would be missed.

Institutional uncertainty is very important in the management of natural resources. The best modelling practices in the world would be pointless if significant implementation uncertainty is present. When discussing an input-controlled fishery, Dichmont et al. (2006) suggested four sources of implementation error: (1) the extent to which management advice is accepted by the decision makers, (2) management decisions fail to be implemented because of different levels of participation in the fishery, (3) restricting fishing mortality may be difficult when a fishery targets multiple species or stocks, and (4) inadequate allowance made for changes in fishing power through time. Time lags between receiving data, generating an assessment, and implementing its recommendations are extremely common in fisheries practice. The extent to which advice is accepted or is effective can be influenced by time delays between advice being given and it being implemented.

Model and implementation uncertainty are both very important in the management of natural resources. However, in this chapter we are going to concentrate on methods that allow us to characterize the confidence with which we can accept the various parameter estimates and other model outputs obtained from given models that have significance for management advice. There are numerous methods for characterizing the uncertainty around any parameter estimates or model outputs. Some methods work best when the underlying data structures are normally distributed, whereas others are more general. We will start with the more classical approaches that relate to estimates of normal standard errors. These lead naturally toward methods that characterize uncertainty by defining the sampling distribution of parameters and outputs and identifying some central percentile range around the distribution's expectation.

8.2 Asymptotic Standard Errors

The classical definition of the 95% confidence intervals around the mean of a sample is given by (Snedecor and Cochran, 1967)

$$\bar{x} \pm t_v \frac{\sigma}{\sqrt{n}} \tag{8.1}$$

where $\bar{x}$ is the mean of the sample, n is the sample size, σ is the sample standard deviation, and t^v is the t distribution with $v = (n - 1)$ degrees of freedom. The term σ represents the variability of the sample about its mean. If we had multiple samples it would be possible to estimate the standard deviation of the group of sample means. The fragment $\sigma/\sqrt{n}$ is known as the standard error of the variable x, and is one way of estimating the expected standard deviation of a set of sample means when there is only one sample. Classical confidence intervals are distributed symmetrically around the mean or expectation, and this is standard when dealing with normal distributions.

While the classical definition works well with single parameters and normally distributed data, there can be complications when fitting a model if there are multiple and possibly correlated parameters being estimated at once. When multiple parameters are fitted to a model, there are a number of ways in which it is possible to generate estimates of standard errors around each parameter value.

Asymptotic standard errors are produced by estimating the variance-covariance matrix for the model parameters in the vicinity of the optimum parameter set. Essentially the gradient of the maximum likelihood or sum of squares surface at the optimum is used to characterize any relationships between the various parameters, and these are used to generate the variance-covariance matrix. Parts of this are then used in conjunction with the residual variance from the optimum fit to produce the estimates of standard error so that Equation 8.1 can be used to generate approximate confidence intervals. The primary assumptions are that the fitted surface near the optimum is regular and smooth, and that the surface is approximately normal near the optimum. This means the standard errors will be symmetric around the optimal solution, which may or may not be appropriate. As a first approximation, however, asymptotic standard errors provide an indication of the inherent variation around the parameter estimates. The methodology will be illustrated with the data and linear regression describing the relationship between catches of king prawns and tiger prawns used in Chapter 6 (Example Box 6.3).

Formally the variance-covariance matrix is the product of the residual variance and the inverse of the Hessian matrix. The residual variance is simply

$$Sy^2 = \frac{\sum \left(y - \bar{y}\right)^2}{n - p} \tag{8.2}$$

where the y values are the observed variable (in this case the king prawn catches), $\bar{y}$ is the expected king prawn catch for each time period (predicted

by the regression on tiger prawns), n is the number of sets of observations (in this case twelve data pairs), and p is the number of parameters being estimated (in this case two, a gradient and intercept). The Hessian matrix describes the local curvature or gradient of the sum of squared residual surface. More formally, this is made up of the second-order partial derivatives of the function describing the sum of squared residuals. Given a function with its parameters, e.g., a, b and the linear regression

$$King = f\left(Tig\right) = a.Tig + b \tag{8.3}$$

the square Hessian matrix of this function f, with two parameters a and b, would be

$$\mathbf{H}\left(f\right) = \begin{Bmatrix} \dfrac{\partial^2 f}{\partial a^2} & \dfrac{\partial^2 f}{\partial a \partial b} \\ \dfrac{\partial^2 f}{\partial a \partial b} & \dfrac{\partial^2 f}{\partial b^2} \end{Bmatrix} \tag{8.4}$$

The variance-covariance matrix (**A**) is estimated by multiplying the elements of the inverse of the Hessian by the residual variance, Equation 8.2:

$$\mathbf{A} = Sy^2.\mathbf{H}^{-1} \tag{8.5}$$

The estimate of the standard error for each parameter in the set θ is then obtained by taking the square root of the diagonal elements of the variance-covariance matrix:

$$StErr(\theta) = \sqrt{diag\left(\mathbf{A}\right)} \tag{8.6}$$

We can approximate the elements of the Hessian numerically by making small increments to the optimum parameter estimates and estimating the gradients of the surface at the optimum (this is the method of finite differences). Thus, for each of the n data points, if we use a small increment (e.g., $\Delta = 0.00001$), we would calculate the gradient of the surface of the predicted king catches with respect to the two parameters individually as

$$\frac{\partial K}{\partial a} \approx \left[\left(\left(a+\Delta\right)Tig + b\right) - \left(\left(a-\Delta\right)Tig + b\right)\right] \Big/ \left(2\,\Delta\right) \tag{8.7}$$

and

$$\frac{\partial K}{\partial b} \approx \left[\left(aTig + \left(b+\Delta\right)\right) - \left(aTig + \left(b-\Delta\right)\right)\right] \Big/ \left(2\,\Delta\right) \tag{8.8}$$

The sums of the squared values and of the cross products of these values are then calculated:

$$\frac{\partial^2 K}{\partial a^2} = \sum^{n} \left(\frac{\partial K}{\partial a} \right)^2 \tag{8.9}$$

$$\frac{\partial^2 K}{\partial b^2} = \sum^{n} \left(\frac{\partial K}{\partial b} \right)^2 \tag{8.10}$$

$$\frac{\partial^2 K}{\partial a \partial b} = \sum^{n} \left(\frac{\partial K}{\partial a} \cdot \frac{\partial K}{\partial b} \right) \tag{8.11}$$

These are substituted into Equation 8.4 and the result inverted (the array function minverse in Excel) and used in Equations 8.5 and 8.6, and then finally in Equation 8.1, to provide confidence intervals around each parameter. These equations are all implemented in Example Box 8.1.

Remember that asymptotic standard errors are quick and convenient but can usually only be regarded as approximate. Nevertheless, they can provide an indication of which parameters are least well estimated and may suggest whether a more detailed examination of uncertainty would be wise.

8.3 Percentile Confidence Intervals Using Likelihoods

There are a number of ways of generating confidence intervals around parameter estimates when the optimum fit is found using a maximum likelihood criterion. It is usually possible to calculate asymptotic standard errors, but there are alternative approaches that can often produce more convincing confidence intervals. Using maximum likelihood, a von Bertalanffy growth curve can be fitted to female Pacific hake (Kimura, 1980; Figure 8.2, Example Box 8.2). A plot of different hypothesized values for the parameter of interest against their respective likelihood (after optimizing the fit of any remaining parameters, e.g., Figure 8.3) demonstrates that as one increments away from the optimum parameter value, the likelihoods decline in a smooth fashion. Such plots of likelihood against the parameter of interest are known as likelihood profiles.

Approximate confidence intervals around parameters may be obtained by generating a likelihood profile, standardizing those likelihoods so they add to 1 (i.e., sum all the separate likelihoods and divide each by the total), and then finding the particular parameter values that encompass the confidence

EXAMPLE BOX 8.1

A linear regression between the bycatch of king prawns and the catch of tiger prawns from the Australian Northern Prawn Fishery in the years 1976 to 1987; data as in Example Box 6.3 (see Figure 8.1). Minimize the sum of squared residual errors in B5 using the solver and varying the two parameters in B2:B3. By plotting column B against column A as data points and adding column C against column A as a line, you should be able to reproduce Figure 8.1.

	A	B	C	D
2	**a**	0.025854		
3	**b**	–11.4826		
4				
5	**SSQ**	=SUM(D11:D22)		
6	**n**	=COUNT(A11:A22)–2		
7				
8	**Sy^2**	=B5/B6		StErr
9				
10	**Tiger**	**King**	**E(King)**	**$Resid^2$**
11	566	10	=$B$2*A11+$B$3	=(B11–C11)^2
12	1437	22	=$B$2*A12+$B$3	=(B12–C12)^2
13	1646	42	=$B$2*A13+$B$3	=(B13–C13)^2
14	2056	33	41.67	75.22
15	3171	64	70.50	42.26
16	Get data	From Example Box 6.3	Copy down to	Row 22

Generating the asymptotic standard errors requires the parameters to be incremented slightly to estimate a gradient, so place =((($B$2+$F$9)*A11+$B$3)–(($B$2–$F$9)*A11+$B$3))/(2*$F$9) into E11, which increments the **a** parameter, and copy down to row 22; put =(($B$2*A11+($B$3+$F$9))–($B$2*A11+($B$3–$F$9)))/(2*$F$9) into F11 to increment the **b** parameter and copy down to row 22. Vary the value of delta in F9 to explore the sensitivity of the process to this value. After filling in cells G11:I22, the estimate of the Hessian matrix **H** is formed in E2:F3. Select cells E5:F6, type =MINVERSE(E2:F3), then press all three keys <Ctrl><Shift><Enter> at once to produce the necessary array function needed to invert the Hessian. Cells H2:I3 constitute the variance-covariance matrix and represent Equation 8.5. The matrix in H5:I6 contains the correlation matrix between the parameters. Put =H2/SQRT(H2*H2) into H5 and =I3/SQRT(I3*I3) into I6, both of which

continued

EXAMPLE BOX 8.1 (continued)

should produce 1.0. The cross-correlations are gained from placing =I2/SQRT(H2*I3) into I5 and =H3/SQRT(H2*I3) into H6, which again should be equal. The standard error estimates are produced in E8:F8; the confidence intervals are generated by placing =B2–TINV(0.05,B6)*E8 into H7 and similarly putting =B2+TINV(0.05,B6)*E8 into H9, with =B3–TINV(0.05,B6)*F8 in I7 and =B3+TINV(0.05,B6)*F8 in I9 for the approximate 95% confidence intervals.

	E	F	G	H	I
1	**H**			**A**	
2	=SUM(G11:G22)	=SUM(H11:H22)		=$B$8*E5	=$B$8*F5
3	=F2	=SUM(I11:I22)		=$B$8*E6	=$B$8*F6
4	$\mathbf{H^{-1}}$			**a**	**b**
5	7.06E–08	–0.00019		1.0000	–0.9271
6	–0.00019	0.593226		–0.9271	1.0000
7	$\mathbf{(Sy^2H^{-1})^{1/2}}$		Lower95	0.0131	–48.3843
8	=SQRT(H2)	=SQRT(I3)	Estimate	=B2	=B3
9	Delta	0.000001	Upper95	0.0386	25.4190
10	**δK/δa**	**δK/δb**	$\mathbf{X^2}$	**XY**	$\mathbf{Y^2}$
11	566.00	1.000	=E11*E11	=E11*F11	=F11*F11
12	1437.00	1.000	=E12*E12	=E12*F12	=F12*F12
13	1646.00	1.000	=E13*E13	=E13*F13	=F13*F13
14	2056.00	1.000	=E14*E14	=E14*F14	=F14*F14
15	3171.00	1.000	=E15*E15	=E15*F15	=F15*F15
16	2743.00	1.000	=E16*E16	=E16*F16	=F16*F16
17	2838.00	1.000	=E17*E17	=E17*F17	=F17*F17
18	Copy	Down to	Row	22	

intervals required (usually the central 95 or 90% of the total distribution). This approach would only be valid if the increments on the parameter axis are relatively small, so that the likelihood distribution approximates a continuous probability distribution. In addition, the likelihoods at the extremes of the parameter values considered should be negligible. This is an empirical likelihood profile method and is illustrated in Example Box 8.2.

The standardized likelihoods should cumulate to 1, and the 95% confidence intervals may be determined by searching for the 0.025 and 0.975 percentiles of the cumulative distribution (Example Box 8.2). An important difference between likelihood profile confidence intervals and those generated using 1.96 times the asymptotic standard errors is that the latter are forced by their nature to be symmetrical. This is not necessarily the case when using likelihood profile confidence intervals (Figure 8.3). A similar

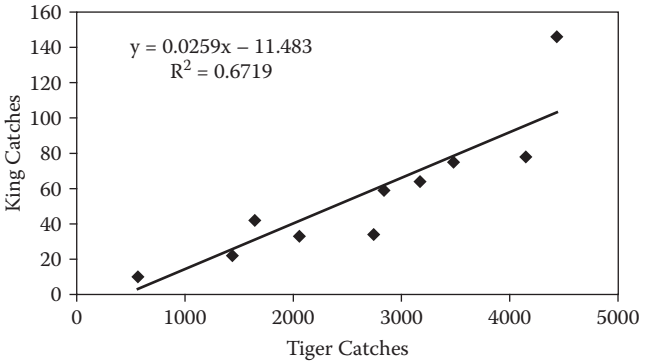

Figure 8.1
The linear regression between king prawn catches and tiger prawn catches, in tonnes, from the Australian Northern Prawn Fishery for the years 1976 to 1987. See Example Boxes 6.3 and 8.1.

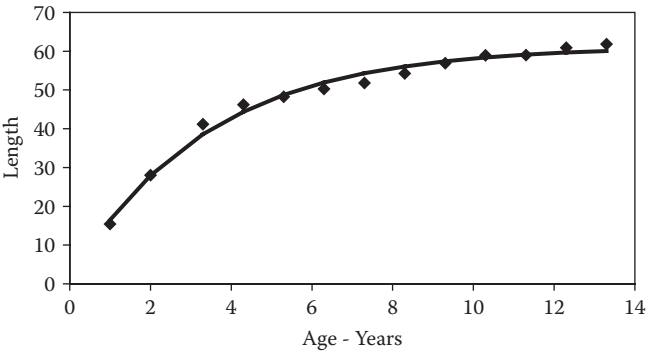

Figure 8.2
The optimal von Bertalanffy curve fitted to data for female Pacific hake (Kimura, 1980; equal weight given to each data point). Optimum parameters: L_∞ =61.22, K = 0.296, t_0 = –0.057, and σ = 1.489 (Example Box 8.2).

process could be applied to each of the model parameters in turn, or even to multiple parameters at once.

8.4 Likelihood Profile Confidence Intervals

An alternative to empirical likelihood profiles is to use a deterministic approximation. Venzon and Moolgavkar (1988) describe a method of obtaining what they call "approximate 1 – α profile-likelihood-based confidence intervals." This simple procedure starts with finding the set of model parameters that generate the maximum likelihood for the data. The method then

EXAMPLE BOX 8.2

Fitting a von Bertalanffy growth curve to length-at-age data for female Pacific hake (Kimura, 1980) using maximum log-likelihood. To calculate the expected lengths in column C, put =$B$2*(1–exp(–$B$3*(A8–$B$4))) into C8 and copy down to C20. Put =Ln(normdist(B8,C8,$B$5,false)) in D8 and copy down to D20. Plot column B against A as a scattergram and add column C as a line (cf. Figure 8.2) to observe the relation between age and length. First use the solver to maximize B6 by altering B2:B5. To generate a likelihood profile fill E8:E73 with 57 to 70 in steps of 0.2. To produce the profile, individually copy each of the possible Linf values from column E into B2 and re-solve for the maximum likelihood, but only by varying B3:B5. In other words, for each value of Linf, optimize for the other parameters and save the log-likelihood value in column F next to its respective Linf value. Do this for all the values in column E (a macro to do this is given in the example box continuation, where the final details of producing the likelihood profile are given; cf. Figure 8.3).

	A	B	C	D	E	F	G
1		**Parameters**					
2	**L∞**	61.2250					
3	**K**	0.2963					
4	**t0**	–0.0573					
5	**Sig**	1.4896					
6	**LL**	=sum(D8:D20)					=sum(G8:G73)
7	**Age**	**Obs(L)**	**Ex(L)**	**Log_L**	**Linf**	**Profile**	**Likelihood**
8	1	15.40	16.467	–1.5741	57.0		=exp(F8)
9	2	28.03	27.945	–1.3191	57.2		=exp(F9)
10	3.3	41.18	38.584	–2.8359	57.4		=exp(F10)
11	4.3	46.20	44.390	–2.0555	57.6		=exp(F11)
12	5.3	48.23	48.707	–1.3688	57.8		=exp(F12)
13	6.3	50.26	51.917	–1.9364	58.0		=exp(F13)
14	7.3	51.82	54.304	–2.7081	58.2		=exp(F14)
15	8.3	54.27	56.079	–2.0549	58.4		=exp(F15)
16	9.3	56.93	57.399	–1.3669	58.6		=exp(F16)
17	10.3	58.93	58.380	–1.3856	58.8		=exp(F17)
18	11.3	59.00	59.110	–1.3201	59.0		=exp(F18)
19	12.3	60.91	59.652	–1.6740	…		=exp(F19)
20	13.3	61.83	60.055	–2.0271	to 70		=exp(F20)

continued

EXAMPLE BOX 8.2 (continued)

Development of likelihood profile for the Linf parameter for the von Bertalanffy curve fitted in Example Box 8.1. Copy each of the possible Linf values from column E into cell B2 and re-solve, maximizing the log-likelihood in B6 by varying B3:B5 (do not include B2), copying the optimum maximum log-likelihood from B6 down into column F. Either do this manually or implement the following macro (much of the macro can be recorded and then edited), and then plot column H against column E (cf. Figure 8.3). Compare this with a plot of the log-likelihood in column F.

```
Sub Do_Profile()
For i = 1 To 66
  Range("E7").Select
  ActiveCell.Offset(i, 0).Range("A1").Select
  Selection.Copy
  Range("B2").Select
  Selection.PasteSpecial Paste:=xlPasteValues
  SolverOk SetCell:="$B$6", MaxMinVal:=1, ValueOf:="0",
    byChange:="$B$3:$B$5"
  SolverSolve (True)
  Range("B6").Select
  Selection.Copy
  Range("F7").Select
  ActiveCell.Offset(i, 0).Range("A1").Select
  Selection.PasteSpecial Paste:=xlPasteValues
Next i
End Sub
```

	E	F	G	H	I
6			=SUM(G8:G73)		
7	**Linf**	**Profile**	**Likelihood**	**Std Like**	**Cumulative**
8	57	–29.6787	=EXP(F8)	=G8/$G$6	=H8
9	57.2	–29.2452	=EXP(F9)	=G9/$G$6	=I8+H9
10	57.4	–28.8136	=EXP(F10)	=G10/$G$6	=I9+H10
11	57.6	–28.3852	4.7E–13	0.00056	=I10+H11
12	57.8	–27.9617	7.18E–13	0.000855	0.002171
13	58	–27.5448	1.09E–12	0.001298	0.003469
…	Down to	…	…	…	…
73	70	–31.0395	3.31E–14	3.94E–05	1

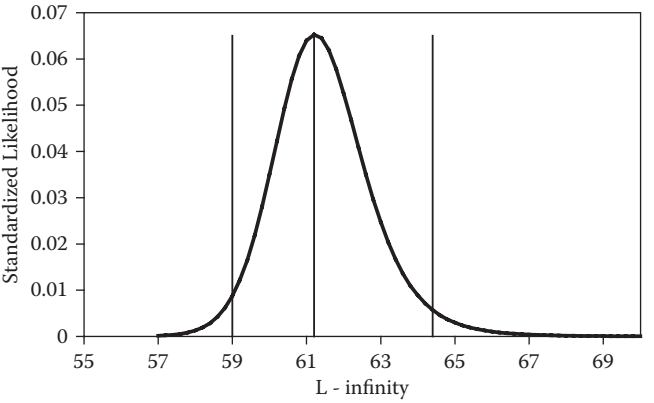

Fig ur e 8.3
Likelihood profile for the L_∞ estimate for the von Bertalanffy curve fitted to data in Example Box 8.2. The outer vertical lines at 59.0 and 64.4 are the closest steps to the 0.025 and 0.975 percentiles enclosing the central 95% around the optimum of 61.2 depicted by the central line. Note the asymmetry of the 95% confidence intervals.

relies on the fact that likelihood ratio tests asymptotically approach the χ^2 distribution as the sample size gets larger (see Chapter 9 for more details on likelihood ratio tests). This means that with the usual extent of real fisheries data, this method is only approximate. Likelihood ratio tests are, exactly as their name suggests, a ratio of two likelihoods, or if dealing with log-likelihoods, the subtraction of one from another, the two are equivalent:

$$\frac{L(\theta)_{Max}}{L(\theta)} = e^{\left(LL(\theta)_{Max} - LL(\theta)\right)} \tag{8.12}$$

where $L(\theta)$ is the likelihood of the θ parameters, the *Max* subscript denotes the maximum likelihood (assuming all other parameters are also optimally fitted), and $LL(\theta)$ is the equivalent log-likelihood.

The expected log-likelihoods for the actual confidence intervals for a single parameter, assuming all others remain at the optimum, are given by the following (Venzon and Moolgavkar, 1988):

$$2 \times \left[LL(\theta)_{Max} - LL(\theta) \right] \leq \chi^2_{1,1-\alpha}$$

$$LL(\theta) = LL(\theta)_{Max} - \frac{\chi^2_{1,1-\alpha}}{2} \tag{8.13}$$

where $\chi^2_{1,1-\alpha}$ is the $(1 - \alpha)$th quantile of the χ^2 distribution with 1 degree of freedom (e.g., for 95% confidence intervals, $\alpha = 0.95$ and $1 - \alpha = 0.05$, $\chi^2_{1,1-\alpha} = 3.84$). For a single parameter θ_i, the approximate 95% confidence intervals are therefore those values of θ_i for which two times the difference between the

EXAMPLE BOX 8.3

Approximate likelihood profile confidence intervals. Extend Example Box 8.2 as below. First use the solver to maximize B6 (the maximum log-likelihood) by changing the parameters in cells B2:B5. Given the optimum parameter estimates, we can calculate the log-likelihood expected at the 95% confidence intervals for Linf. In this instance the optimum LL is –23.6266, and subtracting 1.92 (E4) from that gives –25.5474. Run the solver again, only this time, rather than maximizing the target cell (B6), have the solver vary cell B2 only until the target cell has a value of –25.5474. It may be necessary to slightly vary the value in B2 to start. If the search finds the upper confidence interval, you will need to set the starting value of Linf below the optimum in order that the search will move to the lower confidence interval. Compare the results with the percentile confidence intervals produced in Example Box 8.2. With the true likelihood profiles the other parameters were also optimized rather than staying constant as here.

	A	B	C	D	E
2	**Linf**	62.20844			
3	**K**	0.29631			
4	**t0**	–0.0573	**Chi2**	=CHIINV(0.05,1)	=D4/2
5	**sigma**	1.489579			
6	**LL**	–25.5474			

corresponding log-likelihood and the overall optimal log-likelihood is less than or equal to 3.84 ($\chi^2_{1,0.05} = 3.84$). Alternatively, one can search for the θ_i that generates a log-likelihood equal to the maximum log-likelihood minus half the required χ^2 value (i.e., $LL(\theta)_{Max} - 1.92$; see Equation 8.13, Example Box 8.3).

As there are often parameter correlations within fisheries models (and many other disciplines), the assumption that keeping other parameters at the value for which the overall optimal fit was obtained will introduce bias such that the interval estimates would tend to be smaller than they should be (Figure 8.4). Strictly, because they hold the other parameters constant at their original optimum, these are slices through the likelihood surface rather than profiles; they are only approximations.

The deterministic likelihood profile method may be applied to more than one parameter to avoid the problem of parameter correlation. The search for values that satisfy Equation 8.13 (e.g., with χ^2 set at 5.99 for two parameters, 2 degrees of freedom; i.e., $LL(\theta)_{Max} - 2.995$) is somewhat more complicated, but Excel's solver could be set to search for the requisite values if it were started in the correct direction. A quick and simple method for doing this in Excel is to use a table function to calculate the log-likelihood for many combinations

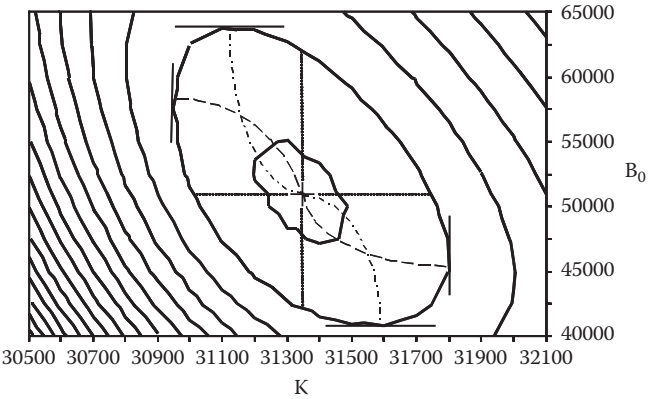

Fig ur e 8.4
Maximum likelihood contours for two parameters from a surplus production model of northern Australian tiger prawns from the Gulf of Carpentaria (see Chapter 11). With no parameter correlations the contours would be perfect circles. The second central ellipse here represents the 95% confidence intervals. The vertical and horizontal fine lines inside this ellipse indicate the approximate likelihood confidence intervals for single parameters (holding the other parameters constant), while the S-shaped dashed lines illustrate the true two-parameter likelihood profile confidence intervals that are wider than the single-parameter intervals.

of the two parameters, and then plot a contour diagram, adjusting the interval in the contours so that one contour matches the required $LL(\theta)_{Max} - 2.995$ (Figure 8.4).

Clearly, when considering more than one parameter, the confidence intervals become wider (Figure 8.4). This only occurs if there is any correlation (covariance) between the parameters. If there were no parameter correlation, then there would be no interference between the parameters and the confidence intervals would be unbiased and circular.

The empirical likelihood profile method produced confidence intervals that were asymmetric around the optimal value (58.9 to 61.2 to 64.4; a ratio of 1:1.47), whereas the deterministic approximation tends to be symmetric (60.24 to 61.2 to 62.21; a ratio of 1:1). Both methods are approximations, but the deterministic approximation has more assumptions and tends to produce narrower or underestimated confidence bounds.

8.5 Percentile Likelihood Profiles for Model Outputs

While the generation of likelihood profiles for model parameters is relatively straightforward, generally the interest in stock assessment models relates to model outputs that are not directly estimated as parameters. For example, an

assessment model might estimate stock biomass, or the maximum sustainable yield, or some other performance measure that is an indirect output from a model. Generating a likelihood profile for such model outputs can be produced by adding a penalty term to the log-likelihood that constrains the likelihood to the optimal target, as in Equation 8.14. In this way, the impact on the log-likelihood of moving away from the optimum can be characterized.

$$LL = LL + w\left(\frac{\text{output} - \text{target}}{\text{target}}\right)^2 \tag{8.14}$$

where *LL* is the log-likelihood, *output* is the variable of interest, *target* is the optimum for that variable, and w is a weighting factor. The weighting factor should be as large as possible to generate the narrowest likelihood profile while still being able to converge on a solution.

As an example, we will use a simple surplus production model similar to those introduced briefly in Chapter 2 and examined in detail in Chapter 11 (Example Box 8.4). These models are sometimes referred to as biomass dynamic models (Hilborn and Walters, 1992) because the stock dynamics are usually represented solely in terms of changing levels of biomass. Thus, the biomass at time $t + 1$ derives from the biomass at time t multiplied by some composite terms reflecting reproduction and growth, minus whatever catch has been taken.

$$B_{t+1} = B_t + rB_t\left(1 - \frac{B_t}{K}\right) - C_t \tag{8.15}$$

where B_t is the stock biomass at time t, r is the intrinsic rate of natural increase, K is known as the carrying capacity, and C_t is the catch taken in period t. The model can be fitted to catch rate data using the relationship

$$\frac{C}{E} = qB_t e^{\varepsilon} \tag{8.16}$$

where C/E is the catch rate data and q is the catchability coefficient (defined as the proportion of the exploitable stock biomass taken with one unit of effort). Given a time series of catches and catch rates, it is possible to suggest a time series of stock biomasses that is consistent with the data, and conclusions can be drawn about the productivity of the resource. Equations 8.15 and 8.16 represent a simple nonequilibrium version of a model first proposed by Schaefer (1954, 1957). For our purposes here, we also need to know that the maximum sustainable yield may be estimated once the model is fitted, using (see Chapter 11)

$$MSY = \frac{rK}{4} \tag{8.17}$$

EXAMPLE BOX 8.4

A nonequilibrium Schaefer surplus production model fitted to abalone data from the southwest of Tasmania (Table 8.1). Copy the data into columns A to C, down to row 33. In D11 put Equation 8.15 as MAX(D10+$B$2 *D10*(1–(D10/$B$3))–B10,100), and copy down to D33. The max function ensures that the stock biomass cannot go extinct when using the solver. In E10 put =((D10+D11)/2)*$B$5, the mid-year biomass in a given year multiplied by the estimate of q, and copy down to E32 (in E33 put =D33*$B$5). In F10 put =Ln(C10/D10) to generate the contributions to the q estimate. In G10 put =(Ln(C10)–Ln(E10))^2 to generate the residuals for lognormal random errors. Select F10:G10 and copy down to row 33. Complete the worksheet by putting =exp(average($F$10:$F$33)) into B5 to calculate the closed form of q (see Equation 8.21). Count the number of observations by putting =count(C10:C33) into B7, and to estimate the standard deviation of the residuals put =sqrt(sum(G10:G33)/B7) into B8 (see Equation 3.17). Plot columns C and E against column A (use different colours) to replicate something like Figure 8.5. The parameters shown are close to optimum. Put =–(B7/2)*(LN(2*PI())+ 2*LN(B8)+1) into B6 to gain the log-likelihood (Equation 11.25 and Appendix 11.3). Maximize B6 using the solver and changing B2:B4. In the solver options, turn on automatic scaling, increase the required precision, and decrease the convergence limits. Cells C1:G7 relate to the MCMC to be developed later (Example Box 8.5). In E5 put =NORMINV (RAND(),0,1) and the other numbers in C1:E5 are constants.

	A	B	C	D	E	F	G
1		**Param**	**Scale**	**Iteration**	15000	**Bcurr**	=D33
2	**re**	0.39065	0.025	**N**	15000	**Bcurr/K**	=G1/B3
3	**Ki**	8844.976	2.5	**np**	3	**MSY**	=B2*B3/4
4	**B0**	3170.305	2.5	**Step**	1	**Output**	
5	**qi**	0.000373	0.0148	**Norminv**	–1.0680		
6	**LL**	41.7623					
7	**n**	24				**Weight w**	0
8	σ	0.04247					
9	**Year**	**CatchT**	**CE**	**Model B**	**E(CE)**	**Expt_q**	**Ln(I–I)**
10	1985	1018.884	1.00000	=$B$4	1.1414	–8.0616	0.0175
11	1986	742.347	1.0957	2946.982	1.1042	–7.8968	0.0001
12	1987	868.023	1.1303	2971.163	1.0908	–7.8742	0.0013
13	1988	715.104	1.1466	2873.924	1.0806	–7.8266	0.0035
	Extend	Down to	Row 33				

continued

EXAMPLE BOX 8.4 (continued)

Development of a likelihood profile for the MSY parameter for the surplus production model fitted in Example Box 8.4. Put potential values of MSY into H10:H55 (650:1,100 in steps of 10). Alter the log-likelihood in B6 to become =–(B7/2)*(LN(2*PI())+2*LN(B8)+1)+G7*((G3–G4)/G3)^2. Copy each MSY value in turn from column H into cell G4 and re-solve, maximizing the log-likelihood in B6 by varying B2:B4, copying the optimum maximum log-likelihood from B6 down into column I. Either do this manually or implement the following macro and then plot column K against H (cf. Figure 8.6). Try different weights in G7. Once you complete the MCMC section below and develop a marginal posterior for the same MSY, you will see that a weight of 100 is satisfactory in this case. The profile likelihoods in column K should sum to 1.0. If they are cumulatively summed (in column L), it is possible to search for the MSY trials that most closely correspond to the 0.025 and 0.975 percentiles, as in Figure 8.6. The optimum values are stored in J3:J5 for reference.

```
Sub Profile()
  Application.ScreenUpdating = False
  For i = 1 To 46      ' or however many trials you decide
    Range("H9").Select
    ActiveCell.Offset(i, 0).Range("A1").Select
    Selection.Copy
    Range("G4").Select
    Selection.PasteSpecial Paste:=xlPasteValues
      ' other terms deleted for brevity
    SolverOk SetCell:="$B$6", MaxMinVal:=1, ValueOf:="0",
      ByChange:="$B$2:$B$4"
    SolverSolve (True)
    Range("B6").Select
    Selection.Copy
    Range("I9").Select
    ActiveCell.Offset(i, 0).Range("A1").Select
    Selection.PasteSpecial Paste:=xlPasteValues
  Next i
  Range("A1").Select
  Application.ScreenUpdating = True
End Sub
```

continued

EXAMPLE BOX 8.4 (continued)

	F	G	H	I	J	K
1	**Bcurr**	5344.535				
2	**Bcurr/K**	0.5687			**Optimum**	
3	**MSY**	882.5			0.39065	
4	**Output**	650			8844.976	
5					3170.305	
6						
7	**Weight w**	100				
8					=SUM(J10:J55)	
9	**Expt_q**	**Ln(I–I)**	**MSY**	**LL**	**Likelihood**	**Profile**
10	–8.1197	0.0119	650	48.23374	=EXP(–I10)	=J10/$J$8
11	–7.9644	0.0001	660	47.65001	=EXP(–I11	=J11/$J$8
12	–7.9473	0.0019	670	47.09313	3.5295E–21	0.000329

TABLe 8.1

The Abalone (*Haliotis rubra*) Catch and Catch Rates, CE (Standardized by Diver, Month, and Reporting Block) for the Years 1985 to 2008, from the Southwest of Tasmania (see Figure 8.5)

Year	Catch t	CE	Year	Catch t	CE
1985	1018.884	1.0000	1997	657.496	2.0515
1986	742.347	1.0957	1998	480.660	2.1244
1987	868.023	1.1303	1999	645.049	2.2148
1988	715.104	1.1466	2000	960.741	2.2531
1989	585.651	1.1873	2001	940.434	2.1051
1990	532.214	1.2018	2002	911.717	2.0820
1991	566.507	1.2652	2003	954.589	2.0087
1992	611.126	1.3199	2004	935.142	1.9226
1993	548.256	1.4284	2005	940.138	1.8703
1994	499.406	1.4772	2006	951.849	1.8776
1995	478.919	1.6853	2007	1030.043	1.8496
1996	427.787	1.9195	2008	984.599	1.7271

In addition, we need a number of other equations to anchor the dynamics to a fishery and to fit the model. Assuming that the error in Equation 8.16 is multiplicative and lognormal with a constant variance (i.e., $I_t = qB_t e^{\varepsilon}$, where $\varepsilon = N(0, \sigma^2)$), then the estimates of the model parameters (B_0, r, q, and K) are obtained by maximizing the appropriate likelihood function:

$$L\left(\text{data}\middle| B_0, r, K, q\right) = \prod_t \frac{1}{I_t\sqrt{2\pi}\hat{\sigma}} e^{\frac{-\left(\text{Ln}\, I_t - \text{Ln}\, \hat{I}_t\right)^2}{2\hat{\sigma}^2}} \tag{8.18}$$

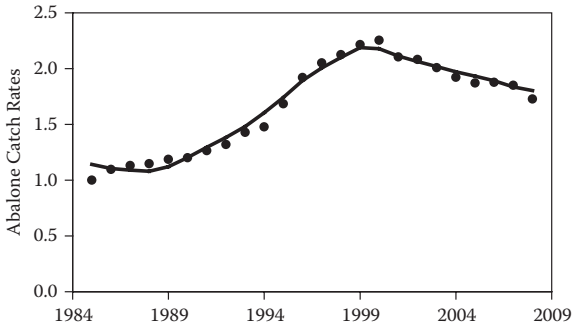

Fig ur e 8.5
Standardized catch rates for abalone from southwest Tasmania with the predicted catch rates from the simple surplus production model in Example Box 8.4. Data from Table 8.1.

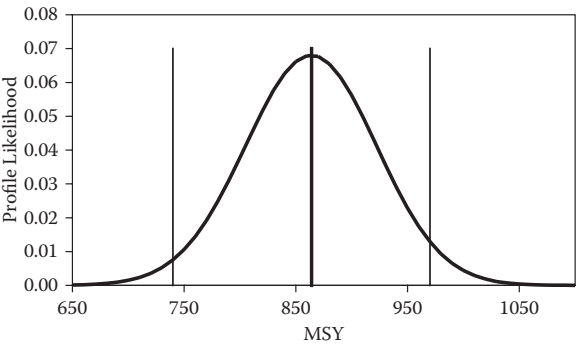

Fig ur e 8.6
The estimated likelihood profile for MSY for abalone from the southwest of Tasmania, as determined in Example Box 8.4.

where $L(\text{data}|B_0,r,K,q)$ is the likelihood of the data given the parameters, the product is overall years (t) for which CPUE data are available, and (Neter et al., 1996, p. 34)

$$\hat{\sigma}^2 = \sum_t \frac{\left(\text{Ln}\, I_t - \text{Ln}\, \hat{I}_t\right)^2}{n} \tag{8.19}$$

and n is the number of observations (maximum likelihood estimate of the variance uses n rather than $n-1$). A simplified version of the same maximum likelihood equation is

$$LL = -\frac{n}{2}\left(Ln(2\pi) + 2Ln\left(\hat{\sigma}\right) + 1\right) \tag{8.20}$$

(see Appendix 11.3 for its derivation).

The catchability, q, simply scales the catch rates to match the units of the estimated biomass levels and, in this case, assumes a linear relationship between catch rates and exploitable stock size. An estimate of q, which maximizes Equation 8.18, is given by the geometric average of the time series of individual q estimates (see Appendix 11.2 for the derivation); thus,

$$\hat{q} = e^{\frac{1}{n}\sum \mathrm{Ln}\left(\frac{I_t}{\hat{B}_t}\right)} \tag{8.21}$$

Alternatively, one could estimate the q value directly using the solver, but the value thus determined should differ from that derived from Equation 8.21 by only a small fraction. Such closed-form estimates, as in Equation 8.21, are valuable because, on average, the model tends to be quicker to fit and more robust when it has fewer directly estimated parameters.

A surprising thing here is that such a simple model can generate seemingly reasonable management advice. It is also surprising that catch rates appear to be meaningful for an abalone species, which because of their aggregated nature and the hand-collected nature of the fishery are notorious for having hyperstable catch rates. However, in Tasmania, catch rates do appear to be informative, although in this example no attention has been given to any improvements in fishing methods. The advent of global position receivers and other practical methods used in this diver fishery has undoubtedly improved catch rates for a given number of hours underwater. What this means is that the catchability, or the effectiveness of a unit of effort, is likely to have increased through time so that, for example, catch rates in the 2000s are of a different quality than those in the 1990s. Such "effort creep" or "technology creep" has the effect of biasing such analyses high so that they become more risk prone. Keep in mind that this simple model is designed simply to illustrate the methods and not to provide specific management advice. Rather, more details than we need here are required to provide usable advice while capturing more of the uncertainty in the system and data.

8.6 Markov Chain Monte Carlo (MCMC)

Parameter optimization can be envisaged as finding the location of the maximum likelihood on the multidimensional likelihood surface implied by a model and a given data set. If the likelihood surface is very steep in the vicinity of the optimum, then the uncertainty associated with those parameters would be relatively small. Conversely, if the likelihood surface was comparatively flat near the optimum, this would imply that similar likelihoods could be obtained from rather different parameter values, and the uncertainty around those parameters would be relatively high. If we assume that

the likelihood surface is multivariate normal in the vicinity of the optimum, then we could use asymptotic standard errors to define confidence intervals around parameter estimates. However, for many variables and model outputs in fisheries that might be a very strong assumption. Ideally, we would use methods that independently characterized the likelihood surface in the vicinity of the optimum solution. If we could manage that, we could use the equivalent of percentile methods to provide estimates of the confidence intervals around parameters. It would be possible to produce a likelihood profile for two parameters at once using the solver as described in the previous section. However, this would be clumsy and become impractical if there were more than two parameters. What is required is a way of integrating across many dimensions at once to generate a multidimensional likelihood profile. Fortunately, there are numerous methods for doing this (Gelman et al., 2004) including Sampling Importance Resampling (SIR) and Markov chain Monte Carlo (MCMC). Here we will focus on the use of a MCMC method. This approach has become much more common with the expansion in the use of Bayesian statistics that has occurred in fisheries modelling (Punt and Hilborn, 1997). As described in Chapter 3, when using Bayesian methods, a common problem is to characterize a multidimensional posterior distribution. If the prior information is omitted from a Bayesian analysis (as a strategy for providing truly uninformative priors), it becomes equivalent to a multidimensional likelihood profile. There are numerous alternative algorithms for conducting an MCMC, but we will focus on a relatively flexible approach called the Gibbs sampler (which is formally termed alternating conditional sampling).

An MCMC uses a Markov chain to trace over the multidimensional likelihood surface. A Markov chain describes a process whereby each state is determined probabilistically from the previous state. A random walk would constitute one form of Markov chain, and its final state would be a description of a random distribution. However, the aim here is to produce a Markov chain whose final state, the so-called stationary distribution, provides a description of the target or posterior distribution. The Markov chain starts with some combination of parameter values and the model being used that defines a location in the likelihood space. Depending on the set of parameter values, the likelihood can obviously be small or large. The MCMC process entails iteratively stepping through the parameter space following a set of rules based on the relative likelihood of each candidate parameter set to determine which steps will become part of the Markov chain, that is, which steps will be accepted and which rejected. Each step of the process entails the production of a new candidate set of parameter values, which is done stochastically (hence Markov chain) and, in the Gibbs sampler, one parameter at a time (Casella and George, 1992). Each of these new candidate sets of parameters, combined with the available data and the model, defines a new likelihood. Whether this new parameter combination is accepted as the next step in the Markov chain depends on how much the likelihood has been

changed. In all cases where the likelihood increases the step is accepted. But where the likelihood decreases it can still be accepted if the ratio of the new likelihood relative to the old likelihood is larger than some uniform random number (between 0 and 1). Thus, a Markov chain can be generated if the likelihood of an initial set of parameters θ_t given a set of data x can be defined as follows:

$$L_t = p\left(\theta_t \middle| x\right) \tag{8.22}$$

If we then select a new trial or candidate parameter set θ^* by randomly incrementing one of the parameters in θ_t($\theta^* = \theta_i + \Delta\theta_i$), this will alter the implied likelihood:

$$L^* = p\left(\theta^* \middle| x\right) \tag{8.23}$$

If the ratio of these likelihoods is greater than 1, then the jump from θ_t to θ^* is accepted into the Markov chain.

$$r = \frac{L^*}{L_t} = \frac{p\left(\theta^* \middle| x\right)}{p\left(\theta_t \middle| x\right)} > 1.0 \tag{8.24}$$

Alternatively, if the ratio is less than 1, then the jump is only accepted if the ratio r is greater than a newly selected uniform random number:

$$\theta_{t+1} = \begin{cases} \theta^* & \text{if } r > U(0,1) \\ \theta_t & \text{otherwise} \end{cases} \tag{8.25}$$

Strictly, only Equation 8.25 is required because where the new likelihood is greater than the original it will satisfy the top clause of Equation 8.25, but both are given to improve the clarity of the algorithm. If the candidate parameter set is rejected, it reverts to the original and the cycle starts again with a new trial parameter set. As the Markov chain develops it should trace out a multidimensional volume in parameter space. After sufficient iterations it should converge on the stationary distribution.

The jumping distribution defines how each parameter is incremented when adopting new candidate parameter sets. There are many options available, but commonly, for each parameter in turn in the set θ, a standard normal random deviate $N(0,1)$ is generated, and this is scaled, α_i, to suit the scale of the parameter, i, being incremented.

$$\theta^* = \theta_{t,i} + N(0,1).\alpha_i \tag{8.26}$$

This scaling by α_i is important because if the jumps in parameter space are too large, then the outcome may be unstable, but if they are too small, it may take an enormous number of iterations to converge on a stationary distribution. There is an element of trial and error in this procedure with no fixed or simple rule as to what scaling factor to use. The acceptance rate of the new candidate values is an indicator of the efficiency of the performance. A simple rule of thumb might be to scale the normal random deviate to approximate between 0.5 and 1.0% of the original parameter value. There are other MCMC algorithms that vary multiple parameters at each iteration; however, if there are parameter correlations, this can complicate the selection of new candidate parameter sets. Cycling through each of the parameters in θ in turn, using something like Equation 8.26, is relatively simple and is known as the Gibbs sampler (Example Box 8.5). There are numerous algorithms for conducting Markov chain Monte Carlo simulations with a growing literature devoted to this subject (Gelman et al., 2004).

An example MCMC run, using Example Boxes 8.4 and 8.5, illustrates that while the individual parameters might show evidence of serial correlation within the MCMC, the model output used (MSY estimate) produces the desired "hairy caterpillar" that indicates an acceptable result (Figure 8.7). That there is a strong correlation between the r and K parameters was explained in Chapter 2. This means that the negative correlation between these two parameters effectively cancels out any serial correlation in the MCMC runs when we consider the MSY output (which is a combination of both the r and K parameters; Figure 8.7).

Rather than simply using graphical indicators (as in Figure 8.7), diagnostic statistics should be used to indicate whether convergence to a stationary distribution has occurred. There are many available, but here we will only mention some simple strategies. As with any nonlinear solver, it is a good idea to start the Gibbs sampler from a wide range of initial values (the output illustrated in Figure 8.7 was generated by starting the Gibbs sampler at the optimum solution). Any diagnostic tests for identifying whether the MCMC has achieved the target stationary distribution would literally consider the convergence of the different Markov sequences. Of course, it would be necessary to discard the so-called burn-in phase before making any comparisons. The burn-in phase is where the Markov chain is traversing the likelihood space before it starts to characterize the posterior distribution. Gelman et al. (2004) recommend the conservative option of discarding the first half of each sequence, but the actual fraction selected should be determined by inspection. The simplest diagnostic statistics involve comparing the means and variances of either different sequences, or different portions of the same sequence. If the mean values from different sequences (or subsets of a sequence) are not significantly different, then the sequence(s) can be said to have converged. Similarly, if the within-sequence (or subset of a sequence) variance is not significantly different from the between-sequence, then convergence can be identified. Using a single sequence may appear convenient,

EXAMPLE BOX 8.5

The macro needed implements a Gibbs sampler to run a MCMC using the Excel sheet set out as in Example Box 8.4. You could delete the parts of the worksheet relating to the likelihood profile (or set the weight in G7 to zero). The *i* is just a counter used in the macro, N is the number of iterations of the Markov chain, np is the number of parameters, and step is how often a candidate parameter set should be considered for acceptance. In Example Box 8.4, Norminv in E5 is a standard normal variate, and the scales are the multipliers used to adjust each normal random deviate to suit the scale of the parameter being varied; see Equation 8.22. The output is placed in B40:F15040.

```
Sub Do_MCMC()
Dim N, np, step, Count, i, j As Integer
Dim func0, func1, trial As Double
Dim Scales(4), pars(4), trialpar(4), arate(4) As Double
  Range("B40:F15040").Select
  Selection.ClearContents
  Range("A1").Select
N = Cells(2, 5).Value       ' iterations in the MCMC
np = Cells(3, 5).Value      ' number of parameters
step = Cells(4, 5).Value    ' how often to store results
For i = 1 To np             ' Get the parameters and scales
  Scales(i) = Cells(1 + i, 3).Value
  pars(i) = Cells(1 + i, 2).Value
Next i
func0 = Cells(6, 2).Value   ' posterior
Count = 1
For j = 1 To N * step       ' Start the MCMC loop
  For i = 1 To np ' loop through parameters
    Calculate
    trial = pars(i) + Cells(5, 5).Value * Scales(i)
    Cells(1 + i, 2).Value = trial
    func1 = Cells(6, 2).Value     ' New log-likelihood
    If (Rnd < Exp(func1 - func0)) Then  ' Equ 8.24
      trialpar(i) = trial
      func0 = func1
      arate(i) = arate(i) + 1
      Else: trialpar(i) = pars(i)
    End If
  Next i                    ' End of Gibbs sampler
```

continued

EXAMPLE BOX 8.5 (continued)

The rest of the macro needed implements a Gibbs sampler to run a MCMC using the Excel sheet set out as in Example Box 8.4.

```
  If ((j Mod step) = 0) Then
    Cells(1, 5).Value = Count
    For i = 1 To np
      Cells(Count + 40, 1 + i).Value = trialpar(i)
      pars(i) = trialpar(i)
    Next i
    Cells(Count + 40, np + 2).Value = func0
    msy = Cells(3, 7).Value
    Cells(Count + 40, np + 3).Value = msy
    Count = Count + 1
  End If
  If ((j Mod 100) = 0) Then              ' controls screen
    Application.ScreenUpdating = True
    Application.ScreenUpdating = False
  End If
Next j ' End of MCMC
For i = 1 To np
  Cells(40, 1 + i).Value = arate(i) / N
Next i
End Sub
```

In the example run illustrated in Figure 8.7, the MSY values ranged from 655.08 up to 1,094.35 (this will differ from run to run). In cell J45 put =min(F41:F15040) and in J46 put =max(F41:F15040). The marginal distribution can be generated (for each variable) by finding the range and setting up a series of bins, in this case 650 to 1,100 in steps of 10. In cell K48 place =countif($F$41:$F$15040,">="&J48)–countif($F$41:$F$15040,">="&J49) and copy down to row 92 to give the relative frequency of occurrence of the different MSY values. In K46 put =sum(K48:K92), to ensure all 15,000 are captured. In cell L48 put =K48/$K$46 and copy down to row 92. Copy the final likelihood profile from the first use of Example Box 8.4 into M47:M93. Plot L48:L92 against J48:J92. Add M48:M92 to this graph. Put =percentile($F$41:$F$15040, 0.025) into cell N48, and in N49 put =N48. In N50 put =median(F41:F15040), and in N52 put =percentile($F$41: $F$15040,0.975). Add the groups in N48:O49, in N50:O51, etc., to the graph to mimic Figure 8.8. Compare a graph with fifteen thousand iterations with one of five thousand iterations.

continued

EXAMPLE BOX 8.5 (continued)

	J	K	L	M	N	O
45	655.0839					
46	1094.346	15000				
47	**MSY**	**Freq**	**Prop**	**Profile**		
48	650	1.00	0.00007	0.00011	752.5	0
49	660	2.00	0.00013	0.00019	=N48	0.075
50	670	3.00	0.00020	0.00033	868.4	0
51	680	4.00	0.00027	0.00056	=N50	0.075
52	690	6.00	0.00040	0.00092	984.5	0
53	700	15.00	0.00100	0.00149	=N52	0.075
54	710	35.00	0.00233	0.00233		
55	720	61.00	0.00407	0.00354		
56	Extend	Down to	Row 93	…		

but if convergence is relatively slow and this was unknown, then relying on a single sequence may provide incorrect answers. The title of Gelman and Rubin (1992) states the problem clearly: "A single sequence from the Gibbs sampler gives a false sense of security." Suffice to say that it is better to use multiple starting points to give multiple sequences along with an array of diagnostic statistics and graphics to ensure that the conclusions that one draws from any MCMC simulation are not spurious (Gelman et al., 2004). Such approaches are termed computer-intensive for a good reason.

The marginal distribution of the MSY (Example Box 8.5, Figure 8.8) is not completely smooth but forms a close approximation to the likelihood profile generated for the MSY in Example Box 8.4 (with a weight of 100 given to the penalty term; Figure 8.8). To obtain a smoother marginal distribution, more than fifteen thousand iterations of the MCMC would be required. In the example illustrated the MCMC was started with the maximum likelihood optimum parameter set; if it had been started elsewhere, an initial burn-in period would have been required, which should be omitted from the marginal distribution as not representing plausible parameter combinations.

In a Bayesian framework the MCMC can be used to characterize the posterior distribution. In this case, when estimating the log-likelihood value for each new trial parameter value it would also be necessary to include the prior probability for each of the parameter values. If the MCMC is used to fit the model, then because it would be unlikely that the MCMC could start near the optimum parameter set, there would need to be some, often long, burn-in series aimed at moving the parameter combinations into the vicinity of the main posterior distribution. With multiparameter models

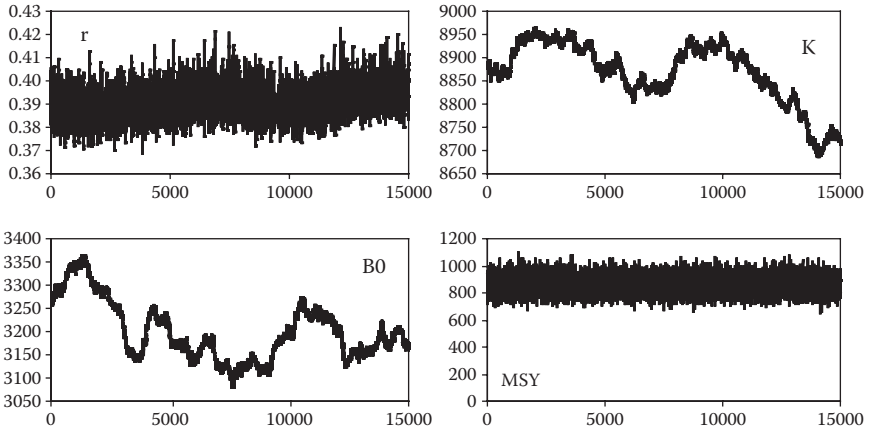

Fig ur e 8.7
Outputs from the Gibbs sampler in Example Boxes 8.4 and 8.5. The top left is the Markov chain for the *r* parameter, the top right is the *K* series, the bottom left is the B0 series, and the bottom right is the model output MSY series.

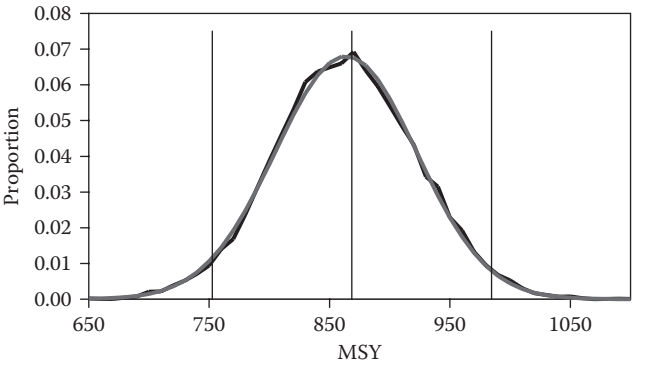

Fig ur e 8.8
The marginal frequency distribution of fifteen thousand iterations of the Gibbs sampler for MSY from Example Box 8.5 (left-hand *Y* axis). The smooth fine line is the likelihood profile from Example Box 8.4 with a weight of 100 (right-hand *Y* axis). The verticals represent the 95% confidence intervals and the median.

the number of iterations required can become enormous, and so this is invariably a very slow process.

In a real analytical situation the use of Excel to run an MCMC would be very slow and inefficient, and it is not recommended for anything except perhaps very small jobs (or, as here, for learning the principles). The algorithm and implementation expressed in Example Box 8.5 can be generalized to most problems, but this would not be sensible in Excel. As a means of understanding the process of stepping through a Markov chain, Excel is

excellent, but for any serious undertaking, one should be realistic and at least use R or WinBugs (both of which are freely available). However, if the algorithm expressed by Equations 8.23 to 8.26 and as code in Example Box 8.5 is compiled into some executable language, the analysis will run much more quickly. With stock assessments involving many parameters it is not uncommon to run millions of iterations to adequately describe the marginal distributions of the outputs of interest, and under such circumstances some serious computing can be required.

8.7 Concluding Remarks

Exactly how one decides to characterize the uncertainty around parameter estimates or model outputs depends on the circumstances and the properties of the model involved. In addition to the methods listed in this chapter, it is also possible to use bootstrapping methods, as described in Chapter 6. Keep in mind that it appears to be more common to underestimate uncertainty than to overestimate it. How such estimates of uncertainty are used varies greatly. But at least it means that it becomes possible to attribute levels of relative risk to different management decisions.