7

Monte Carlo Modelling

7.1 Monte Carlo Models

7.1.1 The u ses of Monte Carlo Modelling

Monte Carlo models and simulations are computer-intensive methods that can be used to test specific hypotheses and to answer less specific "what if" type questions, including projections into the future. Whichever task is at hand, one requires a simulation model of the fishery that can be compared with nature. It can differ from an assessment model in that it may include hypothetical relationships, parameters, and other modifications. Any modelling task that requires a projection or extrapolation of a model's implications into the future is best carried out using Monte Carlo simulation techniques. Risk assessment and the more complicated management strategy evaluation are all relatively recent developments that are becoming more commonly applied to fisheries around the world. All use simulation models to explore the implications of manipulating the management of the fishery concerned. The impact of uncertainty concerning any part of a model may also be investigated using simulations.

In this chapter we will consider aspects of modelling that are required when implementing simulations that include hypothetical or unknown components. Like other computer-intensive methods, Monte Carlo simulations require resampling from statistical distributions. In particular, they require resampling with replacement from theoretical probability density functions (pdfs). Detailed examples will be used to illustrate the principles discussed.

Strict Monte Carlo resampling differs from bootstrapping in that the distributions being resampled are parameterized pdfs and are not just empirically determined. Beware of confusion over terminology; some people refer to Monte Carlo resampling as parametric bootstrapping and to bootstrapping as nonparametric bootstrapping. As always, terminology is unimportant as long as one is clear about what is being resampled and how.

7.1.2 Types of u ncertainty

A major requirement for developing Monte Carlo simulations is to devise methods for including uncertainty about parameter estimates and other

165

components of a model. No ecological process is known perfectly, and this constitutes a problem when attempting to create a simulation that adequately captures a system's dynamic behaviour. This appears to be especially true when attempting to understand fishery dynamics. Rather than work entirely with deterministic models, simulations invariably include stochastic elements in which random variation around some model component(s) is included. In this way the implications of uncertainty may be included in a model analysis. There are a number of different types of uncertainty that can influence ecological modelling:

1. The appropriateness of a model. Fundamentally different, independent models can be used to describe the dynamics of a natural system; it is uncertain which model is the best representation.

2. Process error or uncertainty arises where a model has been defined in a deterministic manner but some of the components actually vary randomly through time (for example, virgin biomass is usually defined as representing a constant, but this index of stock productivity is likely to change with environmental conditions).

3. A model of how a particular system operates in nature is accepted, but values for some of the parameters involved are unknown or cannot be independently estimated; these tend to be given fixed values (e.g., natural mortality is often given an assumed constant value in population models).

4. Observation error or uncertainty relates to how precisely a given model's parameters can be estimated given a particular set of data (this is what is estimated when constructing confidence intervals around a set of parameters).

5. Measurement error or uncertainty, where data being put into the model are only measured with error (e.g., ring count estimates from otoliths). This can easily be confounded with observation error.

When conducting a fishery's stock assessment, large parts of the model are often deterministic (process error, type 2 above, is ignored). Parameters are estimated along with some indication of the precision of the parameter estimates (type 4 above; observation error is recognized). However, very often there are parameters, such as natural mortality, that are only poorly known and are not easily estimated. In these cases values for such parameters are often set at an assumed value (type 3 above; uncertainty). The model can be said to be conditioned on such parameters. One of the advantages of Bayesian methodology, as it is used in fisheries (Punt and Hilborn, 1997), is that it is possible to include a prior distribution that attempts to capture our uncertainty for such parameters. A common alternative is to test the sensitivity of the assessment to these constant conditioning parameters (the data are repeatedly refit to the model, and each time the conditioning parameters are given

different values from a predetermined range of values). If there were more than one conditioning parameter, then testing the sensitivity of all possible combinations of values would become more and more difficult as the number of parameters and their ranges increased. Rather than completing such a set of discrete combinations, it would be more efficient to conduct a simulation in which values for the conditioning parameters were selected at random from defined ranges. Such simulations would be Monte Carlo simulations.

By comparing simulated outcomes against observations from nature it is possible to test hypotheses. Simulation models are generally stochastic models because at least some of the variables or parameters have uncertain values. In these cases, the values for the variables or parameters are determined anew during each *run* of the Monte Carlo model, by sampling "with replacement" from a known probability distribution.

7.2 Practical Requirements

7.2.1 The Model Definition

An obvious requirement for Monte Carlo simulation modelling is a formal simulation model. This may closely resemble an assessment model, but in addition it will have a number of unknown components (variables, parameters, or sequence events) that need to be included. The final requirement is to define the expected probability density functions that are to be used to generate values for each uncertain model component.

7.2.2 r andom Numbers

Fundamental to the process of stochastic simulation modelling is a requirement to obtain random values from under a variety of probability density functions. In Chapter 3, a number of standard pdfs were introduced. Given the parameters of a pdf for a given variable, we saw how to calculate the likelihood for a particular value of that variable. What we now require is a method of inverting that process. We want to be able to randomly select values of the variable concerned once we have selected a parameterized pdf.

There are numerous algorithms for generating random values from given pdfs (Hastings and Peacock, 1975), and modellers who use Monte Carlo simulations tend to collect different techniques for doing this (Press et al., 1989). Fundamental to many methods is the generation of random numbers from a standard uniform variate (i.e., values from 0 to 1 with equal probability). Press et al. (1989) discuss a number of algorithms for generating uniform random numbers.

The pseudorandom number generators used in computers are of variable quality. Such generators are so fundamental to computer-intensive methods that it is a good idea to test any random number generator that you may use. A simple test is to count the number of times the random number generator gives particular values and compare those with the number of times such values would be expected; one could use either a *G* test or a χ^2 test (Example Box 7.1).

In tests of the efficiency of random number generators, if one chooses a significance level of 5% (i.e., $\chi^2 > 16.919$, for 9 degrees of freedom), the comparison of the frequency of classes of random numbers with their expected values would be expected to fail 5% of the time. The random number generator in Excel (Example Box 7.1) will generate failure rates that average 5%; in some particular tests it is greater than 5%, and in others it is less than 5% (e.g., the average of ten trials of one thousand replicates each using Example Box 7.1 was 0.0504).

7.2.3 Nonuniform r andom Numbers

The fundamental idea of sampling under a pdf curve is relatively simple, although how it is implemented for each pdf is not necessarily straightforward. For continuous variates the area under a pdf curve must sum to 1. If there is a method for generating random numbers from 0 to 1, then these can be used to generate target areas under a curve. For example, given a normal distribution, $N(0,1)$, i.e., mean zero and variance = 1, then we know that a random number (area under the curve) of 0.05 will give an *X* variate value of –1.6448, a random area of 0.5 will give a value of 0, and one of 0.9 will give 1.2815 (Figure 7.1). To obtain these inverse pdf values one requires a mechanism by which the pdf curve is integrated until the required random probability is attained (generated from a random number taken between 0 and 1), and then the respective target value of the *X* variate is determined (Figure 7.1).

It must be remembered that for many statistical distributions there are either direct equations for generating the required random numbers (Hastings and Peacock, 1975), or different, more efficient algorithms for their generation.

If someone wishes to write an independent computer program to conduct a simulation, then a formal algorithm will be necessary once it has been decided which pdf to use for the problem being considered (Press et al., 1989). Fortunately, in Excel many inverse pdfs are provided as built-in functions (Example Box 7.2, Figure 7.2), and one could write other user-defined functions.

How one conducts the integration under the curve will depend upon the probability density function being used (Press et al., 1989). This approach to generating random numbers is, in fact, only one out of many possible. It has only been identified here to assist in the development of intuitions about the process.

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 7.1

A Monte Carlo test of the random number generator function rand(). The =rand() function is copied down as far as A1001. Enter the number of Monte Carlo trials in E15. Select cells C2:C11, enter the array function =frequency(A2:A1001,B2:B11), and press <Ctrl><Shift><Enter>. This counts the number of times a value ≤ each respective bin value occurs in the column of random numbers. The χ^2 test is of whether the observed frequencies match the expected and column E has the (Obs – Expected)^2/Expected. This test should fail one time in twenty at $P = 0.05$. Instead of repeatedly pressing the F9 button, one should generate a macro to count the frequency of successes and failures. The only part of the macro that can be recorded is the activesheet.calculate (as <Shift><F9>). The rest of the algorithm must be added. The ability to write simple programs like this is required for Monte Carlo simulation modelling. Does using bins of 0.05 affect the result? Cell E14 is filled by the macro.

	A	B	C	D	E
1	=count(A2:A1001)	Bin	Observed	Expected	Chi2
2	=rand()	0.1	102	=$A$1/10	=(C2-D2)^2/D2
3	=rand()	0.2	97	=$A$1/10	=(C3-D3)^2/D3
4	=rand()	0.3	118	Copy down	3.24
5	Copy down	0.4	84	100	2.56
6	0.693034	0.5	107	100	0.49
7	0.502615	0.6	83	100	2.89
8	0.395583	0.7	99	100	0.01
9	0.230871	0.8	114	100	1.96
10	0.598551	0.9	99	100	0.01
11	0.579494	1	97	100	0.09
12	0.421189			Chi2	=sum(E2:E11)
13	0.705528		Do_Test	P	=chidist(E12,9)
14	0.197339			Significant	
15	0.275854			Replicate	1000

```
Sub Do_Test()                   ' attached to button on sheet
  Dim i As Integer, n As Integer ' good practice
  Randomize                      ' resets random number generator
  n = 0                          ' Sets counter to zero
  For i = 1 To 1000
    ActiveSheet.Calculate ' recalculates sheet
```

continued

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 7.1 (continued)

```
    If Cells(13, 5).Value < 0.05 Then n = n + 1
                            ' no End If as only one line
    Cells(14, 5).Value = n / i' interact with worksheet
    Cells(15, 5).Value = i
  Next i
End Sub
```
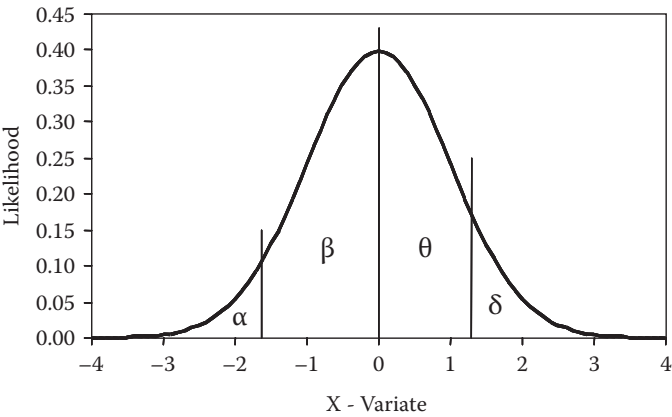

Fig ur e 7.1
Three different values of a variate, *X*, taken from a standard normal distribution N(0,1), using three different cumulative probabilities: $\alpha + \beta + \theta$. The area under the curve for α was 0.05, for $\alpha + \beta$ was 0.5, for $\alpha + \beta + \theta$ was 0.9, and for $\alpha + \beta + \theta + \delta$ was 1.0. The *X* values for each of the first three probabilities were –1.6448, 0, and 1.2815, respectively.

7.2.4 Other Practical Considerations

The selection of a suitable probability density function to represent a particular variable or parameter in a simulation is obviously an important step. There are many pdfs that were not considered in Chapter 3 but which are useful in simulations because they are so flexible and general. As usual, one can use whatever is most appropriate as long as one can defend the selection.

Commonly, Monte Carlo replicates are often more time-consuming than randomization or bootstrap replicates, and in the literature it is often found that the number of Monte Carlo replicates are limited. This is risky, as few replicates rarely produce a smooth distribution of expected outcomes. As with the randomizations and bootstraps, the reduction of noise in the results should improve with increasing numbers of Monte Carlo replicates.

EXAMPLE BOX 7.2

The generation of normally distributed random numbers. Instead of using the variance in their normal equations, Microsoft programmers elected to use the standard deviation. This unusual usage possibly derives from them using the standard normal distribution (N[0,1]) in which the variance equals the standard deviation. The frequency bins are extended down to B37, generating a range from 1 to 9 in steps of 0.25. Select C5:C37, type =frequency (A5:A1004,B5:B37), and press <Ctrl><Shift><Enter> to enter the array formula. Put the following function into D5 and copy down to D37: =($C$3/4)*normdist (B5,$B$1,$B$2,false). The first term accounts for the relative numbers and the bin sizes. Plot column C against B as a histogram and add column D to the same graph, but make the graph type for that data series a line graph (to mimic Figure 7.2). By repeatedly pressing F9 the random numbers will be recalculated, generating new normally distributed random numbers. Note the differences between the sample mean and standard deviation (C1:C2) and the parameters from which the sample was generated (B1:B2). Note how the graph alters each time and how closely or otherwise it resembles the expected. Try reducing the number of normally distributed random numbers (e.g., delete A105:A1004) and note the impact on the quality of the generated normal distribution. Try altering the value given in B2 and note its impact on the quality of the distribution generated. Clearly, large samples are best when there are many frequency categories. If you implement a χ^2 test on this sheet, remember that the test will be more severe than usual, as there will be many expected frequencies of less than 1, as in D5:D9 (Sokal and Rohlf, 1995). Note the use of rand() in column A. Compare C3 with D3.

|	A	B	C	D
1	**Mean**	5	=average(A5:A1004)	
2	**StDev**	1	=stdev(A5:A1004)	
3	Count		=count(A5:A1004)	=sum(D5:D37)
4	Inverse Normal Values	Bins	Frequency	Expected
5	=norminv(rand(),$B$1,$B$2)	1	0	0.017
6	=norminv(rand(),$B$1,$B$2)	1.25	0	0.044
7	=norminv(rand(),$B$1,$B$2)	1.5	0	0.109
8	Copy down to row 1004	1.75	0	0.254
9	4.829436	2	0	0.554
10	5.430025	2.25	0	1.137
11	3.227922	2.5	4	2.191
12	4.424067	2.75	4	3.967
13	6.961762	3	7	6.749

© 2011 by Taylor & Francis Group, LLC
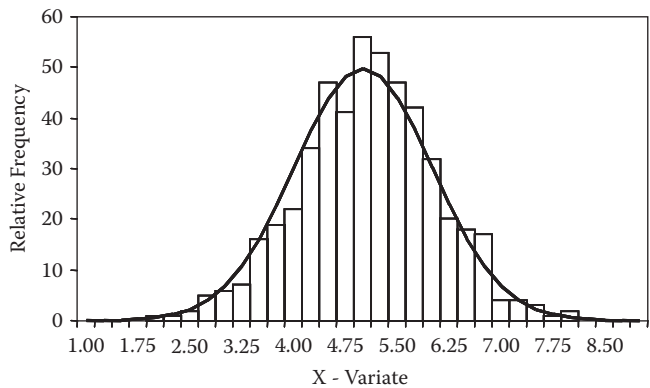
Fig ur e 7.2

A normal distribution generated from five hundred random numbers sampled from under a normal pdf with a mean of 5 and variance of 1. The solid curve denotes the expected relative frequency (Example Box 7.2). Note the imperfections despite there being five hundred random numbers contributing to the frequency distribution. Each sample of five hundred would generate a different relative frequency distribution. This should reinforce the idea that one should try to use as many Monte Carlo replicates as is practically possible.

7.3 A Simple Population Model

As a simple example, we will consider a population growing without density dependence:

$$N_{t+1} = rN_t \tag{7.1}$$

where N_t is the population size at time t, and r is the population growth rate. As was seen in Chapter 2, such a population has three possible behaviours, depending on the value of the growth rate r. If $r > 1$, then the population grows exponentially to infinity; if $r = 1$, the population is stable; and if $r < 0$, it declines exponentially to zero. This assumes that the growth rate is a constant. It seems much more likely that a density-independent species will be greatly affected by the environment and the growth rate is likely to vary above and below a value of 1. The question to be answered is whether a population with a randomly varying growth rate (having a mean value of 1 and a given variation σ) would be able to survive. Put another way, we could ask: How often would we expect such a species to go extinct over a given number of generations, and how is that extinction rate influenced by the variability in growth rate? The model becomes

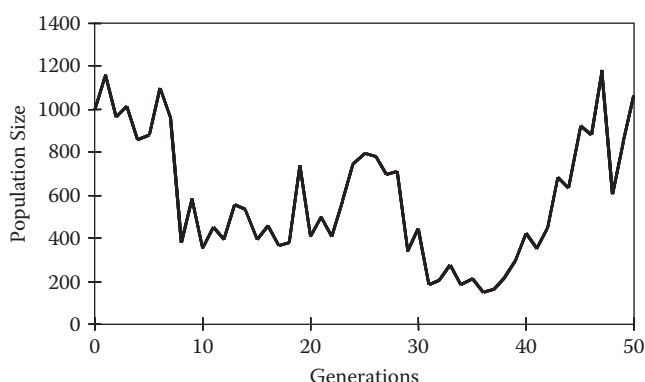$$N_{t+1} = r_{N(1,\sigma)}N_t \tag{7.2}$$

Fig ur e 7.3
A typical population trajectory through fifty generations when the growth rate for Equation 7.2 is set to a mean of 1 and a standard deviation of 0.3. A series of *r* values greater than 1 would lead to massive population growth, and similarly, a series of values less than 1 would lead to a marked reduction. While variation is low, most model runs do not go extinct and the population trajectory reflects the random walk of the *r* values around the mean value.

where the $r_{N(1,)}$ is a random normal deviate having a mean of 1 and a standard deviation of σ; we also need an artificial lower limit to the *r* value, say, 0.01 (Figure 7.3).

We need to limit the value of *r* in the simple model to positive numbers because if unconstrained normal random deviates were used, it would be possible to obtain a negative growth rate, which would mean instant extinction. We also need to define extinction; in this case, if the population falls below a single individual (i.e., $N_t < 1$), it can be regarded as extinct.

With the simulation model definition in Equation 7.2 (and the listed assumptions), it would be possible to generate the information necessary to illustrate the relationship between the proportion of simulation runs that go extinct in a given number of generations and the variability of the growth rate (Figure 7.4, Example Box 7.3). In a real exercise, one would also monitor the range and variability of population sizes, along with other measures of the model's performance. Despite the extreme simplicity of this model investigation the power of the method should be clear (Figure 7.4).

7.4 A Nonequilibrium Catch Curve

7.4.1 Ordinary Catch Curve Analysis

A more complex example will illustrate many of the ideas already discussed. As we saw in Chapter 2, if a single cohort were exposed to a constant
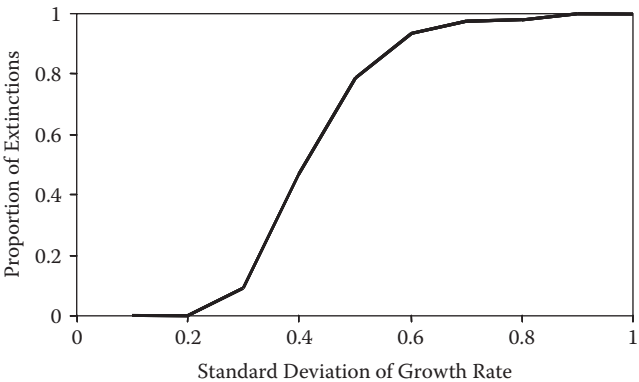
Fig ur e 7.4
Outcome of ten thousand Monte Carlo replicates of a population growing in a density-independent fashion with a growth rate described by a normal distribution with a mean of 1 and a standard deviation of σ. The graph compares the proportion of model runs that went extinct over fifty generations against the standard deviation of the growth rate.

mortality rate, then the numbers would be expected to decline exponentially following the relationship

$$N_t = N_0 e^{-(M+F)\,t} = N_0 e^{-Zt} = N_{t-1} e^{-Z} \tag{7.3}$$

where N_t is the cohort size at time t, and Z is the instantaneous total mortality rate (natural and fishing combined).

   If the total population concerned received constant recruitment and a constant mortality rate each year, then all cohorts would be identical and the numbers in each age class would be expected to decline exponentially. It is possible to follow the fate of a single cohort through a number of years or consider multiple cohorts in a single year; Equation 7.3 would apply in both cases, i.e., whether N_t referred to a cohort or a whole population (Figure 7.5). While it is obviously unlikely that any fished population adheres to any of these conditions (constant recruitment and mortality through time), the logarithmic transform of this relationship is sometimes used to make an estimate of the total mortality:

$$Ln(N_t) = Ln(N_0) - Zt \tag{7.4}$$

Given Equation 7.4 and the assumptions of constant recruitment and mortality, a linear regression of the log of numbers-at-age against age should produce a plot with a negative gradient of slope Z. Such an analysis is referred to as a catch curve (Figure 7.6), and by definition these relate to a fishery in equilibrium.

EXAMPLE BOX 7.3

A density-independent population growth model in which the growth rate is described by a normal distribution: $N(1,\sigma)$. Plot the numbers through time (column B) against time (column A) as a connected scattergram to obtain the equivalent to Figure 7.3. By varying the value in B1 and pressing F9 to recalculate the sheet, the population trajectory will alter and may go extinct. If the population goes extinct, then the text in C4 will appear (format it as bold and red).

Cell B56 is duplicated in B4 for visual convenience. The use of the max function in column C is to prevent negative deviates being produced. Instead of manually replicating model runs, one could use the macro listed at the bottom of the box. This will run one thousand Monte Carlo replicates and update cell B2 when finished. You could try to extend this macro by adding an outer loop so that it automates the investigation of the relationship between σ and the proportion going extinct (Figure 7.4). Investigate how this population model responds to lognormal random deviates, put =loginv(rand(),0,$B$1) into C7, copy down, and repeat your analysis. This is computer-intensive work, so try to automate as much as possible.

	A	B	C
1	**sigma σ**	0.3	*Standard deviation of rate of increase*
2	**Prop**	0.30	*Proportion going extinct*
3	**Initial**	1000	
4	**N50**	=B56	=IF(B4=0,"EXTINCT"," ")
5	**Time**	**Nt**	**Normal Random Deviate**
6	0	=B3	
7	1	=IF(B6*C7>1,B6*C7,0)	=max(norminv(rand(),1, $B$1),0.01)
8	2	=IF(B7*C8>1,B7*C8,0)	=max(norminv(rand(),1, $B$1),0.01)
9	3	=IF(B8*C9>1,B8*C9,0)	=max(norminv(rand(),1, $B$1),0.01)
10	4	Copy down to row 56	Copy down to row 56
11	5	1685.231	0.887778
12	6	1765.256	1.047486

```
Sub Do_Monte()
Dim i As Integer, n As Integer
Randomize
  Application.ScreenUpdating = False
  n = 0
  For i = 1 To 1000
    ActiveSheet.Calculate          ' Only part recordable
    If Cells(4, 2).Value = 0 Then n = n + 1
    Cells(2, 2).Value = n / i
  Next i
Application.ScreenUpdating = True
End Sub
```
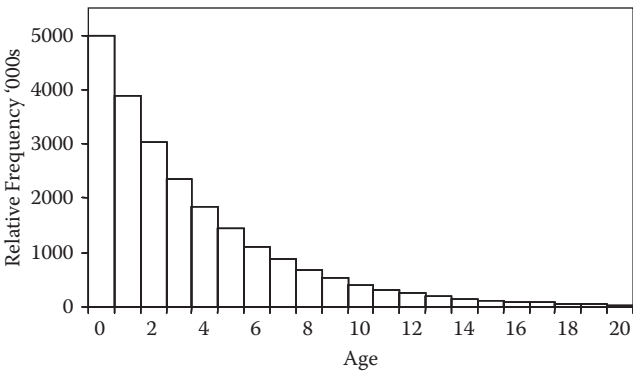
© 2011 by Taylor & Francis Group, LLC

Fig ur e 7.5
A population in equilibrium with a constant recruitment of 5 million animals and a constant rate of mortality, in this case $Z = 0.25$, being applied instantaneously, leading to an exponential decline. As recruitment is constant, the decline in the total population numbers reflects the decline in each cohort, i.e., there is equilibrium.
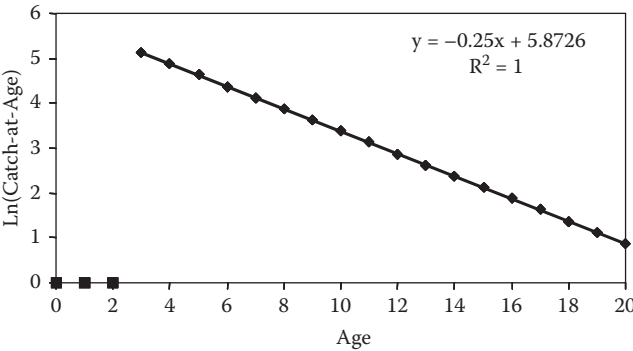

Fig ur e 7.6
A population in equilibrium with a constant recruitment of 5 million animals and a constant rate of mortality of $Z = 0.25$. The first three year classes are not selected by the fishery and hence are not sampled. The log of the relative frequency of each age class (the relative catch-at-age), plotted against age, generates a straight line whose gradient is an estimate of Z. This is the classical catch curve analysis (Example Box 7.4).

7.4.2 The influence of Sampling error

As well as assuming constant, equilibrium levels of recruitment and mortality, standard catch curve analyses also assume that samples precisely represent the population. Sadly, it is difficult to obtain a sample from a population (in which the fish are measured and aged) that provides a perfect representation of the relative abundance of each age class (the catch-at-age). It is natural to ask what difference sampling error would make to the catch curve analysis. The simplest way of determining this would be to take the perfect sample of
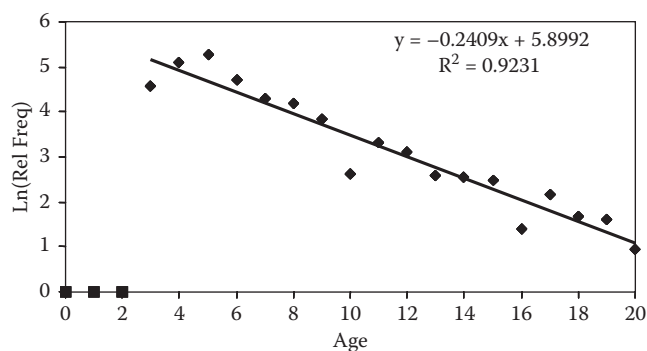
© 2011 by Taylor & Francis Group, LLC

Fig ur e 7.7

The same population data as in Figures 7.5 and 7.6 except that the true relative frequencies in a sample of 750 (cf. Figure 7.6) have had random variation added as described in Equation 7.5, with a σ of 15. Note that, in the Monte Carlo replicate illustrated, the linear regression now has a gradient of only 0.241 instead of the true value of 0.25.

size N, which gives rise to the true relative frequencies f_t for each age class t, and then add some random variation to each estimate of the relative frequency. This variability would need to be smaller for the older age classes in absolute terms; else, unrealistically large relative frequencies of older animals would be possible. Also, no frequencies less than 1 should be permitted:

$$\hat{f}_t = f_t + \varepsilon \times \frac{f_t}{\bar{f}} \tag{7.5}$$

where f_t is the true frequency, f bar is the overall mean frequency, f hat is the expected frequency, and ε is a normal random variate with mean zero and standard deviation σ (i.e., N[0,σ]). The last term weights each normal deviate so that the variation around larger true frequencies is greater than around the smaller true frequencies (i.e., we have a constant coefficient of variation and not a constant variance).

When this is done, the first set of random errors added to the "true" relative frequencies will give rise to a nonperfect straight line with a slightly different gradient from the "true" mortality rate (Figure 7.7, Example Box 7.4).

The actual gradient observed in a single trial has little meaning on its own. However, one can easily imagine taking the same age structure and adding a new set of random errors and repeating the analysis to derive further mortality estimates. If this process were to be repeated many times, we would end up with, say, one thousand Monte Carlo estimates of the Z value. Using these values to form a frequency distribution of outcomes, we could then determine the effect of different degrees of sampling error upon the estimation of Z (Figure 7.8). In addition, as with bootstrapping, we could determine approximate percentile confidence intervals on the estimate of Z (Example Box 7.5).

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 7.4

A classical catch curve analysis with sampling error included. N0 is the constant annual recruitment. B5 contains the age at which fish are first vulnerable to capture. Z is the total mortality (name cell B6 as Z), n is the sample size, and σ_s is the standard deviation of the sampling error. Column D takes the age at first capture into account. Sample size (B8) times the proportion of each age class in the total fished population =IF(D11>0,D11*$B$8/$D$9,"") is the "perfect sample" in E11:E31. To add sampling error, put =if(isnumber(E11),E11+ (norminv(rand(),0,$B$7)*E11/$E$9),0) into F11:F31. The total sample may now be smaller or greater than n. The E11/$E$9 generates the constant coefficient of variation that prevents excessively large numbers occurring in the older age classes (the lower frequencies). In column H, repeat the formula without the "E11+" to see the range of random variation decreasing with age. In G11 enter =if(F11>1,Ln(F11),0), which prevents the sampled frequency dropping below 1 and produces the required log of catch-at-age. Plot G14:G31 against column A as a scattergram; add a linear trend line to the data to mimic Figure 7.7. Enter =linest(G14:G31,A14:A31) into E1:F1 as an array function (<Ctrl><Shift><Enter>) to produce the catch curve regression. By pressing F9, the rand() functions are recalculated and a new sample will be generated with its associated gradient. You can observe the impact of varying the sampling error variance (in B7) on the quality of the observed catch curve. If you alter the age at first capture you will need to alter E1:E2 and the plot.

	A	B	C	D	E	F	G
1				Gradient	−0.269	5.8317	
2			Z	=abs(E1)			
3							
4	N0	5000000					
5	Age$_c$	3					
6	Z	0.25					
7	σ_s	15					
8	n	750		Tot Vuln	Average f		
9				=sum(d11:d31)	=average(e11:e31)		
10	Age	Nt		Vuln N	Perfect Sample	+Error	Ln(n)
11	0	=B4		=if(a11>=b5,b11,0)		0.0	0
12	1	=b11*exp(-Z)		=if(a12>=b5,b12,0)		0.0	0
13	2	=b12*exp(-Z)		=if(a13>=b5,b13,0)		0.0	0
14	3	=b13*exp(-Z)		Copy down	167.8	184.1	5.2155
15	4	Copy down		To row 31	130.7	168.5	5.1267
16	5	To row 31		1432524	101.8	93.6	4.5386

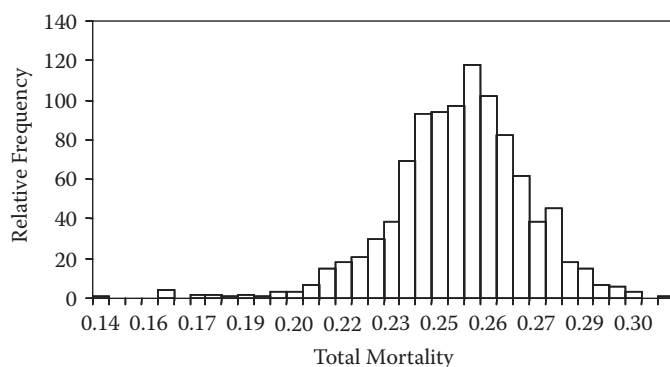© 2011 by Taylor & Francis Group, LLC

Figur e 7.8

Frequency distribution of a simple Monte Carlo simulation of a catch curve analysis. The hypothetical population had constant recruitment and total mortality but was sampled with random error in the estimates of numbers-at-age (Example Box 7.4). The errors had a constant coefficient of variation instead of a constant variance, which allowed for greater variation at high relative frequencies and lower variation at low frequencies. The distribution of Z estimates is slightly skewed but approximately normal. While the median was 0.24997 (the expected was 0.25), the 2.5 and 97.5 percentiles were 0.205 and 0.286, respectively (55.4% below 0.25 and 44.6% above).

EXAMPLE BOX 7.5

A macro that can be added to Example Box 7.4 to automate the execution of multiple repeats of the Monte Carlo simulation. This will generate a column of estimates of Z that can be summarized using the =frequency function. Implement this and attempt to obtain a version of Figure 7.8.

```
Sub montecarlo()
  Dim i As Integer
  ActiveSheet.Calculate
  Application.ScreenUpdating = False
  For i = 1 To 1000
    Range("D2").Select
    Selection.Copy
    ActiveCell.Offset(i, 9).Range("A1").Select
    Selection.PasteSpecial Paste:=xlPasreValues
  Next i                                   ' turned off
  Application.ScreenUpdating = True
  Range("A1").Select
End Sub
```

© 2011 by Taylor & Francis Group, LLC

The catch curve analysis conducted in Example Boxes 7.4 and 7.5 assumed both a constant annual recruitment and total mortality applied to each recruited age class. Sampling error was added in the form of normal random variation with a constant coefficient of variation. By implementing a Monte Carlo simulation of the fishery and the sampling, we were able to investigate the importance of obtaining a representative sample of the relative catch-at-age.

The Monte Carlo simulation in Example Box 7.4 could also be used to investigate the impact of obtaining a biased catch-at-age sample, or of bias in the ageing methodology (ageing error; Lai and Gunderson, 1987), or other sources of error. By investigating which sources of error exert the greatest impact on the final estimates, researchers can identify weak links in their chain of reasoning and focus their research on reducing uncertainty in those areas.

7.4.3 The influence of r ecruitment Variability

The assumption of constant recruitment and equilibrium in the catch curve analysis grates against our intuitions about reality. To investigate the significance of this assumption, we could arrange a Monte Carlo simulation where stochastic recruitment variation was included. In this case, the population model would need to be able to step its cohorts through time with each annual iteration having a randomly selected number of recruits. An efficient method of arranging this would be to use some deterministic stock recruitment relationship, such as the Ricker equation (see Chapter 9), and then adding random error (Equation 7.6):

$$\hat{R}_{t+1} = aS_t e^{-bS_t} e^{N(0,\sigma)} \tag{7.6}$$

where $\hat{R}_{t+1}$ is the expected recruitment at time $t + 1$, S_t is an index of the spawning stock size at time t, and a and b are constants of the Ricker equation (a is the recruits-per-spawner at low stock levels, and b relates to how quickly the level of recruits-per-spawner drops as S increases). With most stock recruitment relationships we should use multiplicative lognormal errors (hence the e^ε). The spawning stock is the sum of all mature-aged animals. In practice, this would entail calculating the deterministic recruitment expected from a given stock size and including the random variation. The impact of total mortality on each age class would be calculated, and then the results copied and stored, ready for the next iteration (Example Box 7.6).

Given such a model, even with no sampling error we would expect to see variation in the relative frequencies of ages in the population. This variation would reflect the nonconstant recruitment to the population. An implication of variable recruitment is that a population is likely to exhibit evidence of strong and weak year classes (Figure 7.9). If this data were used in a catch curve analysis, we would expect to see noise about the hypothetical straight line, even in the absence of sampling error (Figure 7.10, Example Box 7.6).

EXAMPLE BOX 7.6

A nonequilibrium, variable recruitment catch curve analysis. This worksheet is identical to that in Example Box 7.4 except in cells A1:B3, B9, B11:B31, and C12:C31. B11:B31 are now just numbers, and C12:C31 now represent the population in column B suffering mortality. Column D now refers to column C, which represents the latest population structure. Otherwise, the worksheet remains the same. Cells A1:B3 contain the Ricker stock recruitment parameters and the stdev of the recruitment variation. Put =$B$2*D9*exp(-b*D9)*exp(norminv(rand(),0,B3)) into B9 to calculate the expected recruitment. One could have used loginv(rand(),0,B3) to obtain the same result. D9 represents the spawning stock size that is used in the stock recruitment relationship. Try varying both σ_r and σ_s, and determine the effect on the catch curve analysis. Implement a macro to repeat the analysis one thousand times and construct a relative frequency of Z estimates (Example Box 7.7, Figure 7.11). B11 is filled by the macro from B9. B4 is no longer used.

	A	B	C	D
1	Ricker a	6		Gradient
2	Ricker b	1.826E-07	Z	0.2365
3	σ_r	0.3		
4	N0	5000000		
5	Age_c	3	Step_Model	
6	Z	0.25		
7	σ_s	15		
8	N	750		Mature Pop
9	Recruits	6673370		=sum(D11:D31)
10	Age	Nt	Nt+1	Mature N
11	0	8774442		
12	1	5760803	=B11*exp(-$B$6)	=IF(A12>=$B$5,C12,0)
13	2	2726971	=B12*exp(-$B$6)	=IF(A13>=$B$5,C13,0)
14	3	3374618	=B13*exp(-$B$6)	=IF(A14>=$B$5,C14,0)
15	4	2540555	Copy Down	Copy Down
16	5	1520554	to row 31	to row 31
17	6	1293433	1184209	1184209
18	7	878412	1007327	1007327

A Monte Carlo analysis of catch curves with only recruitment variation leads to a symmetric distribution of total mortality estimates, while recruitment and sampling variation leads to a skewed distribution with much wider percentile confidence intervals (although the median remains centrally located about 0.25). The earlier analysis of sampling variation alone (Figure 7.8)

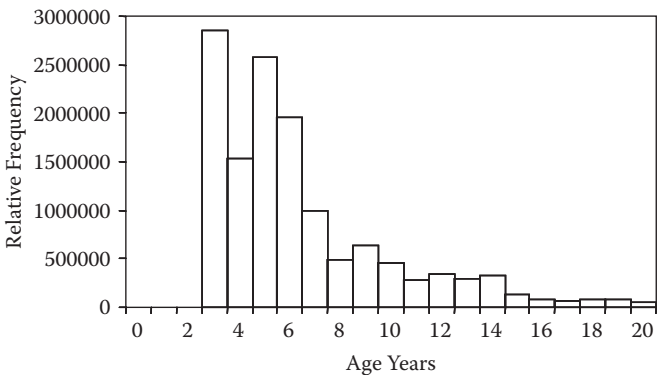© 2011 by Taylor & Francis Group, LLC

Fig ur e 7.9
A nonequilibrium age-structured population model with a constant rate of mortality of $Z = 0.25$ and variable recruitment as in Equation 7.6, with $\sigma = 0.3$. The first three year classes are not selected by the fishery and hence are not sampled. Strong year classes appear evident from three, five, six, and nine years before present (cf. Figure 7.5). If this population were to be sampled, especially if there were sampling error, we would not expect a particularly good fit of a straight line from the log of the relative catch-at-age against age, as required by the standard catch curve analysis (Figure 7.10, Example Box 7.6).
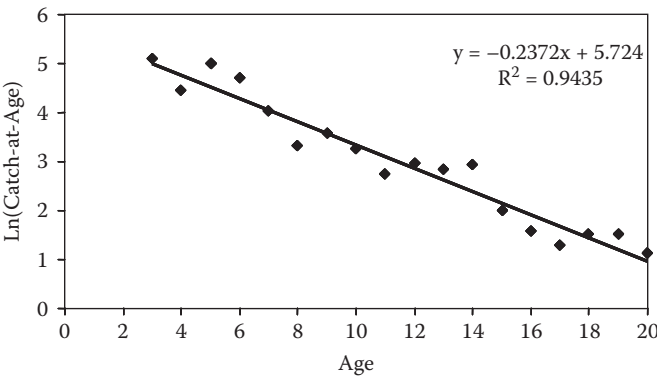

$$y = -0.2372x + 5.724$$
$$R^2 = 0.9435$$

Fig ur e 7.10
Catch curve analysis of the data, given in Figure 7.9, from a nonequilibrium population having a total mortality of 0.25 and random recruitment. There was no sampling error. Note the scatter about the regression line and, in this case, the underestimate of the Z value (Example Boxes 7.5 and 7.6). Repeating this analysis many times would permit a characterization of the effect of recruitment variation on the estimation of Z.

indicated that the distribution of the Z estimates only increased its range in a small way after the addition of recruitment variation. With only sampling error, the 95% confidence intervals were 0.205 and 0.286 (Figure 7.8). With both sampling error and recruitment variation, the confidence intervals around the Z estimate were 0.1995 and 0.292 (Figure 7.11). Recruitment variability does not appear as important as the presence of sampling error on the number-at-age.
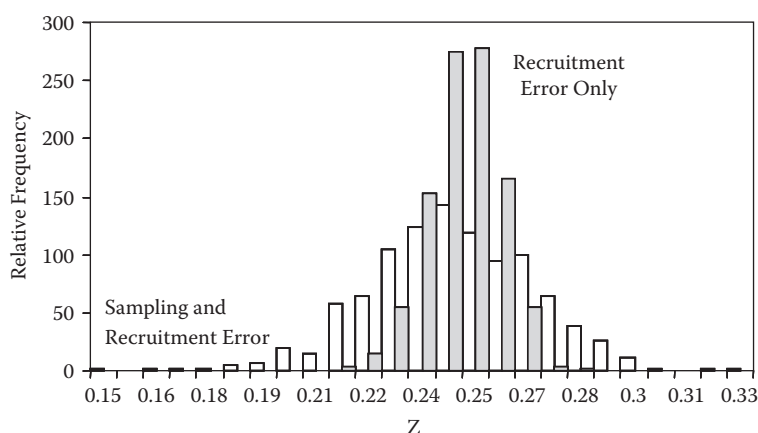
© 2011 by Taylor & Francis Group, LLC

Fig ur e 7.11
The impact of sampling error relative to recruitment variation on the estimation of total mortality using a catch curve analysis (Example Boxes 7.6 and 7.7). The model settings are recruitment variability = 0.3 and sampling variability = 15. The darker columns reflect the catch curve analysis when only recruitment variability is present. The empty bars are where both types of variation are present. The sampling errors dominate the analysis (cf. Figure 7.8).

It would not be valid to draw conclusions after so few investigations. The ranges of variation assumed for the sampling and recruitment error would need to be altered and their interaction more thoroughly determined before any conclusions could be drawn (Example Box 7.7).

7.5 Concluding Remarks

Monte Carlo simulation modelling allows one to investigate many parts of nature that are currently not amenable to direct observation of experiment. Its value in risk assessment and management strategy evaluation is its most important use in fisheries today (Francis, 1992; Smith et al., 2007). Risk assessments involve projecting the stock dynamics forward under different simulated management regimes, and of course, this is beyond experience.

Many skills are needed to conduct such simulation modelling, including knowledge of probability density functions and which to use in what circumstances, as well as a facility with computer programming, in whatever language suits you best. Most important, however, is a need to have an understanding of the present state of knowledge about the system being modelled. Not all simulation models need be realistic, and for those situations a simple mathematical view of the system would be all that was required. However, for a simulation model to be realistic, any hypothetical additions would

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 7.7

A macro and Monte Carlo procedure to add to Example Box 7.6 to automate the simulations and enable percentile analyses of the results. Set the sampling error to 0.001 and the recruitment error to 0.3, and run the Monte Carlo analysis, generate a histogram of the results using the =frequency function, and store the actual numbers away from column K. Then set the sampling error to 15 and rerun the Monte Carlo analysis. Add the Z estimates from this to the histogram already generated to produce something akin to Figure 7.11.

```
Sub Step_Model()
' Conducts a single population growth cycle
  Range("C12:C31").Copy
  Range("B12").Select
  Selection.PasteSpecial Paste:=xlValues
  Range("B9").Copy
  Range("B11").Select
  Selection.PasteSpecial Paste:=xlPasteValues
  Application.CutCopyMode = False
  Range("A1").Select
End Sub

Sub Do_Monte_Carlo()
' automates simulations for Monte Carlo analyses
Dim i As Integer
For i = 1 To 1000
  Application.ScreenUpdating = False
  Step_Model                        ' call the other macro
  Range("D2").Copy
  ActiveCell.Offset(i, 10).Range("A1").Select
  Selection.PasteSpecial Paste:=xlPasteValues ' column K
Next i
  Range("A1").Select
  Application.ScreenUpdating = True
End Sub
```

---

need careful construction to ensure biological plausibility. Even if a model has been inappropriately constructed it will still generate the implications of particular actions. Generally, it is better to investigate a range of model structures than rely on a single formulation.

Whenever a simulation model is used and it contains parts that are strictly invention it must not be mistaken for reality; it is only a model. When conducting "what if" trials, we would be testing what would be the

case if the system operated in a particular way. It is important not to forget that even if the results reflected observable nature closely, this would not imply that the particular model being used is a true representation of the operation of nature. As always, while a good fit certainly describes, it does not necessarily explain.

© 2011 by Taylor & Francis Group, LLC