4

Computer-Intensive Methods

4.1 Introduction

"Computer-intensive methods" is a phrase used to refer to a number of methods that include randomization tests (= permutation tests), jackknife techniques, bootstrap techniques, and Monte Carlo simulations and tests. In this chapter we will briefly introduce all four methods along with some of their uses. Fuller descriptions and examples will be presented in the following chapters.

Computer-intensive can sound daunting, but only if it is confused with meaning intensive computer programming (though some facility with writing relatively simple macros or computer programs is necessary). The word *intensive* refers to the use of available computing power. Such statistics and methods are not new. Randomization tests were discussed by Fisher (1936), but have recently become much more widespread and popular as a direct result of the easy availability of powerful computers (Good, 1994; Manly, 1991, 1997).

The history of statistics is relatively short and has a number of phases. The first developments in statistics of importance to our work here relate to parametric statistics. These statistics are dependent upon known probability distributions (such as the normal or binomial distributions, as in Chapter 3), and when they are used they require that the populations being sampled adhere to assumptions peculiar to the distributions concerned. Parametric statistics form the core of "traditional" statistics (analysis of variance (ANOVA), regression, sample comparison, experimental design, etc.).

The second phase of particular interest to us was the production of nonparametric statistics. Ronald A. Fisher (1936) first mooted the idea of testing hypotheses concerning observed patterns between groups of individuals by using a randomization test. These tests compare an observed pattern among groups with those possible by randomly allocating the available data across the groups of interest. Of course, with any reasonable amount of data such an approach was simply not practical at a time (1930s) when all calculations had to be done by hand or on a mechanical hand calculator. Nevertheless, as a theoretical insight it eventually led to the development of a wide range of nonparametric test statistics. Almost all common nonparametric statistics

Copyright © 2011. CRC Press LLC. All rights reserved.

© 2011 by Taylor & Francis Group, LLC

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

have their basis in replacing the observed test data with ranks and computing test statistics based upon permutation tests (Good, 1994). Using ranks removes the unique or idiosyncratic aspects of each situation so that the same probability distribution of possible arrangements can be applied to every hypothesis test. The term *nonparametric* implies that the parametric form of the underlying population distribution need not be specified exactly. In other words, the data do not have to match the assumptions behind any particular probability density function.

Most recently there has been a growth of interest in general randomization tests, plus the development of the newer methods of jackknife and bootstrap statistics. All of these require far more calculations than classical methods, but with increasing computing power, these approaches may supersede older methods that depend on parametric statistical distributions. They will certainly constitute an important part of the future of statistical analysis, especially of biological data, which often is nonnormal.

4.2 Resampling

The notion of resampling a population is fundamental to many statistics. If a set of data is normally distributed, then given its mean and standard deviation, we can estimate the standard error of the mean analytically:

$$\text{StErr} = \frac{\sigma}{\sqrt{n}} \tag{4.1}$$

where n is the number of observations and σ is the standard deviation.

An alternative to this standard analytical method would be to obtain a number of independent samples from the original population (i.e., resample the population a number of times). The standard deviation of the multiple sample means would also provide an estimate of the standard error. Thus, the notion of resampling should not be considered remarkable. This is fortunate because resampling in one form or another is the foundation of all the computer-intensive methods to be discussed. While resampling is not unusual, what is resampled, and how it is resampled, differs between the methods.

In the following sections and examples we will be considering a range of available computer-intensive methods. Before this, we need to introduce a distinction concerning the approach used when making observations (sampling) that will be important in what follows. Sampling without replacement means making an observation from a population by removing individuals. Further observations from a sample could obviously not contain those particular individuals again. This contrasts with sampling with replacement, where an observation is made nondestructively, i.e., without removing individuals.

Copyright © 2011. CRC Press LLC. All rights reserved.

© 2011 by Taylor & Francis Group, LLC

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

Clearly, when sampling with replacement, subsequent observations could be repeats of observations already made.

4.3 Randomization Tests

Randomization tests are used to test the null hypothesis that an observed pattern is typical of a random event. The null hypothesis suggests that each group of observations merely represents a random sample from a single population. The observed pattern among groups must be characterized by a test statistic (e.g., a mean difference between two groups). To test whether the observed pattern is significantly different from a random pattern, the data from all groups are first combined. The observed test statistic is then compared with that obtained by randomly reallocating individuals from the combined data back to the groups to be compared (i.e., resampling without replacement). Of course, it is of no value to do this only once. The random resampling without replacement into the groups and recalculating the test statistic must be repeated many times (1,000+ being typical). The observed test statistic for the original arrangement is compared with the empirical distribution of that test statistic given the data available. If the pattern observed is not different from random, then the observed test statistic value will be typical of those generated by random groupings. A significant difference is indicated by the observed test statistic lying beyond or at the extremes of the empirical distribution obtained from randomizing the data among the groups.

Permutation tests are good for testing hypotheses, but standard errors cannot be calculated using this approach, and confidence intervals on parameter estimates can only be fitted very inefficiently (Curnow, 1984).

Fisher (1936) first provided a theoretical description of randomization tests (Pitman, 1937a, 1937b). Reviews of randomization techniques, with contrasting views on some methods, are given by Manly (1997) and Edgington (1995). There is a large and expanding volume of literature on randomization tests providing both examples and theoretical developments (Lindley and Novick, 1981). Some controversy has occurred, but this has not prevented continued development (Basu, 1980, plus associated comments; e.g., Hinkley, 1980; Kempthorne, 1980; Lane, 1980; Lindley, 1980). A detailed review and bibliography relating to randomization tests is given by Good (1994).

4.4 Jackknife Methods

The name is reported as coming from considering this statistical method as a flexible tool rather like a multifunction pocketknife. By considering known

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

subsets of the available data, one can produce estimates of standard errors as well as detect whether any parameter estimates are biased. Subsetting the data involves estimating the statistics of interest on all possible combinations of the available data minus one data point: in a data set of n values there will be n subsets of $(n - 1)$ data points, and these are used to calculate the jackknife replicates. Using the jackknife replicates, one can calculate what are known as pseudovalues for the statistic of interest. The difference between the original sample mean and the mean of the n pseudovalues provides the estimate of bias. The value of this methodology comes when one is not estimating the sample mean (which is an unbiased estimate) but some other parameter. The pseudovalues can also be used to calculate jackknife estimates of the parameter of interest and its standard error. Confidence intervals can be fitted using this standard error estimate, but there is a problem deciding how many degrees of freedom to use, so this approach is no longer recommended.

The jackknife methodology was first discussed by Quenouille (1956), who recommended the approach as a method for removing bias from parameter estimates. Tukey gave a paper to a conference of the American Institute of Statistics at Ames in Iowa. He introduced the notion of using the jackknife approach to produce parameter estimates with estimates of standard errors. Only the abstract of the conference talk was printed (Tukey, 1958), but that and his talk were enough to set off a number of developments (Hinkley, 1983). Jackknifing is discussed in Chapter 6.

4.5 Bootstrapping Methods

Data sampled from a population are treated as being (assumed to be) representative of that population and the underlying probability density distribution of expected sample values. Given an original sample of n observations, bootstrap samples would be random samples of n observations taken from the original sample with replacement. Bootstrap samples (i.e., random resampling from the sample data values with replacement) are assumed to approximate the distribution of values that would have arisen from repeatedly sampling the original sampled population. Each of these bootstrapped samples is treated as an independent random sample from the original population. This approach appears counterintuitive to some, but can be used to fit standard errors, confidence intervals, and to test hypotheses. The name *bootstrap* is reported to derive from the story *The Adventures of Baron Munchausen*, in which the baron escaped drowning by picking himself up by his own bootstraps and thereby escaping from a well (Efron and Tibshirani, 1993).

Efron (1979) first suggested bootstrapping as a practical procedure. He states (Efron and LePage, 1992; Lepage and Billard, 1992) that development of the bootstrap began as an attempt to better understand the jackknife but

Copyright © 2011. CRC Press LLC. All rights reserved.

© 2011 by Taylor & Francis Group, LLC

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

quickly developed beyond the potential of the jackknife. The bootstrap could be applied to problems beyond those of estimating bias and standard errors; in particular, it was an approach that could provide better confidence intervals than those from the jackknife. Bickel and Freedman (1981) provided a demonstration of the asymptotic consistency of the bootstrap (convergent behaviour as the number of bootstrap samples increased). Given this demonstration, the bootstrap approach has been applied to numerous standard applications, such as multiple regression (Freedman, 1981; ter Braak, 1992) and stratified sampling (Bickel and Freedman, 1984, who found a limitation). Efron eventually converted the material he had been teaching to senior-level students at Stanford into a general summary of progress to date (Efron and Tibshirani, 1993).

4.6 Monte Carlo Methods

Monte Carlo simulations are carried out with a mathematical model of a situation plus a series of model parameters. Some of the parameters will not be known perfectly (i.e., there is uncertainty), so that instead of a particular value, one would have a probability density distribution from which the parameter values are derived (e.g., normal, lognormal, hypergeometric, etc.). Each run of a Monte Carlo simulation involves randomly selecting a value for the variable, parameter, or data values, from the known distribution(s), and then determining the model's output. This process can be repeated many times to test hypotheses or determine confidence intervals. Such resampling from a theoretical distribution of values is effectively sampling with replacement (one could, in theory, obtain the same value more than once). The probability density distribution can never be exhausted of values.

Monte Carlo testing is often about comparing what was actually observed in a system with what one obtains from a model of the system. It involves an assessment of the properties of the system. In a hypothesis testing situation, if any of the hypotheses included in the model are incorrect, then the model output would not be expected to be consistent with the available observations.

Monte Carlo simulations are also the basis of risk assessment methods in fisheries by projecting the expected path of a fishery when it is exposed to a particular harvest strategy (e.g., a constant fishing mortality rate or constant catch level). When more than one of a model's parameters are each free to vary over a range of values, then the model output also becomes variable. If the model is run enough times, one would expect to be able to generate a frequency distribution of possible outcomes (perhaps the biomass remaining in a stock after some years of exploitation at a given total allowable catch (TAC)). From this distribution one could derive the likelihood of various outcomes (e.g., stock collapse—defined in a particular way, or current biomass falling below defined levels). In New Zealand, for example, such a model

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

was used to determine the impact on orange roughy (*Haplostethus atlanticus*) stocks of different projected catch rates over the next twenty years (Francis, 1992). Thereby, the option of reducing the commercial catch slowly was demonstrated to be more risky for the stock than a rapid decline in catch levels.

4.7 Bayesian Methods

It is becoming common for many fisheries' stock assessments to be conducted using a Bayesian framework. This cannot be well characterized using only a few sentences and will be introduced more completely later. Having said that, Bayesian analyses can be thought of as attempting to describe the uncertainty around model estimates using a multidimensional likelihood profile. Bayesian methods attempt to combine maximum likelihood methods in a formal manner with prior information to produce the final posterior probability distributions that describe the relative probability of different outcomes. This involves combining the probability density function selected for the maximum likelihood analysis of the data used in the stock assessment with the probability density function used to describe the prior probability distribution of the various parameters being estimated. Some combinations of likelihood distributions and prior probabilities (so-called conjugate priors; Gelman et al., 2004) permit an analytical solution for the Bayesian analysis; however, most fisheries' assessments are sufficiently complex that no such simple solution is possible (Walters and Ludwig, 1994). In most fisheries' assessment situations there are so many parameters, each with a prior probability distribution that no simple solution exists. Under these more usual circumstances the analysis becomes a multidimensional integration problem and alternative computer-intensive approaches can be used. Sampling importance resampling (SIR) and Markov chain Monte Carlo (MCMC) are different Monte Carlo methods commonly used to combine the many probability distributions relating to each of the model parameters. They are not in themselves Bayesian methods; they are merely different algorithms used to conduct the multidimensional integration needed to find the posterior probability distributions produced by the Bayesian analysis. An example of how to run a MCMC is given in Chapter 8.

4.8 Relationships between Methods

All the computer-intensive methods we are going to consider may be viewed as different forms of random resampling where the observed sample data or

Copyright © 2011. CRC Press LLC. All rights reserved.

© 2011 by Taylor & Francis Group, LLC

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

TABLe 4.1

Relationships between Computer-Intensive Methods and Their Strategies for Resampling from Probability Distributions

Method of Resampling	Computer-Intensive Method
Resampling a theoretical PDF (e.g., *t* distribution, χ^2 distribution) Implicitly this is sampling with replacement	Parametric statistics (analytically) Parametric Monte Carlo simulations Includes Bayesian methods
Resampling an empirical distribution (as represented by a sample) with replacement	Nonparametric bootstrap Nonparametric Monte Carlo Includes Bayesian methods
Resampling an empirical distribution (as represented by a sample) without replacement	Randomization tests Jackknife statistics

Note: PDF refers to probability density function. Parametric statistics are included for completeness but are not usually considered computer intensive. The nonparametric Monte Carlo and bootstrap methods are equivalent.

its properties are taken to represent the expected range of possible data from the sampled population. So, instead of sampling from a theoretical probability distribution, one can resample from the empirical distribution that is represented by the values in the sample (Table 4.1). Alternatively, one can sample from a parametric statistical distribution whose parameters are estimated from the original sample.

Randomization tests can be considered to be a special case of Monte Carlo testing where the original sample data are resampled without replacement so that each run uses all the available data. The only thing that changes is the assortment of data between groups.

Jackknife analysis is also a special case of Monte Carlo sampling where the available data forms the empirical distribution sampled. In this case it is systematically subsampled without replacement. It is the fact that values are omitted systematically that leads to there being a fixed number ($n - 1$) of jackknife replicates.

Nonparametric bootstrapping is another special case of the Monte Carlo process where the observed sample takes the place of a parametric probability distribution, or even the original population. In this case the situation is much more akin to parametric Monte Carlo sampling. The observed sample is sampled repeatedly, with replacement, just as if it were a continuous probability distribution.

Resampling from either theoretical probability density distributions or empirically derived distributions forms the basis behind the computer-intensive methods used to integrate the multidimensional problems that arise in Bayesian stock assessments.

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.

4.9 Computer Programming

We will implement all of these computer-intensive methods in Excel workbooks using surprisingly little macro coding. Excel macros, however, can often be too slow for serious analyses when they are doing a great deal of work (the whole point of computer-intensive statistics). Ideally, in those cases, we might wish to write an executable program in some programming language such as Pascal, C++, or Fortran (even R, an interpreted language, is far more efficient for these purposes than Excel). Manly (1991) and Edgington (1987) provide program code for subroutines in Fortran for randomization tests, while Efron and Tibshirani (1993) provide the necessary code in the S statistical package for carrying out bootstrapping routines (also available in R). Many statistical packages now include bootstrapping as an option for many of their statistical routines. In addition, many statistical packages can now be macro-programmed into conducting computer-intensive statistics. For our purposes in this book, the Excel spreadsheet and its Visual Basic macro language will suffice.

Copyright © 2011. CRC Press LLC. All rights reserved.

© 2011 by Taylor & Francis Group, LLC

Haddon, Malcolm. Modelling and Quantitative Methods in Fisheries, CRC Press LLC, 2011. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/ucsc/detail.action?docID=1648249.
Created from ucsc on 2021-05-04 11:08:45.