The Role of Simulation Approaches in Statistics

Michael Wood
University of Portsmouth, U.K.

*Journal of Statistics Education* Volume 13, Number 3 (2005),
ww2.amstat.org/publications/jse/v13n3/wood.html

Copyright © 2005 by Michael Wood, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

**Key Words:** Active learning; Approaches to statistical thinking; Bootstrapped confidence intervals; Computer simulation; Probability distributions; Resampling.

**Abstract**

This article explores the uses of a simulation model (the two bucket story)—implemented by a stand-alone computer program, or an Excel workbook (both on the web)—that can be used for deriving bootstrap confidence intervals, and simulating various probability distributions. The strengths of the model are its generality, the fact that it provides a powerful approach that can be fully understood with very little technical background, and the fact that it encourages an active approach to statistics—the user can see the method being acted out either physically, or in imagination, or by a computer. The article argues that this model and other similar models provide an alternative to conventional approaches to deriving probabilities and making statistical inferences. These simulation approaches have a number of advantages compared with conventional approaches: their generality and robustness; the amount of technical background knowledge is much reduced; and, because the methods are essentially sequences of physical actions, it is likely to be easier to understand their interpretation and limitations.

**1. Introduction**

The motivation for this article stems from a program I wrote a few years ago for deriving bootstrap confidence intervals for a mean. Having written the program, it was trivial to adapt it to deal with statistics other than the mean (e.g. median, percentiles, etc), and by coding the presence of a characteristic as 1, and its absence by 0, the program could deal with proportions and categorical data. I then realised the program could do many other things as well: for example, it could simulate the binomial, Poisson, hypergeometric and normal distributions. I then produced a spreadsheet version of the original program, which led to another expansion of the scope of the program without any essential change in the underlying idea—which I will call "the two bucket story" for reasons that are explained below.

The underlying simulation model is obviously very general. It provides the user with a very powerful
tool. Furthermore, it is a tool specified not by mathematical formulae but by a physical process, albeit one performed by a computer. This all has clear implications for both pedagogy and the nature of the statistical knowledge we expect students to learn. The aim of this article is to explore these implications.

There are, of course, many other simulation models that are useful in statistics. Simulation is widely recognised as a useful tool for teaching statistics (Mills 2002; Simon, Atkinson, and Shevokas 1976). Usually it is seen as an aid to learning standard methods. It can also, however, be seen as an alternative to some of these standard methods: if the student has a well-understood and practical simulation method available, is it necessary to learn the conventional method as well? I will start by describing the two bucket story and its applications, and then use this as an example of a simulation model to address these more general issues.

2. The two bucket story and its scope

The two bucket story involves the use of a model with two buckets containing balls. The idea of drawing balls from urns as a metaphor for probability is widely used (e.g. Lindley 1985) because it is easy to visualise what probabilities mean and how they can be calculated in this context (e.g. if 40% of the balls are black, the probability of drawing a black ball is 40%). I have replaced the urns by buckets because these days an “urn” (as used in the UK) typically contains tea, coffee or the ashes of a deceased person—none of which are helpful associations for my purposes here. I have also used the term story to emphasise the fact that the model is a sequence of actions, not a static collection of objects. The word "story" also seems useful here because it may communicate the idea that statistical models never represent the whole truth: they are stories which help us to understand reality but can never provide a perfect match.

I will start by describing the story in fairly abstract terms, and then describe some specific applications. In a teaching context, it would probably be better to reverse this order.

The first bucket, Bucket 1, holds \( b_1 \) balls representing some collection (a sample, a set of possible outcomes, or something else). Each ball has some information stamped on it representing a member of this collection. For example, the collection might be a sample of people and the information stamped on the ball representing each person might be the number of cars owned by the person in question. A random “draw” of \( n \) balls is now taken from Bucket 1, and a statistic (e.g. the mean) is calculated from the information on these \( n \) balls and the answer stamped on a ball that is then put in Bucket 2. This can be done in two ways: either with replacement (replacing each ball after drawing it), or without replacement (which means, of course, that \( n \) must be less than or equal to \( b_1 \) or we will run out of balls).

This is repeated \( b_2 \) times so that there are \( b_2 \) balls in Bucket 2. The contents of Bucket 2 can then be used to analyse the simulated distribution of the statistic (derive probabilities, percentiles, etc). The italicised words all represent variables in the sense that they can be varied between different applications of the model.

The programs I used to implement the two bucket story are Resample.exe and Resample.xls. The first of these can analyse the following statistics: mean, sum, standard deviation, variance, percentiles (including maximum and minimum), median, range, inter-quartile range. The second, a spreadsheet, allows the analysis of bivariate distributions, and so enables the analysis of statistics such as correlation and regression coefficients, and other functions of two variables.

2.1 Bootstrap confidence intervals
Let’s say we have a random sample of 12 measurements from a population of students on a course representing the number of cars owned by each person (3, 2, 0, 0, 1, 1, 5, 1, 1, 0, 4) and wish to derive a confidence interval for the mean from the population. The mean of this sample is 1.5, and we want to know how accurate an estimate of the population mean this is likely to be. The basic idea of bootstrapping is to form a guess of what the population is like, and then to run some sampling experiments on this “guessed population” to gauge the likely extent of sampling error.

Suppose, first, that the population is finite — say of size 48. The first step is to form a suitable guessed population by taking four copies of the sample. We don't know what the real population is like, but this seems a reasonable guess on the basis of the only information we have — the sample. This guessed population is represented by Bucket 1 — which contains 48 balls representing the 48 members of this guessed population. (Bucket 1 contains, for example, 16 balls labelled 0, since there are four 0s in the original sample, and we are taking four copies of this.) Now we take random draws — “resamples” without replacement — of 12 balls from this guessed population in Bucket 1, find the mean of each resample, and stamp these means on separate balls — which we put in Bucket 2. Drawing the balls without replacement means simply that each ball is drawn and not replaced, which is, of course, equivalent to drawing a whole sample of 12 balls.

This models, very directly, the process of sampling from the real population, except, of course, that we only have the guessed population. When I did this, the first resample comprised 5, 0, 1, 0, 3, 5, 0, 1, 1, 0, 0, 2, which has a mean of 1.58. (Note that although all these numbers occur in the original sample, the resampling process means that the frequency with which they occur is not the same.) Repeating this whole process 10000 times gave a total of 10000 resample means in Bucket 2. Ninety five percent of these were between 0.75 and 2.33 (the 2.5 and 97.5 percentiles).

We can then use these results from Bucket 2 to estimate how much means of samples of 12 are likely to differ from the mean of the population from which they are drawn. In this case, as the guessed population is simply four copies of the sample, its mean is the same as the mean of the sample — 1.5. The simulation results showed that the 2.5 and 97.5 percentiles of the distribution of the resample means were 0.75 and 2.33. The lower of these figures (0.75) corresponds to a resample mean that is 0.75 lower than the true mean of the guessed population (1.5), and the upper figure (2.33) is 0.83 above the true mean of the guessed population. These results suggest that there is a 95% probability that the errors in the means of these samples of 12 are less than about 0.8 — the qualifier “about” being necessary because of the slight difference between the errors above and below the true mean. This, in turn, suggests that we can be 95% confident that the true mean of the real population is in the range 1.5 (the mean of the original sample of data) plus or minus about 0.8 — i.e. 0.7 to 2.3. This is roughly the same as the percentiles of the resample means (in Bucket 2), so we can use these percentiles to define a confidence interval for the mean. This is called the bootstrap percentile interval for obvious reasons. The method, and the terminology, can easily be extended to populations of different sizes. Strictly, population sizes that are not integral multiples of the sample size create a problem, but in practice, taking the nearest integral multiple is likely to be good enough.

Infinite populations can be modelled by letting Bucket 1 represent the sample (\( b \) is the sample size), and then assuming that the distribution of the guessed population is identical to that of the sample. This means, for example, that as the value 1 occurs in 25% of the sample measurements, the same will be true of the guessed population. We then draw our sample of 12 balls from Bucket 1, replacing each ball after drawing it so that the distribution remains unchanged for when the next member of the sample is drawn — as it would in an infinite population. In this case the results are similar — the 95% interval derived from the results in Bucket 2 extends from 0.67 to 2.50. This interval is slightly wider than the interval based on the population of 48 — for reasons that should be intuitively obvious if you follow through the...
two processes. (In the extreme, if the population and the sample are the same size, the width of the interval will be zero because there is only one sample that can be drawn.) It is also slightly more asymmetrical. Bootstrapping enables us to model the distinction between finite and infinite populations in a straightforward and transparent manner, whereas conventional methods are generally restricted to infinite populations.

The idea of a guessed population plays a crucial role here. There are other ways of conceptualising the process: Simon (1992) refers to a “pseudo-universe,” and Efron and Tibshirani (1993, p. 87) and Lunneborg (2000) write of the population distribution in the “bootstrap world.” It would be possible to describe the resampling process in purely mechanical terms, but it is important to tell the story in terms of the “guessed population,” or something similar, to clarify the rationale behind the process in intuitive terms, and to assess the assumptions on which the method’s validity depends.

The argument above, in terms of guessed populations, does make a number of assumptions which may not be exactly satisfied in practice — e.g. the sample (or a number of copies of the sample) can be used to form a surrogate population which will give an accurate idea of sampling error, and the extent of sampling error is roughly the same in both directions (see Wood 2003 for more detail of the assumptions implicit in this argument). One of the strengths of this bootstrap method is that it has a relatively simple rationale, so problems and assumptions are relatively clear.

The same method can be used for any other statistic which is calculated from a random sample — e.g. a median, proportion exhibiting some characteristic, various correlation coefficients, regression coefficients, etc. The only difference is the statistic we calculate from each draw from Bucket 1.

There are, of course, more sophisticated bootstrap methods, which may be useful when the assumptions on which the percentile interval is based are unreasonable. For example the sample of 12 above is unlikely to give an accurate idea of rare extreme values — e.g. some people doubtless have 10 cars. If we were interested in these, it may make sense to use the sample to fit a suitable probability distribution, and then use this to generate a guessed population for Bucket 1. More elaborate methods of bootstrapping are discussed in the technical literature on bootstrapping (e.g. Davison and Hinkley 1997; Efron and Tibshirani 1993; Good 2001; Lunneborg 2000).

There are also non-technical explanations of bootstrapping aimed at general readers and beginners (Diaconis and Efron 1983; Gunter 1991, 1992a, 1992b; Simon 1992; Wood 2003), and a few articles on the use of bootstrapping and similar methods for teachers (Braun 1995; Butler, Rothery, and Roy 2003; Duckworth and Stephenson 2002, Ricketts and Berry 1994; Simon, et al. 1976). The Resampling Stats website at resample.com also has links to a range of articles and books on bootstrapping and related ideas.

### 2.2 The binomial, hypergeometric, Poisson and normal distributions

The two bucket model can also be used to simulate the binomial distribution. For example, to simulate the number of heads in 20 tosses of a coin, put two balls in Bucket 1 labelled 1 and 0 representing the possible numbers of heads we get when we toss a single coin. Each draw then involves drawing a ball from this bucket 20 (n) times (replacing the ball after it is drawn), and then finding the sum – which represents the number of heads – and pasting this on a ball in Bucket 2. A large number of such draws will then enable the binomial probabilities for each possible number of heads to be estimated from the contents of Bucket 2. If the probability of success on each trial, \( p \), was a more awkward number, say 0.034, then the contents of Bucket 1 could be designed to give this probability — e.g. 34 ones and 966 zeros, or 17 ones and 483 zeros.
Other possibilities are to simulate distributions that approximate to the Poisson distribution (by taking $p$ small and $n$ large in the binomial simulation), and the normal distribution (by taking $n$ large in the binomial simulation), as described in Wood (2003). And the fact that the distributions in Bucket 2 are frequently normal provides a convincing illustration of the central limit theorem and the ubiquity of the normal distribution.

The model can also deal with the hypergeometric distribution: for example, it can be used to simulate probabilities in the UK National Lottery (Wood 2003 and Example 1 in the Read this sheet in Resample.xls) by coding the six balls selected by a player as 1, and the 43 not selected as 0. Bucket 1 then contains six balls labelled 1 and 43 labelled 0. The lottery can then be simulated by drawing six balls without replacement, and the sum of the numbers (usually many 0’s and occasionally some 1’s) on the six balls drawn represents the number of numbers correctly forecast. Bucket 2 then represents the scores from each lottery ticket, and can be used to estimate the probabilities of the various prizes (although jackpot winners are so rare that a very large value of b2 is required—beyond the capacity of the two programs mentioned above).

2.3 Other possibilities

There are further possible uses. The “birthday” problem of calculating the probability of two people in a room sharing a birthday can be tackled with the addition of a statistic to count the number of different values in a list of numbers (Wood 2003). A very similar method can be used to model the distribution of the number of different DNA types in a sample (the response to a query I received from a geneticist). And in another very different context, the control lines in Shewhart control charts (as used in business quality management) can be derived using a resampling method (Wood, Kaye, and Capon 1999) that can be modelled using the two bucket model.

2.4 Why buckets and balls?

The original program I wrote (Resample.exe) was designed for bootstrapping and uses the language of samples and resampling. However, the terminology of resampling fits awkwardly with a program for simulating the binomial distribution. The “sample” is not really a sample in the ordinary sense of the term, and so the “resample” also cannot be a proper resample. The model needed some more neutral terminology that avoids slotting it into just the one pigeonhole. The two bucket story is such neutral terminology. Bucket 1 may represent a sample, a population or a process, and the balls in it may represent people or possible outcomes of tossing a coin or many other things. The general terminology keeps our options open, and hopefully reduces the danger of potential users interpreting the model too narrowly.

This generality means that the two bucket story is potentially very powerful. Instead of having separate models for deriving confidence intervals for different statistics, and other models for the binomial, Poisson, normal and hypergeometric distributions, and for the birthday problem and for control lines in quality control charts, they are all brought together under one umbrella.

3. Simulation approaches in pedagogy and academia

Simulation approaches offer great opportunities for working out probabilities, confidence intervals and similar concepts. We have looked at the example of the two bucket story: this can be used to estimate a wide variety of probability distributions and confidence intervals. All this can be achieved without any mathematical symbolism: in its place is a physical story and the facility to carry the process out with a computer. Instead of the mathematical theory of confidence intervals, and of the binomial, Poisson,
normal and hypergeometric distributions, we just have a simple physical story.

However, a physical story can be used without any appreciation of its meaning in just the same way that a symbolic argument or formula can be used blindly. For the rationale behind the story to be appreciated so that it can be used intelligently and adapted to new situations, it is necessary for users to think about what is going on, and here suitable terminology is likely to be helpful. In this article I have suggested the phrase “two bucket story” as a general label to avoid prejudging the interpretation of its components. In the particular application to the derivation of confidence intervals, ideas such as a “guessed population” help in the interpretation of the contents of the first bucket, and in clarifying the rationale behind the bootstrapping process. These terms should be taken as suggestions: like any language, the use of words is likely to evolve as they are used in different contexts.

There are, of course, many useful simulation approaches that do not fall under the umbrella of the two bucket story. One example is provided by approximate randomization tests (Noreen 1989; Wood 2003). These are randomization tests (Edgington 1995) which assess significance by “shuffling one variable … relative to another …” (Noreen 1989, page 9). This is a general simulation method that can often be used as a substitute for a number of traditional hypothesis tests — t test, one way analysis of variance, test of the hypothesis that a correlation is zero, Mann-Whitney test, etc. However, it does not fit the format of the two bucket story—it would need a third bucket so that two buckets can be reserved for the data allowing them to be “shuffled” relative to each other. There is a spreadsheet implementation of this shuffling principle at Resamplenrh.xls

The rationale behind many simulation approaches is simple enough to be understood by users without an extensive background in statistics. This means that a “relational” understanding (Skemp 1976) of why the method works, as well as an “instrumental” understanding of how to do it, is a reasonable expectation for most students. This represents a substantial cultural shift. When deriving a conventional confidence interval for the mean, for example, non-mathematical students would not normally be expected to understand where the values of t or z come from: the explanation may be that they are found in tables, or that mathematicians, or computers, have calculated them. These values need to be taken on trust. This is not true of the bootstrap percentile interval. Here, it is possible for the non-mathematical student to follow the whole rationale: there are no gaps to be filled by the mysterious activities of tables, computers or mathematicians.

To put this in different, but more or less equivalent, terms, beginning students are much more likely to be able to take an “active” or a “constructivist” approach with simulation methods: the whole method becomes a story which the student can run through and make sense of. In the words of a participant in one study “… the resampling method makes one feel that we are physically doing it, or actually seeing it being physically done, without having to take any theoretical mathematics into consideration” (Ricketts and Berry 1994, page 43). Simon, et al. (1976) go even further:

“The Monte Carlo method is not explained by the instructor. Rather it is discovered by the students. With a bit of guidance the students invent, from scratch, the procedures for solution” (p. 734)

This is in strong contrast to many formula-based methods, where the story behind the formula or the tables may be too long and complicated for students to “construct” in their minds, and certainly too complicated for them to “discover” for themselves. The virtues of “active learning” are generally accepted by educationalists: see, for example, the review of constructivism in Mills (2002) and the British Higher Education Funding Councils’ journal entitled simply Active learning. Simulation approaches, such as the two bucket story, must be in line with this ethos. They provide a way of seeing a statistical analysis as a physical story, instead of an abstract mathematical model.
The generality of the two bucket story is another important advantage over traditional approaches. Instead of learning a method, and associated formulae, for a confidence interval for a mean, and another for a confidence interval for a proportion, and the theory of the normal, binomial and hypergeometric distributions, there is a just a single method which applies across the board. Furthermore, this single method will cope with problems where there are no well-known methods. This makes the learner’s task far simpler, and gives the successful learner a tool that is far more powerful than a collection of formulae.

These simulation approaches offer the possibility that relatively inexperienced users will be able see the importance of the assumptions on which conclusions are based (random samples, for example), and may be able to adapt methods to new circumstances. The days when statistics was a collection of hazily understood, and often misused recipes, may be replaced by an age when people have a collection of general approaches—like the two bucket story—which they can use, actively and intelligently, to devise ways of tackling problems of current concern.

Unfortunately, there are very few empirical studies of actual benefits. One such study (Simon, et al. 1976) reported the results of “three controlled experimental tests of the pedagogical efficiency of the Monte Carlo [simulation] method.” Most of the results favoured the simulation approach, although there were difficulties in comparing the simulation approach with conventional approaches. They claim that they are not suggesting the simulation approach as an alternative to conventional approaches, but as an “underpinning”. However, this study was performed in the 1970s in the early days of computer technology, so the practical advantage of the simulation approach would have been smaller. More recently, Ricketts and Berry (1994) looked at the experiences of a class using resampling approaches instead of mathematical theory. Unfortunately, the results leave a little to be desired from the statistical point of view, being confined to two enthusiastic comments from students, and a comment from the authors that “our experience suggests that it [the resampling approach] is highly acceptable to students with a range of mathematical abilities.” (p. 44) The ideal would obviously be another controlled trial like the experiments by Simon, et al. (1976) but with modern computing facilities: however these are difficult to organise, particularly as the aims of the new approach might have a different emphasis from the aims of the old approach, with the consequent problems of defining a suitable measure of performance.

From the wider perspective of the academic development of statistics, the difficulty with conventional approaches is that the circumstances in which they work is typically restricted: simulation approaches are widely acknowledged to be more general and more robust (Lunneborg 2000). The formulae may work in interesting special cases, but the best general approach may be the simulation. Sometimes the simulation approach may be the only option. This principle is by no means restricted to statistics: economists, weather forecasters, engineers and many other scientists all make widespread use of simulation. In the words of Stephen Wolfram's “new kind of science ”... “there can be no way to know all the consequences of these rules [which describe the universe], except in effect just to watch and see how they unfold.” (Wolfram 2002, p. 846).

4. Combining simulation and conventional approaches

Despite the advantages of simulation approaches discussed above, there are of course many contexts in which conventional approaches are valuable. It is convenient to divide these into three categories.

4.1 Simulation approaches as a teaching aid

At present, in education, simulation approaches are normally seen as a teaching aid rather than a
substitute for the conventional formulae. The simulation is used to develop intuitions that help to
develop an understanding of conventional formula-based methods, as suggested by Simon, et al. (1976)
more than a quarter of century ago. Mills (2002) for example, provides a detailed review of how
simulation methods can be used to teach statistics, but does not even mention the possibility that
simulation methods may replace some standard methods: resampling and bootstrapping are not
mentioned in the main body of the article despite the fact that several articles on these topics are cited.
Garrett and Nash (2001) consider resampling methods in an article on “teaching the comparison of
variability to non-statistics students” but do not advocate them in this context because of the likely
complexity of the “recipe” for applying the methods, because there are choices within the resampling
framework about the appropriate method, and because they doubt whether the results would yield more
insight. However, it is important to remember that traditional methods are also complex (particularly for
people without an understanding of the background concepts), often involve the imposition of
inappropriate choices (requiring unreasonable assumptions of normality, for example), and the quotation
from the student in Ricketts and Berry (1994) above suggests that they do have the potential to offer
more insight than conventional approaches.

4.2 Simulation approaches for some tasks, conventional approaches for others

I can envisage two reasons for preferring a conventional approach for some tasks: the power or elegance
of the conventional method, and insights it may provide which are not provided by simulation.

As an example of the first, consider the normal probability distribution. It is easy to simulate
distributions which approximate the normal distribution, but the use of the mathematical formula – via
tables or a computer – is likely to be far easier and more elegant. Although simulation is helpful as a
teaching aid, there seems to me to be an extremely strong case for expecting even beginners to use tables
of the normal distribution or an equivalent computer function. Understanding the mathematical rationale
behind the formula (and so the tables and the computer functions) is obviously an unrealistic goal for
most people, but understanding the results empirically as a bell shaped curve which conforms to many
commonly experienced contexts is obviously realistic.

On the other hand, for the computation of confidence intervals, in my judgment, the simulation approach
of bootstrapping should be regarded as a replacement for formula based methods in most circumstances.
This is on the basis of an informal cost benefit analysis (Simon, et al. 1976, p. 738) balancing the costs
of learning about the method (much greater for conventional methods because of the pre-requisite
concepts which need mastering, and the variety of different methods for different statistics) with their
likely benefits (potentially greater for bootstrapping because of the generality of the approach, and
greater likelihood of the results being interpreted accurately). This, however, like the assertion in the
previous paragraph, just reflects my judgment. These judgments could be checked by means of a
controlled trial, although any such experiment would face many obvious difficulties (e.g. quantifying the
costs and the benefits, ensuring the comparison is “fair” in terms of learning resources).

The second reason for preferring conventional methods is that they may, sometimes, yield more insight.
For example, the standard formula for the confidence interval of a mean or a proportion shows how the
width of the intervals depends on the square root of the sample size. This is an aspect of the deep
structure of the statistical universe that cannot be unlocked by simulation. It can be illustrated, but not
proved or explained. Simulation approaches are in a sense cheating: they amount to little more than
some crude experiments with a model of the situation, and in some circumstances such experiments may
provide less insight than the sort of theory provided by a mathematical model.

4.3 Methods that use simulation and conventional approaches
Fairly obviously, some methods are likely to combine both approaches. For example, when using a small sample to derive a bootstrap confidence interval, it may make sense to use the data to fit a mathematically defined distribution and use that as the basis of Bucket 1 (see Section 2.1). Another possibility is that Bucket 2 may be used to estimate a standard error, which can then be interpreted in terms of the normal probability curve.

5. Conclusions

Simulation approaches to probability estimation and statistical inference have substantial advantages compared with conventional approaches: they tend to be more general, require far less technical background knowledge, and, because the methods are essentially sequences of physical actions, it is likely to be easier to understand their interpretation and limitations. This offers opportunities to reduce the technical content of statistics curricula, while at the same time provides students with approaches that are both potentially more powerful and more transparent than the conventional ones.

Despite this, there are obviously circumstances — discussed in the previous section — in which conventional methods have an obvious and useful role to play. A balance needs to be drawn between the two approaches. My feeling is that, in the U.K. at least, the virtues of simulation approaches are under-estimated. In particular, if all that is wanted is a transparent method of deriving answers to specific questions—and many students learning statistics are in this position — simulation approaches do seem adequate for a great many problems, and offer the promise of liberating statistics from the shackles of the symbolic arguments that many people find so difficult (Wood 2001).

References

Braun, W. J. (1995), “An illustration of bootstrapping using video lottery terminal data,” Journal of Statistics Education [Online], 3(2). ww2.amstat.org/publications/jse/v3n2/datasets.braun.html

Butler, A., Rothery, P., and Roy, D. (2003), “Minitab macros for resampling methods,” Teaching Statistics, 25(1), 22-25.

Davison, A. C., and Hinkley, D. V. (1997), Bootstrap Methods and Their Application, Cambridge: Cambridge University Press.

Diaconis, P., and Efron, B. (1983), “Computer intensive methods in statistics,” Scientific American, 248, 96-108.

Duckworth, W. M., and Stephenson, W. R. (2002), “Beyond traditional statistical methods,” The American Statistician, 56(3): 230-233.

Edgington, E. S. (1995), Randomization Tests, 3rd edition, New York: Dekker.

Efron, B., and Tibshirani, R. J. (1993), An Introduction to the Bootstrap, New York: Chapman and Hall.

Garrett, L., and Nash, J C. (2001), “Issues in teaching the comparison of variability to non?statistics students,” Journal of Statistics Education [Online], 9(2). ww2.amstat.org/publications/jse/v9n2/garrett.html
Good, P. I. (2001), *Resampling Methods*, 2nd edition, Boston: Birkhauser.

Gunter, B. (1991, December), “Bootstrapping: how to make something from almost nothing and get statistically valid answers — Part 1: Brave new world,” *Quality Progress*, 97-103.

Gunter, B. (1992a, February), “Bootstrapping: how to make something from almost nothing and get statistically valid answers — Part 2: the Confidence game,” *Quality Progress*, 83-86.

Gunter, B. (1992b, April), “Bootstrapping: how to make something from almost nothing and get statistically valid answers — Part 3: examples and enhancements,” *Quality Progress*, 119-122.

Lindley, D. V. (1985), *Making Decision*, 2nd Edition, London: Wiley.

Lunneborg, C. E. (2000), *Data Analysis by Resampling: Concepts and Applications*, Pacific Grove, CA, USA: Duxbury.

Mills, J. D. (2002), “Using computer simulation methods to teach statistics: a review of the literature,” *Journal of Statistics Education* [Online], 10(1). [ww2.amstat.org/publications/jse/v10n1/mills.html](http://ww2.amstat.org/publications/jse/v10n1/mills.html)

Noreen, E. W. (1989), *Computer Intensive Methods for Testing Hypotheses*, Chichester: Wiley.

Ricketts, C., and Berry, J. (1994), “Teaching statistics through resampling,” *Teaching Statistics*, 16(2), 41-44.

Simon, J. L. (1992), *Resampling: The New Statistics*, Arlington, VA: Resampling Stats, Inc.

Simon, J. L., Atkinson, D. T., and Shevokas, C. (1976), “Probability and statistics: experimental results of a radically different teaching method,” *American Mathematical Monthly*, 83(9), 733-739.

Skemp, R. R. (1976), “Relational understanding and instrumental understanding,” *Mathematics Teaching*, 77, 20-26.

Wolfram, S. (2002), *A New Kind of Science*, Champaign, IL: Wolfram Media.

Wood, M., Kaye, M., and Capon, N. (1999), “The use of resampling for estimating control chart limits,” *Journal of the Operational Research Society*, 50, 651-659.

Wood, M. (2001, May), “The case for crunchy methods in practical mathematics,” *Philosophy of Mathematics Education Journal* [Online], 14. [www.ex.ac.uk/~PErnest/](http://www.ex.ac.uk/~PErnest/)

Wood, M. (2003), *Making Sense of Statistics: A Non-Mathemematical Approach*, Basingstoke: Palgrave.

---

Michael Wood  
Department of Strategy and Business Systems  
University of Portsmouth  
Portsmouth  
U.K.  
*michael.wood@port.ac.uk*