# Part I
**Computational Statistics**

# Computational Statistics: An Introduction

**James E. Gentle, Wolfgang Härdle, Yuichi Mori**

# Computational Statistics
**1.1** # and Data Analysis

To do data analysis is to do computing. Statisticians have always been heavy users of whatever computing facilities are available to them. As the computing facilities have become more powerful over the years, those facilities have obviously decreased the amount of effort the statistician must expend to do routine analyses. As the computing facilities have become more powerful, an opposite result has occurred, however; the computational aspect of the statistician's work has increased. This is because of paradigm shifts in statistical analysis that are enabled by the computer.

Statistical analysis involves use of observational data together with domain knowledge to develop a model to study and understand a data-generating process. The data analysis is used to refine the model or possibly to select a different model, to determine appropriate values for terms in the model, and to use the model to make inferences concerning the process. This has been the paradigm followed by statisticians for centuries. The advances in statistical theory over the past two centuries have not changed the paradigm, but they have improved the specific methods. The advances in computational power have enabled newer and more complicated statistical methods. Not only has the exponentially-increasing computational power allowed use of more detailed and better models, however, it has shifted the paradigm slightly. Many alternative views of the data can be examined. Many different models can be explored. Massive amounts of simulated data can be used to study the model/data possibilities.

When exact models are mathematically intractable, approximate methods, which are often based on asymptotics, or methods based on estimated quantities must be employed. Advances in computational power and developments in theory have made *computational inference* a viable and useful alternative to the standard methods of asymptotic inference in traditional statistics. Computational inference is based on simulation of statistical models.

The ability to perform large numbers of computations almost instantaneously and to display graphical representations of results immediately has opened many new possibilities for statistical analysis. The hardware and software to perform these operations are readily available and are accessible to statisticians with no special expertise in computer science. This has resulted in a two-way feedback between statistical theory and statistical computing. The advances in statistical computing suggest new methods and development of supporting theory; conversely, the advances in theory and methods necessitate new computational methods.

Computing facilitates the development of statistical theory in two ways. One way is the use of symbolic computational packages to help in mathematical derivations (particularly in reducing the occurrences of errors in going from one line to the next!). The other way is in the quick exploration of promising (or unpromising!) methods by simulations. In a more formal sense also, simulations allow evaluation and comparison of statistical methods under various alternatives. This is a widely-used research method. For example, out of 61 articles published in the Theory and

Methods section of the *Journal of the American Statistical Association* in 2002, 50 reported on Monte Carlo studies of the performance of statistical methods. A general outline of many research articles in statistics is

1.  state the problem and summarize previous work on it,
2.  describe a new approach,
3.  work out some asymptotic properties of the new approach,
4.  conduct a Monte Carlo study showing the new approach in a favorable light.

Much of the effort in mathematical statistics has been directed toward the easy problems of exploration of asymptotic properties. The harder problems for finite samples require different methods. Carefully conducted and reported Monte Carlo studies often provide more useful information on the relative merits of statistical methods in finite samples from a range of model scenarios.

While to do data analysis is to compute, we do not identify all data analysis, which necessarily uses the computer, as "statistical computing" or as "computational statistics". By these phrases we mean something more than just using a statistical software package to do a standard analysis. We use the term "statistical computing" to refer to the computational methods that enable statistical methods. Statistical computing includes numerical analysis, database methodology, computer graphics, software engineering, and the computer/human interface. We use the term "computational statistics" somewhat more broadly to include not only the methods of statistical computing, but also statistical methods that are computationally intensive. Thus, to some extent, "computational statistics" refers to a large class of modern statistical methods. Computational statistics is grounded in mathematical statistics, statistical computing, and applied statistics. While we distinguish "computational statistics" from "statistical computing", the emergence of the field of computational statistics was coincidental with that of statistical computing, and would not have been possible without the developments in statistical computing.

One of the most significant results of the developments in statistical computing during the past few decades has been the statistical software package. There are several of these, but a relatively small number that are in widespread use. While referees and editors of scholarly journals determine what statistical theory and methods are *published*, the developers of the major statistical software packages determine what statistical methods are *used*. Computer programs have become necessary for statistical analysis. The specific methods of a statistical analysis are often determined by the available software. This, of course, is not a desirable situation, but, ideally, the two-way feedback between statistical theory and statistical computing dimishes the effect over time.

The importance of computing in statistics is also indicated by the fact that there are at least ten major journals with titles that contain some variants of both "computing" and "statistics". The journals in the mainstream of statistics without "computing" in their titles also have a large proportion of articles in the fields of statistical computing and computational statistics. This is because, to a large extent, recent developments in statistics and in the computational sciences have

gone hand in hand. There are also two well-known learned societies with a primary focus in statistical computing: the International Association for Statistical Computing (IASC), which is an affiliated society of the International Statistical Institute (ISI), and the Statistical Computing Section of the American Statistical Association (ASA). There are also a number of other associations focused on statistical computing and computational statistics, such as the Statistical Computing Section of the Royal Statistical Society (RSS), and the Japanese Society of Computational Statistics (JSCS).

Developments in computing and the changing role of computations in statistical work have had significant effects on the curricula of statistical education programs both at the graduate and undergraduate levels. Training in statistical computing is a major component in some academic programs in statistics (see Gentle, 2004, Lange, 2004, and Monahan, 2004). In all academic programs, some amount of computing instruction is necessary if the student is expected to work as a statistician. The extent and the manner of integration of computing into an academic statistics program, of course, change with the developments in computing hardware and software and advances in computational statistics.

We mentioned above the two-way feedback between statistical theory and statistical computing. There is also an important two-way feedback between applications and statistical computing, just as there has always been between applications and any aspect of statistics. Although data scientists seek commonalities among methods of data analysis, different areas of application often bring slightly different problems for the data analyst to address. In recent years, an area called "data mining" or "knowledge mining" has received much attention. The techniques used in data mining are generally the methods of exploratory data analysis, of clustering, and of statistical learning, applied to very large and, perhaps, diverse datasets. Scientists and corporate managers alike have adopted data mining as a central aspect of their work. Specific areas of application also present interesting problems to the computational statistician. Financial applications, particularly risk management and derivative pricing, have fostered advances in computational statistics. Biological applications, such as bioinformatics, microarray analysis, and computational biology, are fostering increasing levels of interaction with computational statistics.

The hallmarks of computational statistics are the use of more complicated models, larger datasets with both more observations and more variables, unstructured and heterogeneous datasets, heavy use of visualization, and often extensive simulations.

## 1.2 The Emergence of a Field of Computational Statistics

Statistical computing is truly a multidisciplinary field and the diverse problems have created a yeasty atmosphere for research and development. This has been the

case from the beginning. The roles of statistical laboratories and the applications that drove early developments in statistical computing are surveyed by Grier (1999). As digital computers began to be used, the field of statistical computing came to embrace not only numerical methods but also a variety of topics from computer science.

The development of the field of statistical computing was quite fragmented, with advances coming from many directions – some by persons with direct interest and expertise in computations, and others by persons whose research interests were in the applications, but who needed to solve a computational problem. Through the 1950s the major facts relevant to statistical computing were scattered through a variety of journal articles and technical reports. Many results were incorporated into computer programs by their authors and never appeared in the open literature. Some persons who contributed to the development of the field of statistical computing were not aware of the work that was beginning to put numerical analysis on a sound footing. This hampered advances in the field.

## Early Developments in Statistical Computing    1.2.1

An early book that assembled much of the extant information on digital computations in the important area of linear computations was by Dwyer (1951). In the same year, Von Neumann's (1951) NBS publication described techniques of random number generation and applications in Monte Carlo. At the time of these publications, however, access to digital computers was not widespread. Dwyer (1951) was also influential in regression computations performed on calculators. Some techniques, such as use of "machine formulas", persisted into the age of digital computers.

Developments in statistical computing intensified in the 1960s, as access to digital computers became more widespread. Grier (1991) describes some of the effects on statistical practice by the introduction of digital computers, and how statistical applications motivated software developments. The problems of rounding errors in digital computations were discussed very carefully in a pioneering book by Wilkinson (1963). A number of books on numerical analysis using digital computers were beginning to appear. The techniques of random number generation and Monte Carlo were described by Hammersley and Handscomb (1964). In 1967 the first book specifically on statistical computing appeared, Hemmerle (1967).

## Early Conferences and Formation of Learned Societies    1.2.2

The 1960s also saw the beginnings of conferences on statistical computing and sections on statistical computing within the major statistical societies. The Royal Statistical Society sponsored a conference on statistical computing in December 1966. The papers from this conference were later published in the RSS's *Applied Statistics* journal. The conference led directly to the formation of a Working Party on Statistical Computing within the Royal Statistical Society. The first Symposium on the Interface of Computer Science and Statistics was held February 1,

1967. This conference has continued as an annual event with only a few exceptions since that time (see Goodman, 1993, Billard and Gentle, 1993, and Wegman, 1993). The attendance at the Interface Symposia initially grew rapidly year by year and peaked at over 600 in 1979. In recent years the attendance has been slightly under 300. The proceedings of the Symposium on the Interface have been an important repository of developments in statistical computing. In April, 1969, an important conference on statistical computing was held at the University of Wisconsin. The papers presented at that conference were published in a book edited by Milton and Nelder (1969), which helped to make statisticians aware of the useful developments in computing and of their relevance to the work of applied statisticians.

In the 1970s two more important societies devoted to statistical computing were formed. The Statistical Computing Section of the ASA was formed in 1971 (see Chambers and Ryan, 1990). The Statistical Computing Section organizes sessions at the annual meetings of the ASA, and publishes proceedings of those sessions. The International Association for Statistical Computing (IASC) was founded in 1977 as a Section of ISI. In the meantime, the first of the biennial COMPSTAT Conferences on computational statistics was held in Vienna in 1974. Much later, regional sections of the IASC were formed, one in Europe and one in Asia. The European Regional Section of the IASC is now responsible for the organization of the COMPSTAT conferences.

Also, beginning in the late 1960s and early 1970s, most major academic programs in statistics offered one or more courses in statistical computing. More importantly, perhaps, instruction in computational techniques has permeated many of the standard courses in applied statistics.

As mentioned above, there are several journals whose titles include some variants of both "computing" and "statistics". The first of these, the *Journal of Statistical Computation and Simulation*, was begun in 1972. There are dozens of journals in numerical analysis and in areas such as "computational physics", "computational biology", and so on, that publish articles relevant to the fields of statistical computing and computational statistics.

By 1980 the field of statistical computing, or computational statistics, was well-established as a distinct scientific subdiscipline. Since then, there have been regular conferences in the field, there are scholarly societies devoted to the area, there are several technical journals in the field, and courses in the field are regularly offered in universities.

## 1.2.3    The PC

The 1980s was a period of great change in statistical computing. The personal computer brought computing capabilities to almost everyone. With the PC came a change not only in the number of participants in statistical computing, but, equally important, completely different attitudes toward computing emerged. Formerly, to do computing required an account on a mainframe computer. It required laboriously entering arcane computer commands onto punched cards, taking these

cards to a card reader, and waiting several minutes or perhaps a few hours for some output – which, quite often, was only a page stating that there was an error somewhere in the program. With a personal computer for the exclusive use of the statistician, there was no incremental costs for running programs. The interaction was personal, and generally much faster than with a mainframe. The software for PCs was friendlier and easier to use. As might be expected with many non-experts writing software, however, the general quality of software probably went down.

The democratization of computing resulted in rapid growth in the field, and rapid growth in software for statistical computing. It also contributed to the changing paradigm of the data sciences.

## The Cross Currents of Computational Statistics    1.2.4

Computational statistics of course is more closely related to statistics than to any other discipline, and computationally-intensive methods are becoming more commonly used in various areas of application of statistics. Developments in other areas, such as computer science and numerical analsysis, are also often directly relevant to computational statistics, and the research worker in this field must scan a wide range of literature.

Numerical methods are often developed in an ad hoc way, and may be reported in the literature of any of a variety of disciplines. Other developments important for statistical computing may also be reported in a wide range of journals that statisticians are unlikely to read. Keeping abreast of relevant developments in statistical computing is difficult not only because of the diversity of the literature, but also because of the interrelationships between statistical computing and computer hardware and software.

An example of an area in computational statistics in which significant developments are often made by researchers in other fields is Monte Carlo simulation. This technique is widely used in all areas of science, and researchers in various areas often contribute to the development of the science and art of Monte Carlo simulation. Almost any of the methods of Monte Carlo, including random number generation, are important in computational statistics.

## Literature    1.2.5

Some of the major periodicals in statistical computing and computational statistics are the following. Some of these journals and proceedings are refereed rather rigorously, some refereed less so, and some are not refereed.

— *ACM Transactions on Mathematical Software*, published quarterly by the ACM (Association for Computing Machinery), includes algorithms in Fortran and C. Most of the algorithms are available through `netlib`. The ACM collection of algorithms is sometimes called *CALGO*.
www.acm.org/toms/

- *ACM Transactions on Modeling and Computer Simulation*, published quarterly by the ACM.
  www.acm.org/tomacs/
- *Applied Statistics*, published quarterly by the Royal Statistical Society. (Until 1998, it included algorithms in Fortran. Some of these algorithms, with corrections, were collected by Griffiths and Hill, 1985. Most of the algorithms are available through statlib at Carnegie Mellon University.)
  www.rss.org.uk/publications/
- *Communications in Statistics – Simulation and Computation*, published quarterly by Marcel Dekker. (Until 1996, it included algorithms in Fortran. Until 1982, this journal was designated as *Series B*.)
  www.dekker.com/servlet/product/productid/SAC/
- *Computational Statistics* published quarterly by Physica-Verlag (formerly called *Computational Statistics Quarterly*).
  comst.wiwi.hu-berlin.de/
- *Computational Statistics. Proceedings of the xx-th Symposium on Computational Statistics* (COMPSTAT), published biennially by Physica-Verlag/Springer.
- *Computational Statistics & Data Analysis*, published by Elsevier Science. There are twelve issues per year. (This is also the official journal of the International Association for Statistical Computing and as such incorporates the *Statistical Software Newsletter*.)
  www.cbs.nl/isi/csda.htm
- *Computing Science and Statistics*. This is an annual publication containing papers presented at the Interface Symposium. Until 1992, these proceedings were named *Computer Science and Statistics: Proceedings of the xx-th Symposium on the Interface*. (The 24th symposium was held in 1992.) In 1997, Volume 29 was published in two issues: Number 1, which contains the papers of the regular Interface Symposium; and Number 2, which contains papers from another conference. The two numbers are not sequentially paginated. Since 1999, the proceedings have been published only in CD-ROM form, by the Interface Foundation of North America.
  www.galaxy.gmu.edu/stats/IFNA.html
- *Journal of Computational and Graphical Statistics*, published quarterly as a joint publication of ASA, the Institute of Mathematical Statistics, and the Interface Foundation of North America.
  www.amstat.org/publications/jcgs/
- *Journal of the Japanese Society of Computational Statistics*, published once a year by JSCS.
  www.jscs.or.jp/oubun/indexE.html
- *Journal of Statistical Computation and Simulation*, published in twelve issues per year by Taylor & Francis.
  www.tandf.co.uk/journals/titles/00949655.asp
- *Proceedings of the Statistical Computing Section*, published annually by ASA.
  www.amstat.org/publications/

- *SIAM Journal on Scientific Computing*, published bimonthly by SIAM. This journal was formerly *SIAM Journal on Scientific and Statistical Computing*. www.siam.org/journals/sisc/sisc.htm
- *Statistical Computing & Graphics Newsletter*, published quarterly by the Statistical Computing and the Statistical Graphics Sections of ASA. www.statcomputing.org/
- *Statistics and Computing*, published quarterly by Chapman & Hall.

In addition to literature and learned societies in the traditional forms, an important source of communication and a repository of information are computer databases and forums. In some cases, the databases duplicate what is available in some other form, but often the material and the communications facilities provided by the computer are not available elsewhere.

# Why This Handbook                                                                     1.3

The purpose of this handbook is to provide a survey of the basic concepts of computational statistics; that is, *Concepts and Fundamentals.* A glance at the table of contents reveals a wide range of articles written by experts in various subfields of computational statistics. The articles are generally expository, taking the reader from the basic concepts to the current research trends. The emphasis throughout, however, is on the concepts and fundamentals. Most chapters have extensive and up-to-date references to the relevant literature (with, in many cases, perhaps a perponderance of self-references!)

We have organized the topics into Part II on "statistical computing", that is, the computational methodology, and Part III "statistical methodology", that is, the techniques of applied statistics that are computer-intensive, or otherwise make use of the computer as a tool of discovery, rather than as just a large and fast calculator. The final part of the handbook covers a number of application areas in which computational statistics plays a major role are surveyed.

## Summary and Overview; Part II: Statistical Computing                              1.3.1

The thirteen chapters of Part II, Statistical Computing, cover areas of numerical analysis and computer science or informatics that are relevant for statistics. These areas include computer arithmetic, algorithms, database methodology, languages and other aspects of the user interface, and computer graphics.

In the first chapter of this part, Monahan describes how numbers are stored on the computer, how the computer does arithmetic, and more importantly what the implications are for statistical (or other) computations. In this relatively short chapter, he then discusses some of the basic principles of numerical algorithms, such as divide and conquer. Although many statisticians do not need to know the details, it is important that all statisticians understand the implications of

computations within a system of numbers and operators that is not the same system that we are accustomed to in mathematics. Anyone developing computer algorithms, no matter how trivial the algorithm may appear, must understand the details of the computer system of numbers and operators.

One of the important uses of computers in statistics, and one that is central to computational statistics, is the simulation of random processes. This is a theme we will see in several chapters of this handbook. In Part II, the basic numerical methods relevant to simulation are discussed. First, L'Ecuyer describes the basics of random number generation, including assessing the quality of random number generators, and simulation of random samples from various distributions. Next Chib describes one special use of computer-generated random numbers in a class of methods called Markov chain Monte Carlo. These two chapters describe the basic numerical methods used in computational inference. Statistical methods using simulated samples are discussed further in Part III.

The next four chapters of Part II address specific numerical methods. The first of these, methods for linear algebraic computations, are discussed by Čížková and Čížek. These basic methods are used in almost all statistical computations. Optimization is another basic method used in many statistical applications. Chapter II.5 on the EM algorithm and its variations by Ng, Krishnan, and McLachlan, and Chap. II.6 on stochastic optimization by Spall address two specific areas of optimization. Finally, in Chap. II.7, Vidakovic discusses transforms that effectively restructure a problem by changing the domain. These transforms are statistical functionals, the most well-known of which are Fourier transforms and wavelet transforms.

The next two chapters focus on efficient usage of computing resources. For numerically-intensive applications, parallel computing is both the most efficient and the most powerful approach. In Chap. II.8 Nakano describes for us the general principles, and then some specific techniques for parallel computing. Understanding statistical databases is important not only because of the enhanced efficiency that appropriate data structures allow in statistical computing, but also because of the various types of databases the statistician may encounter in data analysis. In Chap. II.9 on statistical databases, Boyens, Günther, and Lenz give us an overview of the basic design issues and a description of some specific database management systems.

The next two chapters are on statistical graphics. The first of these chapters, by Symanzik, spans our somewhat artificial boundary of Part II (statistical computing) and Part III (statistical methodology, the real heart and soul of computational statistics). This chapter covers some of the computational details, but also addresses the usage of interactive and dynamic graphics in data analysis. Wilkinson, in Chap. II.11, describes a paradigm, the grammar of graphics, for developing and using systems for statistical graphics.

In order for statistical software to be usable and useful, it must have a good user interface. In Chap. II.12 on statistical user interfaces, Klinke discusses some of the general design principles of a good user interface and describes some interfaces that are implemented in current statistical software packages. In the development and

use of statistical software, an object oriented approach provides a consistency of design and allows for easier software maintenance and the integration of software developed by different people at different times. Virius discusses this approach in the final chapter of Part II, on object oriented computing.

## Summary and Overview; Part III: Statistical Methodology

Part III covers several aspects of computational statistics. In this part the emphasis is on the statistical methodology that is enabled by computing. Computers are useful in all aspects of statistical data analysis, of course, but in Part III, and generally in computational statistics, we focus on statistical methods that are computationally intensive. Although a theoretical justification of these methods often depends on asymptotic theory, in particular, on the asymptotics of the empirical cumulative distribution function, asymptotic inference is generally replaced by computational inference.

The first three chapters of this part deal directly with techniques of computational inference; that is, the use of cross validation, resampling, and simulation of data-generating processes to make decisions and to assign a level of confidence to the decisions. Wang opens Part III with a discussion of model choice. Selection of a model implies consideration of more than one model. As we suggested above, this is one of the hallmarks of computational statistics: looking at data through a variety of models. Wang begins with the familiar problem of variable selection in regression models, and then moves to more general problems in model selection. Cross validation and generalizations of that method are important techniques for addressing the problems. Next, in Chap. III.2 Mammen and Nandi discuss a class of resampling techniques that have wide applicability in statistics, from estimating variances and setting confidence regions to larger problems in statistical data analysis. Computational inference depends on simulation of data-generating processes. Any such simulation is an *experiment*. In the third chapter of Part III, Kleijnen discusses principles for design and analysis of experiments using computer models.

In Chap. III.4, Scott considers the general problem of estimation of a multivariate probability density function. This area is fundamental in statistics, and it utilizes several of the standard techniques of computational statistics, such as cross validation and visualization methods.

The next four chapers of Part III address important issues for discovery and analysis of relationships among variables. First, Loader discusses local smoothing using a variety of methods, including kernels, splines, and orthogonal series. Smoothing is fitting of asymmetric models, that is, models for the effects of a given set of variables ("independent variables") on another variable or set of variables. The methods of Chap. III.5 are generally nonparametric, and will be discussed from a different standpoint in Chap. III.10. Next, in Chap. III.6 Mizuta describes ways of using the relationships among variables to reduce the effective dimensionality

of a problem. The next two chapters return to the use of asymmetric models: Müller discusses generalized linear models, and Čížek describes computational and inferential methods for dealing with nonlinear regression models.

In Chap. III.9, Gather and Davies discuss various issues of robustness in statistics. Robust methods are important in such applications as those in financial modeling, discussed in Chap. IV.2. One approach to robustness is to reduce the dependence on parametric assumptions. Horowitz, in Chap. III.10, describes semi-parametric models that make fewer assumptions about the form.

One area in which computational inference has come to play a major role is in Bayesian analysis. Computational methods have enabled a Bayesian approach in practical applications, because no longer is this approach limited to simple problems or conjugate priors. Robert, in Chap. III.11, describes ways that computational methods are used in Bayesian analyses.

Survival analysis, with applications in both medicine and product reliability, has become more important in recent years. Kamakura, in Chap. III.12, describes various models used in survival analysis and the computational methods for analyzing such models.

The final four chapters of Part III address an exciting area of computational statistics. The general area may be called "data mining", although this term has a rather anachronistic flavor because of the hype of the mid-1990s. Other terms such as "knowledge mining" or "knowledge discovery in databases" ("KDD") are also used. To emphasize the roots in artificial intelligence, which is a somewhat discredited area, the term "computational intelligence" is also used. This is an area in which machine learning from computer science and statistical learning have merged. In Chap. III.13 Wilhelm provides an introduction and overview of data and knowledge mining, as well as a discussion of some of the vagaries of the terminology as researchers have attempted to carve out a field and to give it scientific legitimacy. Subsequent chapters describe specific methods for statistical learning: Zhang discusses recursive partitioning and tree based methods; Mika, Schäfer, Laskov, Tax, and Müller discuss support vector machines; and Bühlmann describes various ensemble methods.

## Summary and Overview; Part IV: Selected Applications

1.3.3

Finally, in Part IV, there are five chapters on various applications of computational statistics. The first, by Weron, discusses stochastic modeling of financial data using heavy-tailed distributions. Next, in Chap. IV.2 Bauwens and Rombouts describe some problems in economic data analysis and computational statistical methods to address them. Some of the problems, such as nonconstant variance, discussed in this chapter on econometrics are also important in finance.

Human biology has become one of the most important areas of application, and many computationally-intensive statistical methods have been developed, refined, and brought to bear on problems in this area. First, Vaisman describes approaches

to understanding the geometrical structure of protein molecules. While much is known about the order of the components of the molecules, the three-dimensional structure for most important protein molecules is not known, and the tools for discovery of this structure need extensive development. Next, Eddy and McNamee describe some statistical techniques for analysis of MRI data. The important questions involve the functions of the various areas in the brain. Understanding these will allow more effective treatment of diseased or injured areas and the resumption of more normal activities by patients with neurological disorders.

Finally, Marchette discusses statistical methods for computer network intrusion detection. Because of the importance of computer networks around the world, and because of their vulnerability to unauthorized or malicious intrusion, detection has become one of the most important – and interesting – areas for data mining.

The articles in this handbook cover the important subareas of computational statistics and give some flavor of the wide range of applications. While the articles emphasize the basic concepts and fundamentals of computational statistics, they provide the reader with tools and suggestions for current research topics. The reader may turn to a specific chapter for background reading and references on a particular topic of interest, but we also suggest that the reader browse and ultimately peruse articles on unfamiliar topics. Many surprising and interesting tidbits will be discovered!

## The Ehandbook                                                         1.3.4

A unique feature of this handbook is the supplemental ebook format. Our ebook design offers a HTML file with links to world wide computing servers. This HTML version can be downloaded onto a local computer via a licence card included in this handbook.

## Future Handbooks in Computational Statistics                         1.3.5

This handbook on concepts and fundamentals sets the stage for future handbooks that go more deeply into the various subfields of computational statistics. These handbooks will each be organized around either a specific class of theory and methods, or else around a specific area of application.

The development of the field of computational statistics has been rather fragmented. We hope that the articles in this handbook series can provide a more unified framework for the field.

# References

Billard, L. and Gentle, J.E. (1993). The middle years of the Interface, *Computing Science and Statistics*, 25:19–26.

Chambers, J.M. and Ryan, B.F. (1990). The ASA Statistical Computing Section, *The American Statistician*, 44(2):87–89.

Dwyer, P.S. (1951), *Linear Computations*, John Wiley and Sons, New York.

Gentle, J.E. (2004). Courses in statistical computing and computational statistics, *The American Statistician*, 58:2–5.

Goodman, A. (1993). Interface insights: From birth into the next century, *Computing Science and Statistics*, 25:14–18.

Grier, D.A. (1991). Statistics and the introduction of digital computers, *Chance*, 4(3):30–36.

Grier, D.A. (1999), Statistical laboratories and the origins of statistical computing, *Chance*, 4(2):14–20.

Hammersley, J.M. and Handscomb, D.C. (1964). *Monte Carlo Methods*, Methuen & Co., London.

Hemmerle, W.J. (1967). *Statistical Computations on a Digital Computer*. Blaisdell, Waltham, Massachusetts.

Lange, K. (2004). Computational Statistics and Optimization Theory at UCLA, *The American Statistician*, 58:9–11.

Milton, R. and Nelder, J. (eds) (1969). *Statistical Computation*, Academic Press, New York.

Monahan, J. (2004). Teaching Statistical Computing at NC State, *The American Statistician*, 58:6–8.

Von Neumann, J. (1951). *Various Techniques Used in Connection with Random Digits*, National Bureau of Standards Symposium, NBS Applied Mathematics Series 12, National Bureau of Standards (now National Institute of Standards and Technology), Washington, DC.

Wegman, E.J. (1993). History of the Interface since 1987: The corporate era, *Computing Science and Statistics*, 25:27–32.

Wilkinson, J. H. (1963). *Rounding Errors in Algebraic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.