Journal of Mathematical Research & Exposition
Jan., 2009, Vol. 29, No. 1, pp. 153–163
DOI:10.3770/j.issn:1000-341X.2009.01.020
Http://jmre.dlut.edu.cn

Brief Introduction to and Review on Elements of Computational Statistics

XU Li Quan[1], *WANG Zi Kun*[2]
(1. Wuxi Teaching & Research Center, Jiangsu 214001, China;
2. Department of Mathematics, Beijing Normal University, Beijing 100875, China)
(E-mail: Liquanwx@sohu.com)

Abstract A brief introduction and review of new book about the Elements of Computational Statistics is given.

Keywords James E. Gentle; Elements of Computational Statistics; computationally intensive; Monte Carlo methods.

Document code A
MR(2000) Subject Classification 62-XX; 65-XX
Chinese Library Classification O213

1. Briefing on terms about Computational Statistics and introduction of the background

Recently, the selected books of mathematics famous series from abroad have come out by the Committee of Experts of China. These reprints have been authorized by Springer-Verlag for sale in the People's Republic of China. No doubt about it, the "Elements of Computational Statistics"[1] among them is a book that gives ample evidence of the author's scholarship. It helps the statistical workers keep abreast of current development in the field and the complete picture at home and broad.

In recent years, statistics and computer science advance in paralleled columns each assisting and stimulating the other–the computer science or the "computational" sciences are called into play when requiring the mass data storage and processing technology in statistics and they depend upon visualization of many projections of the data; on the other hand, application of computers in statistics stimulate its development and broaden its foundation. Therefore, many new developments of statistics rely on the development of computer science and technology. Many of the interesting statistical methods are in vogue now, amoung them are the computationally intensive methods. The theories which have computationally intensive characteristic statistical methods and support of these constituted a curriculum which is called "Computational Statistics". "Computational Statistics" is different from "statistical computing", but it is indissolubly linked with that of the statistical computing methods. From a statistician's point of view, it

Received date: 2006-09-18; **Accepted date**: 2008-06-20

means "computational methods, including numerical analysis, for himself". Its domain involved duplicate sampling, classification, and multiple transform of datasets, maybe yet to make use of artificial data random generating, and so on. The computationally intensive methods of modern statistics can not do without the development of general statistical computing methods and numerical analysis.

With other computational sciences, such as computational physics, computational biology, etc., computational statistics have two characteristics: one of which is a characteristic of the methodology, i.e., computationally intensive, and the other is the nature of the tools of discovery. The general methods of science discovery are deductive reasoning (logical method) or plausible reasoning (involving observation, experiment, and conjecture etc.), or the combination of both. In addition, computer simulation has become a new means of discovery to use computer exploring greater number of specific events at present. This method is similar to experimentation in some respects.

The development and progress in computing hardware and software are changing now or have changed the nature of routine work of statisticians. Those data analysts and applied statisticians make use of computers for storage of data, analysis of the data, and making analysis reports. The statisticians and even probabilists make use of computers for symbolic manipulations, evaluation of expressions, ad hoc simulations, and production of research reports and papers. Statistics is only to pay close attention to the collections and analysis of scientific data among the mathematics sciences. In recent years, every advanced statistician can sense the impact of remarkable increase brought by the amount of dada. Undoubtedly, the affect of data is sensitive for statisticians, such as the change from the use of "critical values" of test statistics and the use of "p-values". Nevertheless, there are some essential elements more than the former, such as making use of multivariate and/or nonlinear models to substitute univariate linear models, etc. They will play a more important role in mathematical statistics. Perhaps we merely focus on how to compute their approximation of interested data, i.e., how to compute easier? Then it would be better to give consideration of effects after the computers are put into service now. Recently, computational inference using Monte Carlo methods has been replacing asymptotic approximations. "Another major effect that developments in computing have had on the practice of statistics is that many Bayesian methods that were formerly impractical have entered the mainstream of statistical applications."[1]

The development and efficiency in computing technology in the 1990's makes the Bayesian methods be able to be realized in very wide models. The challenges in the coming decades are to fully research and develop the vital nexus linking the Bayesian methods and modern nonparametric as well as semi-parametric statistics methods, involving the research combining the Bayesian methods with frequency theory methods as much as possible.

An apparent result is that the conceptions of unbiasedness and approximate unbiasedness will be utterly useless in regard of a great deal of multivariable data models with enormous capacity, for example, MLE (maximum likelihood estimate). Because the connotative conception of data integrated in statistical methods shall become meaningless with the complexity and variability

of unbiased methods. It is for this reason that we need wider or more comprehensive "biased estimate theories" and other new theories to process a great deal of multivariable tremendous amounts of data. Besides increasing of Monte Carlo methods in modeling practices, obviously an in-depth study of Monte Carlo methods used in deduction is needed as well.

The experimentation has entered into statisticians' "tool-box". The emergence of a large scale of computing system makes statisticians come to understand the essence of discovery in a new light. Science always develops with its new discovery. Exploratory data analysis has been widely used all over the world; data mining techniques (i.e. seeking for information and knowledge from material) have enabled statisticians to increase the rate of finding things that are not being sought.

Data mining is a cross subjects, concentrated by distinct subjects and domain such as database, artificial intelligence, and statistics, visualization, and parallel computing, etc.. It has caused wide public concern over the recent years. The aims of data mining are to find laws behind them or relationship between them from a great quantity data, so that it serves decision-making. The display of large quantities of data is an important aspect of data mining. As far as visualization of the data system is concerned, the analysts are easily confused in face of the data. Visualized tool of data mining is able to path through fruitful starting point of exploration, and present data through proper metaphor, which provides good assistance for data analysts. The advent of the data mining provides a whole new application area for statistics, and it also brought forward fresh subjects for theoretical research of statistics. It will undoubtedly promote vigorous development of statistics[2, 3].

The computing has been taken as a means of discovering in the statistics. The computer plays a role of not only storing data, performing computing tasks, and producing pictures and tables, etc.; furthermore, it provides the scientists models and theories to choose from. The displayed graphic exhibitions offered by given data set usually reflects the general character of computational statistics. Another significant characteristic of computational statistics is the intensivism of computational methods. Even for those data sets of medium size, high-performance computers may be required to perform the computations, because the multiple analyses may result in a large number of computations. How to distinguish size of datasets? Reinhold Huber's categories about size in 1994 may sound cliché, but it is a fact of life: very small 10^2, small 10^4, medium 10^6 , big 10^8, very big 10^{10}, where the unit is bit. For example, in the Center of Stanford Linear Accelerator, the size of a database is about 5×10^{15} bits while making a uniparticle physical test by using BaBar detector[2]. At present, what is the most pressing is to be accustomed to meet the demand of such database for the modern statistics, because they are so complex and so huge that new ideas and methods are required.

2. Briefing on the book "Elements of Computational Statistics"

James E. Gentle, author of the book "Elements of Computational Statistics", is a Prof. of Dept. of CALC-Sta., George Mason University and a member of the American Statistical

Association (ASA) and International Statistical Association. He owns several of national level firms in ASA; he is subeditor of the journal ASA and editors of several well-known journals on statistics and computational sciences around the world, as well. Moreover, he is author of the book "Numerical Linear Algebra for Applications in Statistics and Random Number Generation and Monte Carlo Methods". The second edition is greatly improved on the first edition[4, 5].

James E. Gentle emphasized application methods and technology of Computational Statistics in this book, chiefly concerning some applications in every direction such as the density estimation, confirming data structures, and model building so on. Although there are no expositions particularly of statistical computing methods, this book expounded in all fields of data conversion, approximating functions, and numerical technique of data optimization. The author points up: "The book grew out of a semester course in 'Computational Statistics' and various courses called 'Topics in Computational Statistics' that I have offered at George Mason University over the past several years. The book is part of a much larger tome that also covers many topics in numerical analysis. See http://www.science.gmu.edu/ jgentle/cmpstbk/. Many of the topics addressed in this book could easily be (and are) subjects for full-length books"([1], Preface). In this book, the author laid particular emphasis on exploring a general manner and commonalities among the data processing system. "An example of a basic tool used in a variety of settings in computational statistics is the decomposition of a function so that it has a probability density as a factor." This technique is arranged in Chapter 2 (Monte Carlo Methods for Statistical Inference, see page 52), in Chapters 6 and 9 (function estimation), and in Chapter 10 (projection pursuit).

This book discussed mainly statistical methods and their applications are computationally intensive. This is the reason why we think of the domain is called Computational Statistics. However as we have mentioned above, statistical analysis that people work on is of the characteristic of statistical computing. The computing usually considers an experiment, and the computer has been used as a powerful means for discovery.

The first part of this book describes a general method and technique. Chapter 1 reviewed some points of the basic ideas and the computing methods. We may regard a data-generating process as the subject of statistical analysis; the direct analytic object is a dataset brought out from this process. We may use too many normative statistical means to make analysis for the data, proceeding to deduction for the process. Simulation of the data-generating process is one of the most important means on computational statistics. To deduce for computing by simulating, some of standard principles of statistical inference are employed in computational inference.

The second chapter is about Monte Carlo Methods for statistical inference and its application in computational inference, including Monte Carlo tests. Markov chain Monte Carlo (MCMC) is important method on analogy computing. The research in MCMC theory sets up a broad prospect for practical applications of statistical model. Since the 1990s, many application problems suffer from the difficulties in analyzing complicated subjects and figuring out model's structure. According to MCMC theory at present, a lot of the complex nature of the problems can be solved by means of a special software system to simulate MCMC. In addition, benefited

from MCMC theory, the use of Bayes statistics is restored to a flourishing state. The previous statistical methods believed impossible for computing have now become common[2, 3].

Chapters 3 and 4 are focused on computational inference using resampling and partitioning of a given dataset. These methods include randomization tests, jackknife techniques, and bootstrap methods, and the Monte Carlo sampling. Chapter 5 discusses methods of projecting higher-dimensional data onto lower dimensions; Chapter 6 covers some of the general issues in function estimation and numerical technique in data optimization; and Chapter 7 presents a brief overview of some graphical methods, especially those concerned with multidimensional data. The more complicated the structure of the data and the higher the dimension, the more ingenuity is required for visualization of the data; however, in this case graphics is most important. The orientation of the discussion on graphics is that of computational statistics; the emphasis is on discovery; and the important issues that should be considered in making presentation graphics are not addressed. The tools discussed in Chapter 5 will also be used for clustering and classification, and, in general, for exploring structure in data.

Identification of interesting features, or "structure", in data is an important activity in computational statistics. In part II, the author considered the problem of identification of structure and the general problems of estimation of probability densities. The most useful and complete description of a random data generating process is the associated probability density, if it exists. Estimation of this special type of function is the topic of chapters 8 and 9, building on general methods discussed in earlier chapters, especially Chapter 6. If the data follow a parametric distribution, or rather, if we are willing to assume that the data follow a parametric distribution, identification of the probability density is accomplished by estimation of the parameters. Nonparametric density estimation is considered in chapter 9[1], Preface.

Although the CDF in some ways is more fundamental in characterizing a probability distribution (it always exists and is defined the same for both continuous and discrete distributions), the density probability function is more familiar to most data analysts. Important properties such as skewness, modes, and so on, can be seen more readily from a plot of the probability density function than from a plot of the CDF. We are therefore usually more interested in estimating the density p, than the CDF P. Some methods of estimating the density, however, are based on estimates of the CDF. The simplest estimate of the CDF is the empirical cumulative distribution function, the ECDF, which is defined as

$$P_n(y) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,y)}(y_i).$$

As we know, the ECDF is pointwise unbiased for the CDF.

The derivative of the ECDF is called the empirical probability density function (EPDF),

$$p_n(y) = \frac{1}{n}\sum_{i=1}^{n} \delta(y - y_i)$$

where $\delta$is the Dirac delta function, which is just a series of spikes at points corresponding to the observed values. It is not very useful as an estimator of the probability density. It is, however,

unbiased for the probability density function at any point.

In the absence of assumptions about the form of the density p, the estimation problem may be computationally intensive. A very large sample is usually required in order to get a reliable estimate of density. The goodness of the estimate depends on the dimension of the random variable. Heuristically, the higher the dimension, the larger the sample required to provide adequate representation of the sample space[1], p194.

Observing data from various perspectives often involves transformations such as projections onto multiple lower-dimensional spaces. Features of interest in data include clusters of observation and relationships among variables that allow a reduction in the dimension of the data.

A new theory is to use "encapsulate" (compressing) as an instructional method in data analysis. An ideal structure to comprehend data easily is that it is not only used in compression stored data but also used in unwrap compression and almost back to original information, the wavelets is un-optimization, e.g., in images, signal, and the field of data, actually when it shows and compresses curve boundary in images. This will require us to need new representation system in order to compress better. Chapter 5 summarized some of the fundamental measuring methods, on this basis, chapter 10 specifically discussed again the methods of identification of structure. Higher-dimensional data have some surprising and counterintuitive properties, and the author discussed some of the interesting characteristics of higher dimensions.

Chapter 11 discussed asymmetric relationship among variables. For such problems, the objective is often to estimate or forecast the given dataset or predictive variables, or to identify the type to which objects observed belong. The approach is to use a given datasets to develop a model or a set of rules that can be applied to new data. Because of the number of possible existing forms or because of the recursive partitioning of the data used in selecting a model, the statistical modeling is considered possible computationally intensive. In computational statistics, the emphasis is on building a model rather than just estimating the parameters in the model. Parametric estimation of course plays an important role in building models.

In Chapters 10 and 11, the author discussed also the development of the clustering and classification methods in various disciplines. The author held this opinion, i.e., by means of a model to describe a generation mechanism for data: using the model to simulate artificial data, and then using the model to analyze data; examining the artificial data for conformity to our expectations or to some available real data. It helps us understand the role of the individual component of the model: its functional form, the parameters, and the nature of the stochastic component.

In the research literature Monte Carlo methods are widely used to evaluate properties of statistical methods. Appendix A addresses some of the considerations that apply to this kind of study. It is emphasized that the study uses an experiment, and the principles of scientific experimentation to be observed. Appendix B describes some of the software and programming issues that may be relevant to conducting a Monte Carlo study[1], Preface.

Above we give a brief summary of the basic content of the book. Just as the author mentioned in Preface, still many contents relating to computational statistics have not been mentioned

in this book. Details of many chapters and sections of this book which are capable of being dispensed with or done without were omitted. A few details are provided by his students in class, or else, some of details were worked out by students themselves. This way may be more effective. Some important topics such as FFTs and wavelets are only mentioned in this book. Other topics, such as the bootstrap, classification methods, and model-building, are discussed only in an introductory manner, the first is most obvious of all. If somebody is going to engage in thorough discussions for these topics, then a greater number of new books will be published followed by the book Elements of Computational Statistics. His goal has been to introduce a number of topics and devote an appropriate proportion of pages for each. He makes a series references list so that his students might carry out notes for the class successfully and more widely; this includes computational statistics with other computational sciences.

In the book, many exercises are needed to be processed by computer because in some cases routine calculations or experiments on simulated data are to be performed. The exercises range from the trivial or merely mechanical to the very challenging though the author has not attempted to indicate specifically. Some of the Monte Carlo studies suggested in the exercises were arranged for innovative research and publications.

He stated further that the text covers more material than can reasonably be included in a one-semester course. However, it will be an effective way to use it step by step at the beginning and proceed sequentially. For students with more background in statistics, Chapter I can be skipped. The book can be served as text for two courses in computational statistics if more emphasis is placed on the student projects and/or on numerical computations.

In most classes the author has been teaching computational statistics, giving exercises A.3 in Appendix A (see the original page 348, the reprint page 348) as a term project. "It is to replicate and extend a Monte Carlo study reported in some recent journal articles. Each student picks out some material to use. The statistical methods studied in the article must be ones that the students understand, but that is the only requirement as to the area of statistics addressed in the article. Teacher has varied the way in which the project is carried out, but it usually involves more than one student working together. A simple way is for each student to referee another student's first version (due midway through the term) and to provide a presentation. Each student is both an author and a referee. In another variation, the students are required to work in pairs. One student selects the article and designs and performs the Monte Carlo experiment, and the other one writes the article, in which the main content is description and analysis of the Monte Carlo experiment."[1], Preface

This is really a "teaching methods of investigative learning", and also a convincing model which teaches content associated closely with its organizational form (i.e., combinatory pedagogies of combining class, group and individual person). It fully embodies teaching principle of learning while teaching, and "teaching, research in step with discovery", which will promote the growth of students' intelligence (if you are interested in such information, refer to the paper. Li Zhuyu. To Promote the Combinatory Pedagogics, and Completely Increase the Student's Character in statistics. Education on Statistics, 2006, 04:4-5 (in Chinese)).

When talking about the software resource which this book uses, the author emphasizes:

The first, what software systems a person needs to use depends on the kinds of problems addressed and what systems are available. In this book, it is not intended to teach any software system. Though we cannot presume one is competent with any particular system, software system are not explained, however, examples from various systems, primarily S-Plus, are provided. Most of the code fragments will also work in R.

Secondly, some exercises suggest or require a specific software system. In some cases, the required software can be obtained from either statlib or netlib (see the Bibliography). The online help system should provide sufficient information about the software system required. As with most aspects of computer usage, a spirit of experimentation and adventure makes the effort easier and more rewarding.

Finally the author mentioned specially two seminars, which reminded him of Friday afternoon seminars on computational statistics, in particular, during the visit of Cliff Sutton, for enjoyable discussions on statistical methodology, both during the seminar and in post-seminar visits to local pubs. For the past few years, he has enjoyed pleasant Tuesday lunches with Ryszard Michalski, and he is sure some of their discussions have affected parts of this book. He is convinced that these discussions have exerted an influence for writing this book. Hence one can see the importance of the "seminars" in promoting achievements in scientific research.

Material relating to courses he teaches in computational sciences is available over the World Wide Web at the URL, http://www.science.gmu.edu/

Notes on this book, including errata, are available at

http://www.science.gmu.edu/∼jgentle/cmstbk/

From what we mentioned above, we can see clearly that James E. Gentle the author of this book is not only highly skilled in sphere of learning, but also hitherto unreached by anybody else in statistical teaching research and scientific methodology.

3. Comparison of "Elements of Computational Statistics" with those of other books and its influence on statistics

According to the above introduction as have elucidated in the foregoing, we have already sketched the outline of the subject of computational statistics.

Computational statistics is a specialized distinct field of statistics, it must be built on the basis of mathematics theories and statistical methods. A lot of computing methods among them are built on such traditional statistics fields as mathematical statistics, linear model, sampling survey, and time series, etc., which deal with data mining, numerical computation methods of statistical analysis (statistical computing), and bootstrap methods, and Monte Carlo Methods (MCMC theories), data simulation, discovering technique of data structure, and data storage with enormous capacity and processing technology, and involving significant research projects which build statistical models and statistical computing methods boasting "computer science" characteristics (for example curve fitting). It is generally much more hands-on exploring statis-

tical computing methods. The computationally intensive method sorts through those enormous, nonlinear, and multidimensional data and turns up interesting and useful knowledge and decision; an important part is establishment and evaluation of models among them.

As we have known, many statistical computing methods have been developed in the application field relating closely to it. Some of statisticians with masterly skill or technique in specific application of statistics have made important contributions to this field, but this does not rule out scientists of other subject areas in statistics application and theoretical research. The basic tasks of statistical research are to develop or invent new tools to allow it for use as the frontiers of science and technology.

Since the 1950s, the statisticians, mostly from the United States, have made some outstanding achievements in this field, whose representative is C.R.Rao from Pennsylvania State University. He has many of pioneering contribution in the multivariate statistical analysis, and has solved the problem on researching multi-dimensional data of complex structure. Another master is Princeton's J.Tukey, who is considered the father of modern data analysis. With help of computers, the most successful methods at the end of the 20th century, such as the bootstrap methods and proportional hazards model, collect more numerous or more complex data in expressive power, which give us the future with an exciting, challenging, and more fundamental results. Not many books can be found in our country, however. The following are broadly representative works from abroad, relating to computational statistics.

I. Principles of Data Mining, David Hand, Heikki Mannila and Padhraic Smyth, The MIT Press, 2001. The Chinese version, the translators, ZHANG Yinkui, etc. China Machine Press, 2003. This book is a scholarly treatise on data mining from a statistical point of view, where the definition of data mining technology is given as follows:

It is a science which refines the useful information from huge datasets or large databases. This book introduces the elementary knowledge of data mining, the component and algorithm of data mining.

II. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2001. Chinese version: China Electronic industry Press, 2003. The translators, FAN Min, etc. This book introduces such important concepts of new domains as data mining, Machine Learning, and Bioinformatics, and so on. It contains subjects on a wide range of neural networks, support vector machine, and classification tree and lifting, etc., which are most fully explained among others of its kind. The authors are all statistics professors of Stanford University, who have made outstanding contribution in this field.

III. An Introduction to Markov Chain Monte Carlo Methods and their Actuarial Applications. By David P.M.Scollnik. Department of Mathematics and Statistics, University of Calgary. Please log in: www.casact.org/pubs/proceed/proceed96/96114.pdf. The Courses of MCMC from Britain and the United States. MCMC have been chiefly used in physics in the past, but at present, settlement of a greater mass of bioinformatical problems are all related to the theories of MCMC.

IV. SIMULATION. By Sheldon M. Ross. Copyright © 2006. Fourth Edition. Elsevier Inc. From a book series entitled "Turing Mathematical Statistics". The reprint of P.R. China: Simulation (Statistical Simulation). China Posts & Telecom Press, Feb, 2007. This book introduces some of practical methods and techniques on statistical simulation, and analysis data, such as Bootstrap Methods, the variance reduction techniques, etc.. It introduces how to use statistical simulation to judge if the random model fits the actual data, MCMC theories and some of most up-to-date technology and thesis on statistical simulation.

V. Monte Carlo Strategies in Scientific Computing (Hardcover). by Jun S. Liu (Author). © 2001 Springer-Verlag New York, Inc. The reprint of P.R.China. Copyright © 2005-06-01 haotushu.com. The Monte Carlo method is a statistical spline method based on computer used to solve numerical problems. The author is a professor at Harvard University and the Changjiang Scholar of Beijing University, who won the top honor of American Statistical Association, the President Medal, U.S. This book expounded completely and systematically Monte Carlo methods and its fundamental rules, sequential Monte Carlo theories, the functioning of sequential Monte Carlo methods, Metrololis algorithm and other, Gibbs sampling, cluster sampling used in Ising model, average condition sampling, molecular dynamics and mixed Monte Carlo methods, multilevel sampling and optimization methods, the Monte Carlo methods based on population, Markov Chain and convergence, collection of selected theoretical thesis, the fundamental theories of mathematical statistics.

VI. Computational Statistics, Geof H. Givens, Jennifer A. Hoeting, John Wiley and Sons, Inc. February 2005. It casts light on the relations and distinction between modern statistical computing with computational statistics, and provides some of daily application software for us.

VII. An Introduction to Statistical Methods and Data Analysis. R.Lyman Ott, Micheal T.Longnecker, Duxbury Press, 2001. The Chinese version. This book is in two volumes, China Science Press, June 2003. The translators, ZHANG Zhongzhan, etc. The book brings together a large mass of figures and examples to help readers understand the basic thoughts and features to do with statistical method, understand the characteristic of data, summarize data's methods, and try to get the essence of statistical methods and models. The author attached importance to the role of statistics in solving practical problems. Reading the book does not need the other respects of advanced mathematics.

We list a series of reference books one by one because it is not hard to see that their authors did not consciously draw up a complete and perfect scientific system of computational statistics, although these books belong to the computational statistics category. The domains above are obviously overlapping one another. The common traits or characteristics are quantization, complexity, and modifiability of data. Computing (usually including drawing shapes) has become an important aspect of realizing every thought. James E. Gentle, author of Elements of Computational Statistics made it different. He not only conducted research on the core domain of computational statistics, such as applications of computationally intensive, the density, confirming data structures, model building and estimating, etc., but also made statement with such an addition in general statistical methods, which deals with data conversion, approximating func-

tions, and numerical technique of data optimization. The object of author is to try to establish the basic assumptions or principles of the subject, its concepts and general methods (including calculating devices, too). No matter what we say, these methods are all based on a basic theory framework; this theory has been developed in order to respond to the needs of other sentient fields.

The book introduced comprehensively and systematically mathematical theories of the latest development. It embraces major content on computational statistics. It has established the theory of this subject, and is real "elements" of computational statistics. We firmly believe that its publication will help greatly promote the development of the applications of methods of computational statistics in various fields of research. Though the book is a scholarly treatise, the method of presentation is entirely applicable to all other "computational" sciences. Not only will it definitely promote the further development of statistical science in such practical realm as Bioinformatics, Climatology, intrusion detection system (IDS) and finance; but also it is conducive to the development of computational physics, computational biology, and so on. Its content has covered almost all effective statistical computing methods, up to now.

Another prominent feature of the book is that there is a lot of useful information at the back of the book, including examples, exercises, and Web sites or email address relating to text, as well as ALGQL (algorithmic language) relating to computing method, and the appropriate applications software. It merges technicality, maneuverability, and exploratory into one whole, which play an essential role in the realm. As we have already said, this book is characteristic of both teaching methodology and research methodology. Of course, it is also a distinguishing feature or characteristic in American and European statistics teaching that they concern themselves with training professional knowledge and grasping of depth, breadth and skill of knowledge, and developing their ability to solve practical problems. They advocate the Combinatory Pedagogics on statistical professionals in light of the characteristics of training personnel, which will benefit students, be able to effectively develop a vision of the future, exploit one's self-potential, stay open-minded, never stop absorbing new knowledge and apply it to the areas of science and technology. Only in this way can we contribute to cultivating multi-functional talents who master statistical professional knowledge.

References

[1] GENTLE J E. *Elements of Computational Statistics* [M]. Springer-Verlag, New York, 2002.
[2] LINDSAY B G, KETTENRING J, SIEGMUND D D. *A report on the future of statistics* [J]. With comments. Statist. Sci., 2004, **19**(3): 387–413.
[3] David Hand, Heikki Mannila and Padhraic Smyth. Principles of Data Mining[M]. The MIT Press, 2001. The Chinese version, the translators, ZHANG Yinkui, etc. China Machine Press, 2003.
[4] GENTLE J E. *Numerical Linear Algebra for Applications in Statistics* [M]. Springer-Verlag, New York, 1998.
[5] GENTLE J E. *Random Number Generation and Monte Carlo Methods* [M]. Springer-Verlag, New York, 2003, 1998.
[6] LIU J S. *Monte Carlo Strategies in Scientific Computing* [M]. Springer-Verlag, New York, 2001.