International Statistical Review

International Statistical Review (2020), 88, 1, 252–264

Book Reviews

Editor: Ananda Sen

Random Circulant Matrices
Arup Bose and Koushik Saha
CRC Press, 2019, xix + 192 pages, \$174.95, hardcover
ISBN: 978-1-1383-5109-7

Readership: Graduate students and researchers interested in random matrices.

Chapters: 1. Circulants, 2. Symmetric and reverse circulant, 3. LSD: normal approximation, 4. LSD: dependent input, 5. Spectral radius: light tail, 6. Spectral radius: k-circulant, 7. Maximum of scaled eigenvalues: dependent input, 8. Poisson convergence, 9. Heavy-tailed input: LSD, 10. Heavy-tailed input: spectral radius, 11. Appendix.

A k-circulant $A_{k,n}$ ($1 = k = n-1$) is an n-square matrix whose each row is obtained from the previous by k circular shifts to the right. Its first row $(x_0, \ldots, x_{n-1})$ is called the input. Nothing is said about the x_i's in the definitions, but I guess that they are real numbers. The matrix $A_{k,n}$ is a circulant C_n if $k = 1$ and a reverse circulant RC_n if $k = n - 1$. (The term 'k-circulant' has also another meaning. It is often defined that a matrix is a k-circulant if it is obtained from C_n by multiplying with k its all entries below the main diagonal.) If the sequence $(x_1, \ldots, x_{n-1})$ is palindromic, then C_n is symmetric, denoted by SC_n. Also, RC_n is symmetric.

The empirical spectral distribution function (ESD) of an n-square matrix is obtained by putting a mass $1/n$ at its each eigenvalue. The limiting spectral distribution function (LSD) of a sequence of n-square matrices is the weak limit (if it exists) of the sequence of their ESDs. In the case of random matrices, this limit is understood in some probabilistic sense.

A description of some main points of Chapters 1–8 and 11 follows. The titles of Chapters 9 and 10 are informative enough.

Chapter 1. The matrices $A_{k,n}$, C_n, RC_n, and SC_n are defined. Call them 'circulant-type matrices'. A formula for the eigenvalues of $A_{k,n}$ is given. This formula is fundamental in what follows (except Chapter 2).

Chapter 2. The ESD and LSD are defined. A general technique to find the LSD of symmetric random matrices, based on the moment method, is introduced and applied to RC_n and SC_n when the input is i.i.d. (i.e. it consists of independent and identically distributed random variables).

Chapter 3. Using normal approximation, the LSD of C_n and, more generally, of $A_{k,n}$ (for certain pairs k,n) with independent input is obtained.

Chapter 4. The results of Chapters 2 and 3 are extended to the case when the input follows a stationary linear process.

Chapter 5. Using a sharper normal approximation, the limiting behaviour of the spectral radius of C_n, RC_n, and SC_n is studied when the input is i.i.d.

Chapter 6. The results of Chapter 5 are extended to $A_{k,n}$ when $n = k^g + 1$.

Chapter 7. The results of Chapters 5 and 6 are extended to the case when the input is as in Chapter 4.

Chapter 8. The joint behaviour of the eigenvalues of random circulant-type matrices with light-tailed i.i.d. input is studied via the point process approach.

Chapter 11. The formula for the eigenvalues of $A_{k,n}$ is proved. Certain notion and results in probability theory are summarised. Three auxiliary theorems, used repeatedly, are presented.

This book seems to be a useful ‘state-of-the-art’ on random circulant-type matrices. The authors are experts in this field, and several results are due to them. The reader is supposed to have much knowledge in advanced probability theory. Chapter 11 helps him or her if needed in this regard. The exercise sections, completing Chapters 1–10, are instructive, but they would have become more instructive if solutions or hints to the problems had been given.

What is the motivation of a book on random circulants?

The preface states: ‘Circulant matrices have been around for a long time and have been extensively used in many scientific areas. The classic book *Circulant Matrices* by P. Davis, has a wealth of information on these matrices. New research on, and applications of, these matrices are continually appearing everywhere’. This is repeated in the introduction, and some applications are mentioned. For example (p. xv), the periodogram of a sequence is a function of the eigenvalues of a suitable circulant. But all these are applications of non-random circulants.

As far as applications of random circulants (in fact, of random matrices) are concerned, the authors only say (p. xvi) that the behaviour of the eigenvalues of such matrices with large dimension ‘has attracted considerable interest in physics, mathematics, statistics, wireless communication and other branches of sciences’. Certain important books on random matrices do not appear in the bibliography.

In the beginning of Chapter 1, the applications, already mentioned in the introduction, are repeated. In the beginning of Chapters 2–10, no applications are mentioned. Neither did I find them elsewhere.

A book on random circulants that would become a classic (as Davis' book is on non-random circulants) is, regardless of the merits of the book under review, yet to be written. Such a book would contain applications and would also provide a comprehensive survey that ties the theory of random circulant-type matrices to the more general theory of patterned random matrices and, still more generally, to the theory of random matrices.

Jorma K. Merikoski: *jorma.merikoski@tuni.fi*
Faculty of Information Technology and Communication Sciences
FI-33014 Tampere University
Finland

The Road to Quality Control
Homer M. Sarasohn
Translated by N. I. Fisher and Y. Tanaka from the original Japanese text by Kagaku Shinko Sha
John Wiley & Sons, 2019, 160 pages, £89.95, hardcover
ISBN: 978-1-1195-1493-0

Readership: Researchers and practitioners in the area of quality control.

I have been involved in the field of quality and reliability engineering for more than 40 years. I have taught courses at both undergraduate and graduate levels, taught many short courses for a wide range of industries, conducted research, authored and coauthored books and papers in this field and was a consultant for many companies. Yet, to my astonishment, I was not aware of Homer M. Sarasohn's work in quality until I received a copy of this book. The book is a historical account of the subject of *quality control* by one of the key individuals who had worked through the Industry Branch of the American occupation army's Civil Communications Section (CCS) to improve the quality of the Japanese manufacturer of the radio and telephone communications. His work with Charles Protzman (both industrial engineers) has started the transformation of the quality of the Japanese products from quite basic to sophisticated and cutting-edge like what we observe in modern times, especially in the field of electronics. The initial efforts in quality control in Japan began under Sarasohn. His persistent teaching of management and control of industry paved the way to the then 'modern' statistical quality control based on Shewhart's book and related material. He authored the book *The Road to Quality Control* in Japanese in 1950 and used its material to teach Japanese managers the principles of quality control, sampling and quality inspection methods. The book under review is a translation of this Japanese version. The book as such does not serve as a text for teaching quality control course since its technical coverage is brief and is confined only to Chapters II to V. Of course, the tools of the modern quality control have changed dramatically over the years due to the advances in sensing, computation power and materials processing. The contribution and influence of Shewhart and Sarasohn's work in the early development in quality are undeniable, nonetheless. The seeds of the quality control in Japan were planted before Deming's arrival. This excellent book presents the principles of statistical quality control alongside a detailed history of evolution of the quality of Japanese products. I recommend this book as a supplement to those working in and teaching the subject of quality control.

Elsayed Elsayed: *elsayed@soe.rutgers.edu*
Department of Industrial & Systems Engineering, School of Engineering
Rutgers, The State University of New Jersey
Piscataway, NJ 08854, USA

Introduction to Probability: Models and Applications
N. Balakrishnan, Markos V. Koutras and Konstadinos G. Politis
John Wiley & Sons, 2020, xiii + 608 pages, $140.00, hardcover
ISBN: 978-1-1181-2334-8

Readership: Teachers and students of a first course in probability.

Chapters: 1. The Concept of Probability, 2. Finite Sample Spaces – Combinatorial Methods, 3. Conditional Probability – Independent Events, 4. Discrete Random Variables and Distributions, 5. Some Important Discrete Distributions, 6. Continuous Random Variables, 7. Some Important Continuous Distributions. Appendices: A. Sums and Products, B. Distribution Function of the Standard Normal Distribution, C. Simulation, D. Discrete and Continuous Distributions.

Each chapter
… contains a section 'Computational Exercises' (some are examples).

… ends in a section 'Applications' (actually 'Application'). They are 1. System Reliability, 2. Estimation of Population Size: Capture-Recapture Method, 3. Diagnostic and Screening Tests, 4. Decision Making Under Uncertainty, 5. Overbooking, 6. Profit Maximization, 7. Transforming Data: The Lognormal Distribution.

… begins by introducing some pioneer of probability who contributed to the topic discussed in the chapter. They are 1. Kolmogorov, 2. Pascal, 3. Bayes, 4. Markov, 5. Bernoulli, 6. Laplace, 7. Gauss.

… includes an abundance of exercises. Each section ends to exercises, mostly classified into two groups: A (routine) and B (more advanced).
… contains a section titled 'Self-assessment Exercises' (true-false and multiple choice questions) and a section containing 'Review Problems'.

The first author has much experience in writing textbooks. The second author has written in Greek a book on probability, which is the predecessor of the current book. So the goal the authors put forth in the preface certainly seems reachable: 'It is our sincere hope that instructors find this textbook to be easy-to-use for teaching an introductory course on probability, while the students find the book to be user-friendly with easy and logical explanations, plethora of examples, and numerous exercises (including computational ones) that they could practice with!'

A *plethora* of examples and exercises indeed exists (including many good ones). In my opinion, this is the main merit of the book. But the exercises have no solutions, even no answers or hints. The purpose of adopting this strategy is unclear. If the authors think that there is no space left for them, they could have cut down on the material in favour of some worked out solutions and hints to selected problems. If the authors plan to publish a solution manual separately, this plan should have been mentioned in the preface.

This book is indeed *user-friendly.* It gives *easy and logical explanations* to basic topics. For students of an introductory course, this may suffice so the book definitely helps them to have success. Of course, the reader encounters proofs and definitions that are more complex and advanced in nature. It seems to me that the authors have managed quite well with most of them, too.

A few exceptions appear, however, of which I give two examples. I enjoyed reading about the definition of probability (relative frequency, axiomatic and classical) in Chapters 1 and 2 but did not feel the same in reading about the definition of a continuous random variable in Section 6.1. Although absolute continuity is beyond the scope of the present book, I think that 'this is absolute continuity—let's forget it' is not enough. A (very intuitive) brief discussion on absolute continuity (and also on the Lebesgue integral) would have increased the value of the book.

I also enjoyed reading about the normal distribution in Section 7.2 but excluding the (very intuitive) proof of

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}$$

on pp. 503–504. The authors speak about the proof of $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}\theta\,\mathrm{d}r$, which, however, is not an exact mathematical formula and therefore has no proof. Two figures, where a circle is divided into $\mathrm{d}x\,\mathrm{d}y$'s by horizontal and vertical lines in the first, and into $r\,\mathrm{d}\theta\,\mathrm{d}r$ by concentric circles and their radii in the second, would have enabled easy visualization of the change of variables. This reviewer finds the book *easy-to-use* and looks forward to the appearance of its second volume (informed in the preface) for the second course in probability.

Jorma K. Merikoski: *jorma.merikoski@tuni.fi*
Faculty of Information Technology and Communication Sciences
FI-33014 Tampere University
Finland

Causal Inference in Statistics: A Primer
Judea Pearl, Madelyn Glymour and Nicholas P. Jewell
John Wiley & Sons, 2019, 156 pages, $46.75, paperback
ISBN: 978-1-1191-8684-7

Readership: Graduate students and researchers interested in causal inference.

This book, initially originating from course notes, covers the basics of causal inference in statistics. Often classical statistical methods fail to uncover the intrinsic mechanisms that lead to the data, staying on the side of shallow somewhat descriptive interpretations. Especially in observational studies, the design of the experiment and the statistical model built to make inferences from it do not match the set of mechanistic questions the researcher would like to answer. When this is the case, the researcher can often formulate a different (ideal) design of the experiment and a different mechanistic model supporting it that would be capable of answering the mechanistic questions. Rather than declare the current experiment useless, we may ask ‘Is it possible to make inferences from the current data that would relate to the meaningful parameters of the mechanistic model?’. Over the last two decades we have seen an explosion of interest in such questions handled by the causal inference methodology. The book is meant to be an introduction to this theory. The field is very young, so even a primer book needs to refer to fairly recent literature to explain the concepts and the methods of causal inference. An introductory book in this field is a challenge as many concepts are debated, and the classic historical core of the theory may still be fluid. However, the book admirably navigates this challenge and does help a novice in the field understand its principles and concepts. This is made possible by keeping the mathematical methods confined to basic probability and by always motivating and illustrating the methods by real-life examples. I especially enjoyed a collision of the immediate intuition of a novice trying to make sense of an example with the real picture that starts to emerge as the example is peeled deeper using causal concepts.

The book starts with Chapter 1 that motivates the rest of the book by introducing the Simpson’s paradox illustrating the phenomenon of confounding brought to the extreme. It is followed by a description of the basic probability and statistics tools that are used in the rest of the book. An introduction to statistical dependence follows with reference to their applications in the context of different models. Basic model structures are expressed through equations as well as through graphs. The book explains how these representations are linked and how they can be recognised based on the data at hand.

Chapter 2 presents very useful graphical tools that summarise the key ingredients and structure of the causal models. Well-defined graphical elements such as forks, chains, colliders, and the idea of d-separation allow us to dissect the structure of a causal model relevant to dependencies without having to provide a full probabilistic quantitative specification of the model.

Chapter 3 brings these tools to use to ascertain the causal effects of interventions. A concept of graph surgery is introduced to transform the current model for the data into a model of the ideal experiment that allows for clear unconfounded assessment of the causal effect of the intervention. Adjustments are worked out that allow us to use the statistics based on the current model to predict the relevant quantities (structural parameters) in the causal model. A detailed application for linear models is presented.

Chapter 4 introduces a general approach of counterfactuals. In some sense, this approach generalises the methods presented in the book earlier. This chapter is somewhat more advanced conceptually, and some of the facts are presented without proof.

While maintaining mathematical rigour, the book skillfully avoids complex mathematics. This makes it suitable for statisticians who are looking to start using causal modelling in their work or to epidemiologists who are comfortable with basic probability and statistics tools. However, this does not mean that the book is easy reading. The flow of new concepts is pretty dense and concise for somebody naive to causal inference. However, overall, the book is an outstanding introduction to an exciting and very useful subject.

Alexander Tsodikov: *tsodikov@umich.edu*
Department of Biostatistics
University of Michigan
Ann Arbor, MI 48109, USA

High-dimensional Statistics: A Non-asymptotic Viewpoint
Martin J. Wainwright
Cambridge University Press, 2019, xvii + 552 pages, £57.99, hardback
ISBN: 978-1-1084-9802-9

Readership: Statistics/machine learning graduate students and researchers.

This is an excellent book. It provides a lucid, accessible and in-depth treatment of non-asymptotic high-dimensional statistical theory, which is critical as the underpinning of modern statistics and machine learning. It succeeds brilliantly in providing a self-contained overview of high-dimensional statistics, suitable for use in formal courses or for self-study by graduate-level students or researchers. The treatment is outstandingly clear and engaging, and the production is first-rate. It will quickly become essential reading and the key reference text in the field.
Conventional, classical statistics, as developed in the early 1900s, is founded on an asymptotic regime in which the dimension p of the parameter in the statistical model remains fixed as the sample size n grows to infinity. Standard laws of large numbers and the central limit theorem then furnish a general suite of inferential techniques, typically based on the asymptotic consistency, normality and efficiency of the maximum likelihood estimator. Such inferential techniques, which served as the bedrock of statistical analysis for decades, have been extended, from the 1980s onwards, by the development of refined, likelihood-based and bootstrap methods of distributional approximation: some of the key elements of this substantial theory of 'higher-order asymptotics' are described in the brief review article Young (2009). Typical focus in classical theory concerns a parametric model $F(y;\theta)$ indexed by a $p-$dimensional parameter $\theta = (\psi, \lambda)$, where ψ is a scalar interest parameter and λ is a $(p-1)-$dimensional nuisance parameter. Inference on ψ is based on a random sample Y of size n from $F(y;\theta)$, and for instance, it is required to construct a confidence interval $I_\alpha(Y)$ for ψ, of nominal coverage $1-\alpha$. Bootstrap or analytic approximation is made for the sampling distribution of a 'pivot' $T(Y,\theta)$, such as the signed square root of the likelihood ratio statistic or some modification thereof. This estimated sampling distribution is then used to construct an accurate confidence set $I_\alpha(Y)$, with the property, valid assuming only correctness of the model $F(y;\theta)$, that

$$Pr_\theta\{\psi \in I_\alpha(Y)\} = 1-\alpha + O(n^{-r}),$$

for quantifiable r, typically $r = 1$ or $r = 3/2$. Though the error term $O(n^{-r})$ will typically depend on unknown quantities, so such a result is an asymptotic statement, the operational interpretation is immediate: as $n \to \infty$, the confidence set yields exactly the nominal coverage $1-\alpha$ under repeated sampling. In cases where λ is of low dimension, it is a rule of thumb that if $r = 3/2$, the confidence set will give essentially exact coverage for modest sample size, say $n = 10, 20$, though there are no guarantees from the theory of the magnitude of error for any finite n.

But, the data sets which arise in many areas of modern science and engineering generally have parameter dimension p of the same order, and often exceeding, sample size n, and for such problems classical statistical theory may fail to provide useful estimation or prediction, or simply break down completely. While investigations have established that accurate inference may often be obtained from higher-order asymptotics methodology in circumstances where the parameter dimension p is relatively large compared to available sample size n (see evidence contained in Barndorff-Nielsen & Cox, 1994, for instance), there has been relatively little systematic direct theoretical examination of what Wainwright terms 'high-dimensional asymptotics', where the pair (n, p) are taken to infinity simultaneously, in such a way that some scaling function of (n, p), and possibly other problem parameters, remains fixed or converges to some finite limit. Instead, focus in modern statistical theory has emphasised non-asymptotic results in high-dimensional problems. In this theory, the pair (n, p) as well as other problem parameters are viewed as fixed, and high-probability statements, say about the error of a parameter estimator, are made as a function of them. As its title suggests, non-asymptotic, theoretical results of this type are the focus of this book. Chapter 1 gives a beautiful overview, illustrating through key examples involving linear discriminant analysis, covariance estimation and nonparametric regression, what can go wrong with classical statistics in high dimensions, and motivating persuasively the non-asymptotic viewpoint. The kind of non-asymptotic theory developed in the book, founded on obtaining bounds on the tails of a random quantity, or concentration inequalities which provide bounds on how a random variable deviates from some value, such as its mean, has its main value in being able to be used to predict some aspects of high-dimensional asymptotic phenomena, such as limiting forms, as (n, p) grow, of error probabilities in the linear discriminant problem. Further, the scaling functions that emerge in a non-asymptotic analysis can suggest the appropriate high-dimensional asymptotic analysis to perform in order to reveal relevant limiting distributional behaviour. The non-asymptotic analysis is typically not used, or immediately available, to provide precise statements about behaviour, say, of an estimator for a given, finite (n, p).

As an example of a non-asymptotic result, we take illustration from Wainwright (Example 7.14), concerning the classical linear model

$$Y = X\beta + \epsilon,$$

where the design matrix $X \in \mathbb{R}^{n\times p}$ is deterministic and the noise vector ϵ has independent elements, identically distributed as $N(0, \sigma^2)$. Assume that X satisfies some (checkable, since X is fixed) 'restricted eigenvalue' condition and that it has $\max_{j=1,\ldots,p} \frac{\|X_j\|_2}{\sqrt{n}} \leq C$, where X_j is the jth column of X. Suppose further that the vector β is supported on a subset $S \subseteq \{1, 2, \ldots, p\}$ with $|S| = s$, so that β is sparse, with s non-zero elements. If we define the Lasso estimator $\hat{\beta}$ of β as the minimiser of

$$\frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1,$$

we have that, for constant K,

$$\|\hat{\beta} - \beta\|_2 \leq K\sqrt{s}\left\{\sqrt{\frac{2\log p}{n}} + \delta\right\},$$

with probability at least $1 - 2e^{-n\delta^2/2}$, for any $\delta > 0$, if we set $\lambda = 2C\sigma\left(\sqrt{\frac{2\log p}{n}} + \delta\right)$. Then we see that, provided K, which is defined in terms of C, σ^2 and the eigenvalue condition, stays fixed as (n, p) increase, the Lasso estimator $\hat{\beta}$ is consistent, as long as $\log p$ is dominated by n, if the size of the true β, as determined by s, remains fixed. Such a result predicts rather little, though, about the finite sample behaviour of the estimator $\hat{\beta}$: the bound on $\|\hat{\beta} - \beta\|_2$ depends, inter alia, on the unknown error variance σ^2 and the unknown true β, through its sparsity level s. As is the case for classical statistical theory developed for the fixed p regime, non-asymptotic results of this kind only offer precise guarantees, therefore, asymptotically. But, the beauty of the theory of high-dimensional statistics as described by Wainwright is precisely that non-asymptotic results can yield strong operational support to practical methods of data analysis, such as the Lasso described above. For instance, while the Lasso estimator is not universally optimal, it comes close to mimicking the properties of an oracle estimator (which knows the true state of nature) in many sparse settings, in estimation of the mean $E(Y)$ of Y when X is given. The Lasso predicts $E(Y)$ almost as well as an oracle which knows which of the elements of β are non-zero.

Chapter 1 of Wainwright's book gives also a very clear account of what enables statistics in the high-dimensional setting. What saves us is the reasonable expectation that high-dimensional data are actually endowed with some form of low-dimensional structure, typically some form of sparsity, which might crudely be expressed as meaning that only s of the p parameters of the model are non-zero or non-negligible, where s is much smaller than p. Much of high-dimensional statistics therefore involves constructing models of intrinsically high-dimensional phenomena, but where the models incorporate some implicit form of low-dimensional structure, which can be successfully revealed from sample data. The introductory chapter is, in its own right, a tour de force, but sets the scene for a marvelous account of the mathematics and methodology of all the main elements of high-dimensional statistical theory. Beautifully sign-posted, the subsequent material of the book really divides into two types. Foundational material on tools and techniques, such as concentration inequalities, concentration of measure, uniform laws of large numbers, notions of covering and packing, reproducing kernel Hilbert spaces and techniques for obtaining minimax lower bounds is elegantly described. Of mathematical interest in its own right, this material is directed here to derive theory that is broadly applicable in high-dimensional statistics. Crucially for those interested in statistical practice, the book also provides a thorough account of the models and estimators used in data analysis. The text includes a series of chapters each focused on a particular class of statistical estimation problems, including covariance estimation, the sparse linear model, principal component analysis, estimators based on decomposable regularisers, estimation of low-rank matrices, graphical models and least squares estimation in a nonparametric setting: the principal aim is to shed light on the theoretical guarantees that are offered by widely used estimation techniques. Careful attention is paid throughout to computational considerations, such as the gains that are made by deriving from an initial otherwise NP-hard optimization problem a convex criterion that can be optimised efficiently, while ensuring that the resulting statistical procedure is almost as good as that initially considered. Ideas and formalities are illustrated throughout the text with carefully chosen examples, primarily of a theoretical, rather than data analytic, nature. A minor quibble of this reviewer is the lack of a glossary of notation: someone who is not immersed in the area might wonder what precisely $\asymp$ and $\lesssim$ mean.

In summary, this is an authoritative, scholarly and highly useful summary of high-dimensional statistics. It is dense and not for the faint of heart, though the author offers excellent routemaps through the material. As a text, it stands natural comparison with Giroud (2015) and Bühlmann and van de Geer (2011), but to suggest any preference here would be invidious.

References

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. London: CRC Press.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York: Springer.

Giroud, C. (2015). *Introduction to High-Dimensional Statistics*. London: CRC Press.

Young, G.A. (2009). Routes to higher-order accuracy in parametric inference. *Aust. N.Z. J. Stat.*, **51**, 115–126.

G. Alastair Young: *alastair.young@imperial.ac.uk*
Department of Mathematics
Imperial College London
180 Queen's Gate, London SW7 2AZ, UK

Matrix Differential Calculus with Applications in Statistics and Econometrics, 3rd Edition
Jan R. Magnus and Heinz Neudecker
John Wiley & Sons, 2019, 504 pages, $115, hardcover; $92.99 ebook
ISBN: 978-1-1195-4116-5

Readership: Graduate students, practitioners and researchers interested in calculus, matrices, optimisation problems, statistical models and/or their applications.

The book has become the standard reference on matrix differential calculus since first published in 1988. This is the third edition of the book, with seven parts devoted to the theory and application of matrix differential calculus. Matrix differential calculus was pioneered and promoted by the authors and developed on with many other contributors in the last several decades, especially the late Professor Heinz Neudecker. As stated in the preface to this new edition, 'Heinz Neudecker must be regarded as its founding father'.

The first part is an introductory review on matrices. Chapter 1 includes basic definitions and results such as matrix addition and multiplication, rank, inverse, determinant, trace, eigenvalue and eigenvector, Schur's decomposition theorem and singular-value decomposition. Chapters 2 and 3 cover selected, useful results for later chapters, involving Kronecker product, Hadamard product, vec operator, Moore-Penrose inverse, commutation matrix and duplication matrix.

The second part is on differentials in theory. Chapter 4 introduces the mathematical preliminaries. Chapters 5–7 cover the uniqueness and existence of the differential, the first and second identification theorems, the mean-value theorem, remarks on notation, the chain rule, partial derivatives, the Hessian matrix, unconstrained and constrained optimisation, necessary and sufficient conditions for a local minimum without or under constraints, three versions of the implicit function theorem, among others which are fairly fundamental.

The third part explores differentials in practice. It principally shows how to implement the theory and use important differentials with definitions, rules, propositions, theorems with proofs and provided examples. The first and second identification theorems incorporate Jacobians and Hessian matrices for scalar, vector and matrix functions of scalar, vector and matrix variables. The differentials with Jacobian and Hessian matrices are presented extensively, including matrix trace, determinant and eigenvalue functions, the eigenvector function, as well as Kronecker and Hadamard products.

The fourth part is a collection of some key inequalities, including the Cauchy–Schwarz inequality, its matrix versions (matrix determinant and trace versions), Fischer's min–max theorem, Poincarè separation theorem, Hadamard's inequality, Karamata's inequality, Hölder's inequality, Minkowski's inequality, Schlömilch's inequality, the least squares and generalised least squares with their restricted counterparts.

Part five discusses several topics in linear model. Chapter 12 overviews the statistical preliminaries, cumulative distribution function, joint density function, expectation, variance and covariance, independence of random variables and the normal distribution. Chapters 13 and 14 present the most important topics in the theory of linear models, namely, the Gauss–Markov theorem, the methods of ordinary, generalised and restricted least squares, Aitken's theorem, estimable functions, linear constraints, singular variance matrix, best quadratic unbiased and invariant estimators of the variance, best linear unbiased predictors, and local sensitivity of the posterior mean and precision.

Part six applies the maximum likelihood methods to various models and issues and especially the full and limited information maximum likelihood methods used in simultaneous equations. The topics in one-, two- and multi-mode component analysis, principal components, factor analysis, canonical correlation, correspondence analysis and linear discriminant analysis are widely used in psychometrics and multivariate statistics. Chapter 17 provides additional examples applying the various results obtained, including those on Kronecker and Hadamard products, eigenvalues and eigenvectors, matrix determinant and trice optimisation problems and matrix calculus.

Part seven offers a summary of the essentials of matrix calculus. It has a basic introduction followed by a coverage on the useful definitions and results for differentials, chain rule for differentials, vector calculus, matrix calculus, Kronecker product and vec operator, commutation and duplication matrices, least squares and maximum likelihood, and so forth, with examples.

This edition continues to be successful in maintaining those distinguishing features in the previous editions. The book is well written and lucid. It is self-contained. It builds on good notation, a unique approach of using differentials (rather than derivatives) and the vectorial operation amidst the chain rule procedure. It covers the theory and a wide range of topics with applications in not only statistics, econometrics and psychometrics but also related areas in biosciences, social and behavioural sciences and many others. The proofs of many key theorems are both rigourous and easy to follow. The examples and exercises are extremely helpful, with newly added exercises in this edition linking to research problems and requests from readers. The bibliographical notes are informative for understanding the relevant literature and especially beneficial for those readers interested in further study or research.

This edition has updated materials and references throughout the book, especially in sections involving matrix functions, complex differentiation, Jacobians of transformations, differentiation of eigenvalues and eigenvectors and Hessian matrices for scalar functions. It has made two new sections on correspondence analysis and linear discriminant analysis, respectively. Obviously, a new chapter at the end of the book presents a collection of the essential definitions and results for matrix differential calculus. It is devoted to a practical yet self-contained and handy summary of the whole book and can be studied by readers independently from the rest of the text, that is, without going into theoretical details. Some pertinent extensions are absent, such as the Khatri–Rao product (related to Sections 2.2 and 3.6), the matrix versions of the Cauchy–Schwarz and Kantorovich inequalities (in the Löwner partial ordering) and their applications in efficiency comparisons (related to Sections 11.2, 11.3, 11.29–11.32, 13.18 and 13.19), the local sensitivity of generalised least squares (related to Sections 14.15 and 14.16) and the pseudo likelihood estimation (related to Sections 15.2 and 15.3).

As proven by various applications in different areas and as noted in the preface to the third edition, the technique of matrix calculus through differentials 'is still a remarkably powerful tool'. Recently, matrix differential calculus has been found to be useful (as we should expect of such a powerful tool) in data analytic areas like deep learning and even ecological areas like matrix population models. More concrete uses in newer areas could be explored in the next edition.

In summary, this book is highly recommended as an advanced text or a reference handbook to anyone who uses or is interested in matrices, calculus, optimisation problems, statistical models and/or their applications in individual studies.

Shuangzhe Liu: *shuangzhe.liu@canberra.edu.au*
Faculty of Science and Technology
University of Canberra,
Canberra ACT 2601, Australia

Model-based Clustering and Classification for Data Science
Charles Bouveyron, Gilles Celeux, T. Brendan Murphy and Adrian E. Raftery
Cambridge University Press, 2019, 427 + xvii pages, £59.99, hardcover
ISBN: 978-1-1084-9420-5

Readership: Graduate students and researchers in statistics.

Table of contents. Chapter 1: Introduction, Chapter 2: Model-based Clustering: Basic Ideas, Chapter 3: Dealing with Difficulties, Chapter 4: Model-based Classification, Chapter 5: Semi-supervised Clustering and Classification, Chapter 6: Discrete Data Clustering, Chapter 7: Variable Selection, Chapter 8: High-dimensional Data Chapter, 9: Non-Gaussian Model-based

Clustering, Chapter 10: Network Data, Chapter 11: Model-based Clustering with Covariates, Chapter 12: Other Topics

Model-based clustering and classification are important modern topics. This advanced text explains the underlying concepts clearly and is strong on theory. In an ideal world, this excellent theoretical basis would be accompanied by insightful data analyses to illustrate the methods, informative graphics to complement the analyses and supporting R code. Unfortunately, the data analyses in the current text are disappointing and are occasionally flawed; the graphics are weak with little use of modern display, often too small, with poor choice of colour and symbols, and the relevant accompanying text two or more pages away. The R codes are often incomplete and sometime fail to run, a surprising and unfortunate feature in this age of reproducible research.

If you are interested in the model-based approach (and you should be, if you are interested in clustering and classification), then this book is for you. I congratulate the authors on the theoretical aspects of their book, it's a fine achievement. The application component of the book does not match up to the theoretical counterpart. Details of the issues the reviewer has found have been sent to the authors so that they are able to rectify the errors in the next edition. Once that is accomplished, this book will make a fine contribution in the area of clustering and classification.

Antony Unwin: *unwin@math.uni-augsburg.de*
Institut für Mathematik
Universität Augsburg,
D-86135 Augsburg, Germany