# Book Reviews





Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Book Reviews
DOI: 10.1198/jasa.2009.br0909

that improves the ability of practicing statisticians and biostatisticians to formulate, select, and use models is worth its weight in gold. Konishi and Kitagawa have written such a book.

Joseph CAVANAUGH
*The University of Iowa*

## REFERENCES

Burnham, K. P., and Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), New York: Springer.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, Dordrecht: D. Reidel Publishing Company.

## Intermediate Probability: A Computational Approach.

Marc S. PAOLELLA. Chichester, U.K.: Wiley, 2007. ISBN 978-0-470-02637-3. xiii + 415 pp. $160.00 (H).

This book is the sequel to *Fundamental Probability: A Computational Approach* by the same author (Paolella 2006). The first volume contains a thorough coverage of the topics typically associated with a first course in probability for upper-level undergraduates or first-year graduate students. The sequel is a natural extension of the topics contained in its predecessor. For the most part, *Intermediate Probability* can be used without access to *Fundamental Probability*, with the exception of an occasional reference. A review of *Fundamental Probability* was provided by Harvill (2008).

Like its predecessor, *Intermediate Probability* provides a thorough highly accessible coverage of topics in probability. Naturally the topics are more advanced than those in *Fundamental Probability* and are appropriate for graduate students or highly advanced undergraduates with a sound background in mathematics and introductory probability. Both volumes use MatLab software to provide computational illustrations of probabilistic concepts. The incorporation of a computational aspect enhances the accessibility and students' learning experience through clear illustrations of applicability or of the strengths or weaknesses of the underlying theory. For readers who prefer R to MatLab, translations of MatLab programs into the R language by Sergey Goriatchev are available from the books' companion websites.

The presentation style in *Intermediate Probability* is the same as that in *Fundamental Probability*. The book is divided into three main sections. Section I, "Sums of Random Variables," contains three chapters; Section II, "Asymptotics and Other Approximations," contains three chapters; and the Section III, "More Flexible and Advanced Random Variables," contains four chapters. At the end of every chapter except Chapter 8 is a good set of exercises ranging in difficulty from low (no stars) to high (two stars). Some of the exercises are theoretical, some are computational, and some are a blend of theory and computation. Solution manuals for instructors are available from the publisher. Excellent reference lists are provided throughout the book.

Section I, covering sums of of random variables, begins by emphasizing the use of characteristic functions for calculating the density and distribution of random variables. To aid students without a background in complex variables, an introduction to complex numbers, Fourier series, and discrete Fourier transforms is provided. Chapter 2 discusses the use of characteristic functions for determining the distribution of sums of random variables. It begins with the typical examples (e.g., independent binomial, independent gamma), then tackles cases in which identifying a resulting characteristic function is not so simple. To evaluate the probability density (or mass) functions for these examples, the chapter illustrates the analytical or numerical use of inversion formulas introduced in the sections on complex variables. Chapter 3 contains a thorough exposition on the multivariate normal distribution.

The topics of Section II, Asymptotics and Other Approximations, are introduced through a rigorous but intuitive description of convergence concepts. I particularly liked Chapter 4, which begins by presenting many important inequalities for random variables, results for convergence of sequences of sets, and a brief discussion of convergence of sequences of random variables. The next four sections contain nice discussions of convergence in probability, almost sure convergence, convergence in $r$th mean, and convergence in distribution. Most of the discourse in these sections is technical, with some intuitive explanations and no computational illustrations. This changes in the final section of the chapter with the presentation of the central limit theorem, where an intuitive discussion leads into a more rigorous presentation. The section also contains numerous graphical illustrations of simulated and analytical comparisons of the normal approximation to the true underlying distribution or to empirical distributions of simulated means in the form of a histogram. Overall, this chapter on convergence concepts is very well done.

Chapter 5 presents a discussion of saddlepoint approximations for estimating the distribution of a statistic when the normal approximation breaks down. The chapter includes a good set of important examples for when breakdowns occur and provides a nice list of references containing proofs of the results stated in the following sections. It also gives a detailed exposition of the univariate case, covers some aspects of the multivariate case, and concludes with an introduction to the hypergeometric functions that are needed in several exercises in subsequent chapters.

Chapter 6, on order statistics, contains the standard fare, beginning with a discussion of the distribution theory of order statistics for independent and identically distributed samples in both the univariate and multivariate cases, followed by a presentation of the distribution of the sample range and midrange. Scattered throughout the chapter are some nice computational illustrations. The chapter ends with a nice set of additional examples and a brief discussion on distribution theory for dependent samples.

Section III, "More Flexible and Advanced Random Variables," begins with Chapter 7, which starts with an overview of how to classify some of the hundreds of distributions available via the ideas of nesting, generalizing, and asymmetric extensions. Mixture distributions are discussed in detail, leading to the derivation of the variance-gamma distribution. The chapter is replete with worked examples. The applicability of some examples is immediately evident, whereas other examples are more theoretical. Throughout the chapter, the discussion is enhanced with useful illustrative graphs. Example 7.19 (pp. 260–264) is a good illustration of how the author addresses many dimensions of a problem. He begins with the derivation and illustration of the LaPlace distribution arising from the mixture relationship $X|V \sim \mathrm{N}(\mu, v)$ and $V \sim \text{exponential}(\lambda)$, resulting in $X \sim$ LaPlace. Via moment-generating functions, he provides the details for working through these simple relationships and continues on to more general densities. He provides illustrative graphs and MatLab code to generate iid location-0, scale-1 LaPlace random variates via the mixture relationship.

Chapter 8 is a basic introduction to the stable Paretian distribution, with emphasis on its computation, basic properties, and uses. The chapter ends with the generalized central limit theorem. This brief but nicely done chapter is the only one that does not contain a set of exercises.

Following through with the material from Chapters 7 and 8, Chapter 9 is dedicated to the generalized inverse Gaussian and generalized hyperbolic distributions and their connections. Finally, Chapter 10 provides a detailed account of the singly and doubly noncentral $F$, $t$, and beta distributions. For each, several methods for the exact calculation of the distribution are provided, as is a discussion of approximate methods, most notably saddlepoint approximations, once again underscoring their value. The appendixes contain an extremely useful list of tables, including those for abbreviations, special functions, general notation, generating functions, and inversion formulas, along with distributional naming conventions, distributional subsets, Student $t$ generalizations, noncentral distributions, relationships among major distributions, and mixture relationships.

A slight drawback is the presence of typographical errors throughout the text. This is not surprising, especially given the book's length and level of detail. An errata sheet is available on the book's website. The errors should not deter one from purchasing the text. Another drawback is the sheer volume of material presented in the (two) volume(s). In the preface to *Fundamental Probability* and at the end of the Preface in *Intermediate Probability*, Paolella freely admits that there is too much material to cover in lectures; however, especially in a two-semester sequence, the material is not so overwhelming that students cannot read the material not specifically covered in lectures. As in *Fundamental Probability*, here Paolella's approach continues to be rich, interesting, informative, and entertaining. Students will be engaged by the practical application of material that can be daunting.

The book has a companion website at *http://www.wiley.com/go/intermediate/*. Along with a brief description of the book, the website provides links to two archived compressed (zip) files, one file containing the programs in the book written in R and the other containing the programs in MatLab. It also provides a link to an Adobe Acrobat (PDF) file containing an errata sheet. Finally, there is a link to a form to complete to gain permission to access a set of lecture notes that complement the text and a complete solutions manual for the exercises at the end of each chapter.

I thoroughly enjoyed *Intermediate Probability*. I was so thrilled with it (along with *Fundamental Probability*) that I have shared it with some of my colleagues. They have called it a "gold mine" of problems and resources, and describing it as "amazing." In our collective opinion, these two volumes are a resource that will enrich any library and are well worth the cost. For an instructor with a clear direction not daunted by selecting from an abundance of topics, the book would make an excellent text for a course in probability and random variables, especially when used in tandem with its predecessor. At the least, it should be listed as an excellent reference for any such course. I highly recommend it.

Jane L. HARVILL
*Baylor University*

## REFERENCES

Harvill, J. L. (2008), Review of *Fundamental Probability: A Computational Approach*, by M. S. Paolella, *The American Statistician*, 62, 179–180.

Paolella, M. S. (2006), *Fundamental Probability: A Computational Approach*, Chichester, U.K.: Wiley.

### A Matrix Handbook for Statisticians.

George A. F. SEBER. Hoboken, NJ: Wiley, 2008. ISBN 978-0-471-74869-4. xix + 559 pp. $123.95 (H).

When I was a graduate student, the advanced course in linear models used Seber's (1977) text. Although I found the text enjoyable and have referred to it time and again, it is his Appendix A (innocuously entitled "Some Matrix Algebra") that I have consulted most frequently. As a de facto handbook for matrix results, it served admirably. The current text is well over 500 pages and contains a wide range of results. (The complete list of topics is extensive; those interested can view the table of contents at the publisher's website, *http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471748692,descCd-tableOfContents.html*.) It is certain to supplant the 1977 text on my bookshelf when I need a matrix result.

The book accomplishes what it purports to do and nothing more. Simply put, it is a handbook of matrix results (albeit with a bent toward those that statisticians might find most useful). As such, there is little in the way of exposition, which is not to say there is no prose or unifying presentation. Rather, as a concession to space, proofs are forgone, with references to more traditional texts provided. Taken in isolation, this lack of proofs may be disappointing, but considered in the larger context of a handbook, it is a necessary decision. Pragmatically, and speaking from personal experience, it is often of interest to know immediately whether a conjecture is true, with the proof deferred to later. This text is ideal for such questions. In keeping with this approach, the exposition is concise, with attention focused on consistency and precision of notation.

Those looking for a handbook of matrix results will find this text an unequivocal success, whereas those seeking a more illustrative book for a graduate course in regression or linear algebra will find it ill-suited to that purpose. However, depending on the reader's familiarity with matrix manipulations and statistical theory, it may be an excellent choice to supplement a text.

David H. ANNIS
*Wachovia Bank*

## REFERENCE

Seber, G. A. F. (1977), *Linear Regression Analysis*, New York: Wiley.

### The Minimum Description Length Principle.

Peter D. GRÜNWALD. Cambridge, MA: MIT Press, 2007. ISBN 978-0-262-07281-6. xxxii + 703 pp. $47.00 (H).

The minimum description length (MDL) principle is a relatively recent method for model selection or, more generally, for inductive inference. The insight behind this principle is that statistical inference is equivalent to finding regularities in the data, which can be used to compress the data more efficiently, leading to a shorter code length. First developed by Rissanen in a series of articles starting in 1978, the MDL principle is widely used, especially in engineering, perhaps because of its roots in information theory.

Despite its popularity, however, MDL remains a mystery to those outside a small research community. MDL is difficult for beginners to learn. Some review articles have been published, but these are either too advanced, requiring a background in both information theory and measure-theoretic probability, or too introductory, providing little guidance for further exploration and in-depth reading. In the past 10 years, numerous advances and breakthroughs have been made in the field of MDL; however, these are spread over various journals and conference proceedings and have not received sufficient attention in the scientific community. This is why many practical applications of MDL are still based on the "old" method, and why the common misunderstanding that "MDL is just BIC" persists. All of these considerations have generated a call for a comprehensive and accessible review on MDL and the related theory of stochastic complexity and universal coding. That is precisely what *The Minimum Description Length Principle* is all about.

The book comprises four parts. Part I provides a basic introduction to MDL and a brief review of prerequisites from statistics and information theory. For readers from statistics without previous exposure to information theory, the key concepts that they need to understand are that a code is equivalent to a probability distribution, and that short code lengths correspond to high probabilities. Thus the MDL principle is closely related to the maximum likelihood principle. However, in the context of model selection, MDL naturally includes a penalty for model complexity. Suppose that I want to transmit some data to a friend, and suppose that we agree in advance to use a certain model (i.e., a parametric family of distributions $\{p_\theta, \theta \in \Theta\}$) to encode the data. I might be tempted to use the code corresponding to the distribution $p_{\hat{\theta}}$ (recall the equivalence between a code and a probability distribution), where $\hat{\theta}$ is the maximum likelihood estimate, because this gives rise to the shortest code length. But my friend would not know which distribution to use to decode the data, because he would not know the value of $\hat{\theta}$ without seeing the data. Therefore, I have to transmit $\hat{\theta}$ to my friend first, which adds an extra term in the overall description length of the data. This extra term—the code length used for describing $\hat{\theta}$—usually increases with respect to the complexity of the model (apparently it will cost more to describe a three-dimensional parameter than a one-dimensional parameter), and thus it plays the role of a complexity penalty.

The specific form of MDL introduced in Part I is not optimal, however. The current refined form of MDL is based on the idea of universal coding, which receives detailed coverage in Part II. The four main types of universal codes introduced in Part II are the two-part, the Bayes, the prequential plug-in, and the normalized maximum likelihood. I appreciate the author's effort to make this highly technical topic accessible to readers from a diverse range of areas; for each type of universal code, he starts with the simple finite case and gradually builds up the theory to general parametric and even nonparametric cases.

Part III is dedicated to refined MDL for model selection, density estimation, and prediction, as well as related theoretical results, such as consistency and rate of convergence. Part IV reviews the statistical theory of exponential families, providing the necessary background knowledge for some of the proofs. This part can be read separately as an appendix.

Who might be the potential readers of this book? I would definitely recommend the book to anyone wanting to know about MDL. No other book has its depth and detail. The book also would be a wonderful reference for any researcher in this field, providing excellent coverage of the last 30 years of MDL. Although the book is not intended to be a textbook (it has no exercises), I would use it for a high-level topics course on model selection or information theory.

I have some minor criticisms of the book, which could be easily addressed in a second edition. First, the length of roughly 700 pages might scare away some potential readers, especially first-timers. It could be shortened. For example, definitions for some elementary concepts, such as closed intervals, open intervals, and one-to-one functions in Section 2, could be omitted; at the beginning of some chapters is a paragraph explaining the contents of that chapter, which because it is roughly the same as the section titles, I would suggest omitting. The author also might consider breaking the book into two volumes or moving some material online. Second, in the Preface the author clearly expresses his intention to make the book accessible to researchers from biology, experimental psychology, and other applied sciences; however, to make the book appealing to researchers from applied sciences, more real examples and applications are needed, especially in Part I, which serves as an introduction to a general audience. Third, I personally think that the MDL approach for clustering is an