Frontiers in Probability and the Statistical Sciences

György Terdik

Multivariate Statistical Methods

Going Beyond the Linear

Springer

Frontiers in Probability and the Statistical Sciences

The “Frontiers” is a new series of books (edited volumes and monographs) in probability and statistics designed to capture exciting trends in current research as they develop. Some emphasis will be given on novel methodologies that may have interdisciplinary applications in scientific fields such as biology, ecology, economics, environmental sciences, finance, genetics, material sciences, medicine, omics studies and public health.

More information about this series at http://www.springer.com/series/11957

György Terdik

Multivariate Statistical Methods

Going Beyond the Linear

– Vector-Moments and Vector-Cumulants
– Nonlinear Statistics of Normal Multivariates
– Testing Skewness and Kurtosis

Springer

György Terdik
Faculty of Informatics, Dept. of IT
University of Debrecen
Debrecen, Hungary

ISSN 2624-9987 ISSN 2624-9995 (electronic)
Frontiers in Probability and the Statistical Sciences
ISBN 978-3-030-81391-8 ISBN 978-3-030-81392-5 (eBook)
https://doi.org/10.1007/978-3-030-81392-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To
Judit,
and
Bori and Bálint and Ábel

Foreword

As the author rightly points out, much of what we call linear modeling and inference in multivariate statistics is closely tied to the multivariate Gaussian distribution and allied topics. Indeed, the classic text by Professor C.R. Rao is aptly titled Linear Statistical Inference (1973, Wiley) as is one of the books recently co-authored by me which is titled Linear Models and Regression with R (2020, World Scientific Press). Non-Gaussian multivariate distributions are characterized by nonlinearity, where the cumulants play a prominent role. This monograph is a very deep and substantial piece of work analyzing the cumulants and related statistical measures like the skewness and kurtosis for the most commonly discussed non-Gaussian multivariate distributions. This is also a great place to learn about and polish mathematical prerequisites like multilinear algebra and tensor products for higher-order partial derivatives of vector-valued functions. A large number of exercises at the end of each chapter is an added bonus for those who want to use this as a text for an advanced course. The extensive set of nearly two hundred references, which range from a manuscript that goes back to 1931 to the most recent papers appearing in 2020 and 2021, demonstrates the breadth and depth of coverage.

This monograph will be a standard reference for those who are interested in multivariate distribution theory that goes beyond the multidimensional normal, and will remain so for many years to come!

Santa Barbara, CA, USA
February 25, 2021

S. Rao Jammalamadaka
Distinguished Professor,
Dept. of Statistics and Applied Probability
University of California, Santa Barbara

Preface

Linear theory in statistics is closely connected to the normal distribution. One of the main reasons for this is that the best predictor is linear when the random variables are jointly normal, and an important characterization of the normal distribution is that all cumulants of order higher than 2 are zero. The linear theory is pretty much well defined, unlike the nonlinear theory, which can proceed in many different ways. A careful study of the cumulants is a necessary and typical part of nonlinear statistics. Such a study of cumulants for multivariate distributions is made complicated by the index notations. One solution to this problem is the usage of tensor analysis. In this book we offer an alternate method, which we believe is simpler to follow. The higher-order cumulants with the same degree for a multivariate vector can be collected together and kept as a vector. To be able to do so, we introduce a particular differential operator on a multivariate function, called the T-derivative, and use it to obtain cumulants and provide results which are somewhat analogous to well-known results in the univariate case. We demonstrate this technique through the characterization of several multivariate distributions via their cumulants and by extending the discussion to statistical inference for multivariate skewness and kurtosis.

The book is organized as follows:

Chapter 1 introduces some basic notions and methods which are used in permutations, matrix theory, multilinear algebra, and set partitions.

Chapter 2 deals with the method of tensor products for the higher-order partial derivatives of vector-valued functions. Faà di Bruno's formula is also discussed here.

Chapter 3 concerns the basic theory of T-moments and T-cumulants. Besides connections between cumulants and moments, there are results which relate the cumulants of products to products of cumulants, conditional cumulants, etc.

Chapter 4 covers the elementary theory of the nonlinear Hilbert space of Gaussian variates. Multivariate vector-valued Hermite polynomials are introduced with their basic properties and their moments and cumulants are derived.

Chapters 5 and 6 deal with various applications of the material of the previous chapters. In particular, Chap. 5 deals with the cumulants of multivariate skew-distributions, including skew-normal, skew-spherical, skew-t, scale mixtures of skew normal, skew-normal-Cauchy, and Laplace distributions. The final Chap. 6 is devoted to statistical inference for multivariate skewness and kurtosis. We study the estimators of higher-order cumulant vectors and their asymptotic normality. Explicit formulae for the asymptotic covariances of estimated skewness and kurtosis are given.

The results presented in the book can be followed rather easily because specific references to the formulae that have been used previously are given step by step. Nevertheless, a close reading of the core Chaps. 3 and 4 is required for a proper understanding.

The exercises are of service to practising the methods; some of them are alluded to as simple facts in the text but left unproved. Selected solutions, some useful formulae, as well as basic notations are listed in the Appendix of the book under the titles Solutions, Formulae, and Notations, respectively.

The book is designed both as text for advanced graduate-level courses in multivariate statistical analysis and as a reference book for research workers interested in this area. The R package MultiStatM is available for applications of the theory considered in this book.

During the summer of 2001, when Professor Tata Subba Rao visited the University of Debrecen, we discussed the idea and agreed that statisticians need to pay more attention to cumulants in general. We proposed to write a book on the elementary treatment of cumulants, but unfortunately that work was never continued. In the present book, we use some of the ideas and even some material on cumulants for scalar-valued variates that we discussed at that time.

It is a great pleasure to thank Professor Sreenivasa Rao Jammalamadaka at the University of California, Santa Barbara, and Professor Emanuele Taufer at the University of Trento for numerous discussions and joint work concerning multivariate skewness and kurtosis.

Special thanks go to Professor Sreenivasa Rao Jammalamadaka for a careful reading of the manuscript and useful suggestions which improved the presentation of the book.

Debrecen, Hungary
January 2021

György Terdik

Contents