MULTIVARIATE STATISTICS
MULTIVARIATE STATISTICS

High-Dimensional and Large-Sample Approximations

YASUNORI FUJIKOSHI
Chuo University
Tokyo, Japan

VLADIMIR V. ULYANOV
Moscow State University
Moscow, Russia

RYOICHI SHIMIZU
Institute of Statistical Mathematics
Tokyo, Japan
Contents

Preface ................................................................................................................. xiii
Glossary of Notation and Abbreviations ....................................................... xvii
1 Multivariate Normal and Related Distributions ................................. 1
  1.1 Random Vectors ..................................................................................... 1
     1.1.1 Mean Vector and Covariance Matrix ........................................... 1
     1.1.2 Characteristic Function and Distribution .................................... 5
  1.2 Multivariate Normal Distribution ......................................................... 6
     1.2.1 Bivariate Normal Distribution ..................................................... 6
     1.2.2 Definition ...................................................................................... 8
     1.2.3 Some Properties .......................................................................... 10
  1.3 Spherical and Elliptical Distributions .................................................... 15
  1.4 Multivariate Cumulants ........................................................................ 19
     Problems .................................................................................................. 24
2 Wishart Distribution ..................................................................................... 29
  2.1 Definition ................................................................................................ 29
  2.2 Some Basic Properties .......................................................................... 32
  2.3 Functions of Wishart Matrices ............................................................... 36
  2.4 Cochran’s Theorem ............................................................................... 39
  2.5 Asymptotic Distributions ...................................................................... 40
     Problems .................................................................................................. 43
3 Hotelling’s $T^2$ and Lambda Statistics ...................................................... 47
  3.1 Hotelling’s $T^2$ and Lambda Statistics ................................................ 47
     3.1.1 Distribution of the $T^2$ Statistic ................................................... 47
     3.1.2 Decomposition of $T^2$ and $D^2$ ..................................................... 49
  3.2 Lambda Statistic .................................................................................... 53
     3.2.1 Motivation of the Lambda Statistic .............................................. 53
     3.2.2 Distribution of the Lambda Statistic ............................................ 55
  3.3 Test for Additional Information ............................................................. 58
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1 Decomposition of the Lambda Statistic</td>
</tr>
<tr>
<td>Problems</td>
</tr>
<tr>
<td><strong>4 Correlation Coefficients</strong></td>
</tr>
<tr>
<td>4.1 Ordinary Correlation Coefficients</td>
</tr>
<tr>
<td>4.1.1 Population Correlation</td>
</tr>
<tr>
<td>4.1.2 Sample Correlation</td>
</tr>
<tr>
<td>4.2 Multiple Correlation Coefficient</td>
</tr>
<tr>
<td>4.2.1 Population Multiple Correlation</td>
</tr>
<tr>
<td>4.2.2 Sample Multiple Correlation</td>
</tr>
<tr>
<td>4.3 Partial Correlation</td>
</tr>
<tr>
<td>4.3.1 Population Partial Correlation</td>
</tr>
<tr>
<td>4.3.2 Sample Partial Correlation</td>
</tr>
<tr>
<td>4.3.3 Covariance Selection Model</td>
</tr>
<tr>
<td>Problems</td>
</tr>
<tr>
<td><strong>5 Asymptotic Expansions for Multivariate Basic Statistics</strong></td>
</tr>
<tr>
<td>5.1 Edgeworth Expansion and its Validity</td>
</tr>
<tr>
<td>5.2 Sample Mean Vector and Covariance Matrix</td>
</tr>
<tr>
<td>5.3 $T^2$ Statistic</td>
</tr>
<tr>
<td>5.3.1 Outlines of Two Methods</td>
</tr>
<tr>
<td>5.3.2 Multivariate t-Statistic</td>
</tr>
<tr>
<td>5.3.3 Asymptotic Expansions</td>
</tr>
<tr>
<td>5.4 Statistics with a Class of Moments</td>
</tr>
<tr>
<td>5.4.1 Large-Sample Expansions</td>
</tr>
<tr>
<td>5.4.2 High-Dimensional Expansions</td>
</tr>
<tr>
<td>5.5 Perturbation Method</td>
</tr>
<tr>
<td>5.6 Cornish–Fisher Expansions</td>
</tr>
<tr>
<td>5.6.1 Expansion Formulas</td>
</tr>
<tr>
<td>5.6.2 Validity of Cornish–Fisher Expansions</td>
</tr>
<tr>
<td>5.7 Transformations for Improved Approximations</td>
</tr>
<tr>
<td>5.8 Bootstrap Approximations</td>
</tr>
<tr>
<td>5.9 High-Dimensional Approximations</td>
</tr>
<tr>
<td>5.9.1 Limiting Spectral Distribution</td>
</tr>
<tr>
<td>5.9.2 Central Limit Theorem</td>
</tr>
<tr>
<td>5.9.3 Martingale Limit Theorem</td>
</tr>
<tr>
<td>5.9.4 Geometric Representation</td>
</tr>
<tr>
<td>Problems</td>
</tr>
<tr>
<td><strong>6 MANOVA Models</strong></td>
</tr>
<tr>
<td>6.1 Multivariate One-Way Analysis of Variance</td>
</tr>
<tr>
<td>6.2 Multivariate Two-Way Analysis of Variance</td>
</tr>
<tr>
<td>6.3 MANOVA Tests</td>
</tr>
</tbody>
</table>
Contents

6.3.1 Test Criteria ......................................................... 157
6.3.2 Large-Sample Approximations .................................. 158
6.3.3 Comparison of Powers ........................................... 159
6.3.4 High-Dimensional Approximations .............................. 161
6.4 Approximations Under Nonnormality .............................. 163
   6.4.1 Asymptotic Expansions ....................................... 163
   6.4.2 Bootstrap Tests ................................................. 167
6.5 Distributions of Characteristic Roots ............................ 170
   6.5.1 Exact Distributions ........................................... 170
   6.5.2 Large-Sample Case ............................................ 172
   6.5.3 High-Dimensional Case ....................................... 174
6.6 Tests for Dimensionality .......................................... 176
   6.6.1 Three Test Criteria .......................................... 176
   6.6.2 Large-Sample and High-Dimensional Asymptotics ............ 178
6.7 High-Dimensional Tests ............................................ 181
   Problems ................................................................ 183

7 Multivariate Regression ............................................. 187
  7.1 Multivariate Linear Regression Model ........................... 187
  7.2 Statistical Inference .............................................. 189
  7.3 Selection of Variables ............................................ 194
    7.3.1 Stepwise Procedure ......................................... 194
    7.3.2 Cp Criterion .................................................. 196
    7.3.3 AIC Criterion ................................................ 200
    7.3.4 Numerical Example ......................................... 202
  7.4 Principal Component Regression .................................. 203
  7.5 Selection of Response Variables .................................. 206
  7.6 General Linear Hypotheses and Confidence Intervals ........... 209
  7.7 Penalized Regression Models ..................................... 213
   Problems ................................................................ 213

8 Classical and High-Dimensional Tests for Covariance Matrices 219
  8.1 Specified Covariance Matrix ...................................... 219
    8.1.1 Likelihood Ratio Test and Moments ......................... 219
    8.1.2 Asymptotic Expansions ...................................... 221
    8.1.3 High-Dimensional Tests ..................................... 225
  8.2 Sphericity .......................................................... 227
    8.2.1 Likelihood Ratio Tests and Moments ....................... 227
    8.2.2 Asymptotic Expansions ...................................... 228
    8.2.3 High-Dimensional Tests ..................................... 230
  8.3 Intraclass Covariance Structure .................................. 231
16.1 Linear Discriminant Function ............................................ 467
  16.1.1 Representation as a Location and Scale Mixture ............... 467
  16.1.2 Large-Sample Approximations ..................................... 472
  16.1.3 High-Dimensional Approximations .............................. 474
  16.1.4 Some Related Topics ............................................. 476
16.2 Profile Analysis ......................................................... 479
  16.2.1 Parallelism Model and MLE ..................................... 479
  16.2.2 Distributions of $\gamma$ ...................................... 481
  16.2.3 Confidence Interval for $\gamma$ ................................. 486
16.3 Estimators in the Growth Curve Model .................................. 487
  16.3.1 Error Bounds .................................................. 487
  16.3.2 Distribution of the Bilinear Form ............................. 488
16.4 Generalized Least Squares Estimators .................................... 490
  Problems ...................................................................... 492
Appendix ........................................................................ 495
A.1 Some Results on Matrices .................................................. 495
  A.1.1 Determinants and Inverse Matrices ............................. 495
  A.1.2 Characteristic Roots and Vectors ............................... 496
  A.1.3 Matrix Factorizations ............................................ 497
  A.1.4 Idempotent Matrices ............................................. 500
A.2 Inequalities and Max–Min Problems ...................................... 502
A.3 Jacobians of Transformations .............................................. 508
Bibliography .................................................................. 513
Index .......................................................................... 527
Many multivariate methods are based on large-sample approximations. These results can be found in the books in multivariate statistical analysis. See, for example, the books by Anderson (2003), Muirhead (1984) and Siotani et al. (1985). However, these approximations become increasingly inaccurate as dimension $p$ of observations increases while sample size $n$ remains fixed. On the other hand, in last years we encounter more and more problems in applications when $p$ is comparable with $n$ or even exceeds it. Some examples of high-dimensional data include curve data, spectra, images, and DNA micro-arrays.

Therefore, it becomes essential to revise the classical multivariate methods in order to make them useful in wide range of relations between $p$ and $n$ and to extend multivariate statistical theory in high-dimensional situations. One way of overcoming a weakness of classical large-sample approximations is to derive approximations under high-dimensional framework when $p/n \to c \in (0, 1)$ or $(0, \infty)$.

Another problem related to multivariate approximations concerns their errors. Most results supply the so-called order estimates only. However, such estimates do not give information on actual errors for given values $n, p$ and other parameters. Ideally, we wish to have computable error bounds, in addition to order estimates. It is made already for some multivariate statistics.

In multivariate methods it is important to reduce a set of original variables and canonical variables, so that we can make statistical inference more accurate and its interpretation more simple and effective. However, it is difficult to choose an appropriate subset of variables, or an appropriate number of canonical variables. For such problem, a model selection approach is developed, in addition to traditional testing methods or sequential procedures.

Our book is focusing on high-dimensional and large-sample approximations. At the same time we describe many basic multivariate methods and derive the exact distributional results related to the methods too. For many approximations, its detailed derivation will take a lot of space. Therefore, we give mainly their outlines. In order to solve the above-mentioned problems, we consider in the book

1. high-dimensional as well as large sample approximations for classical multivariate statistics,
2. approximations for high-dimensional statistics,
3. explicit error bounds for large-sample and high-dimensional approximations,
4. selection of variables by model selection approach,
(5) basic multivariate methods and related exact distributions

This book is designed as a reference book for researchers interested in multivariate statistical analysis. However, we believe that it will be useful for graduate level courses as well, since it contains in first twelve chapters many basic facts and methods from multivariate analysis.

Broadly speaking, chapters 1 ~ 12 deal with multivariate analysis focussing on (1), (2), (4) and (5). The last four chapters (Chapters 13 ~ 16) concern with explicit error bounds for some large-sample and high-dimensional approximations. Chapter 1 gives basic properties of multivariate normal distributions and elliptical distributions. Sample covariance matrix and various sums of squares and products matrices have Wishart distributions, when their underlying distributions are normal. In Chapter 2 we describe properties of Wishart distributions. In Chapter 3 the Hotelling $T^2$ and the Lambda statistics are treated. We also study the likelihood ratio test for additional information when several mean vectors are compared. Definitions, inferences, and sampling distributions of several correlations (except canonical correlations) are discussed in Chapter 4. Covariance selection model which is related to partial correlations is discussed as well.

In Chapter 5 we summarize some methods and theories on asymptotic expansion of multivariate statistics. High-dimensional as well as large-sample approximations are discussed. In this chapter the reader will find the topics on Edgeworth, Cornish-Fisher and bootstrap approximations, and their validities. MANOVA problems are discussed in Chapter 6. The distributions of MANOVA tests and characteristic roots are treated. Multivariate regression and linear models are discussed in Chapter 7. We give $C_p$ and AIC criteria for selection of the response variables as well as the explanatory variables. Classical and high-dimensional tests on covariance matrices are considered in Chapter 8.

In Chapter 9, discriminant analysis is studied. The concept of discriminant analysis is given, including a decision-theoretic approach and Fisher's method. Significance tests for discriminant functions and evaluation of probabilities of misclassifications are discussed. Further, the problems of selecting the canonical discriminant variables as well as the original variables are considered, based on model selection criteria. Principal component analysis and canonical correlation analysis are treated in Chapters 10 and 11, respectively. Some inferential problems on dimensionality are treated as the ones of selecting special types of covariance structures. Large-sample approximation is obtained for distributions of canonical correlations as special case of high-dimensional approximations. The growth curve model, which is a model for repeated measures data, is discussed in Chapter 12. Theoretically it can be considered as a multivariate linear model under a conditional set-up. Using this relation various inferential methods are derived, including the methods of defining a degree of polynomial in growth curve model.

Chapters 13 through 16 are concerned with explicit and computable error bounds for asymptotic approximations. We suggest a general approach to approximation of scale mixtures, including special cases of normal and chi-square mixtures and their multivariate extensions. In Chapter 14 we give the results on a location and scale mixture, the maximum of multivariate $t$- and $F$- variables, a mixture of $F$-distribution, and non-uniform error bounds. The applications of these basic results are discussed in Chapters 15 and 16. Error bounds are given in Chapter 15 for
large-sample approximations of Hotelling’s $T^2_0$ (Lawley and Hotelling criterion) and Lambda-statistics. In Chapter 16 we construct error bounds for large-sample and high-dimensional approximations of linear discriminant function. Furthermore, the estimators in profile analysis, growth curve analysis, and generalized linear model are treated.

We express our sincere thanks to Mr. Tetsuro Sakurai, Chuo University, Tokyo, for his kind help with numerical computations of some examples and the preparation of this book.

Our thanks also go to the authors, editors, and owners of copyrights for permission to reproduce the following materials: Table 7.3.1 (the chemical data introduced by Box and Youle, 1995 and examined by Rencher, 2002), Table 12.1.1 (the data on ramus height of 20 boys, taken from Elston and Grizzle, 1962), and Table 12.3.1 (the data of dental measurements on $n_1 = 11$ girls and $n_2 = 14$ boys, taken from Potthoff and Roy, 1964).

We believe and hope that the book will be useful in the future developments in multivariate analysis.

Yasunori Fujikoshi
Hiroshima, Moscow, Tokyo
November, 2009

Vladimir V. Ulyanov
Ryoichi Shimizu
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{R}^p$</td>
<td>$p$-dimensional Euclidean space ($\mathbb{R}^1 = \mathbb{R}$)</td>
</tr>
<tr>
<td>$a, a'$</td>
<td>column and row vectors, respectively</td>
</tr>
<tr>
<td>$a = (a_1, \ldots, a_p)'$</td>
<td>column vector with components $a_1, \ldots, a_p$</td>
</tr>
<tr>
<td>$1_p = (1, \ldots, 1)'$</td>
<td>column vector consisting of $p$ ones</td>
</tr>
<tr>
<td>$A: p \times q$</td>
<td>matrix with $p$ rows and $q$ columns</td>
</tr>
<tr>
<td>$(a_{ij})$</td>
<td>matrix with elements $a_{ij}$'s</td>
</tr>
<tr>
<td>$I_p$</td>
<td>unit matrix of order $p$</td>
</tr>
<tr>
<td>$A'$</td>
<td>transposed matrix of $A$</td>
</tr>
<tr>
<td>$\text{diag}(\theta_1, \ldots, \theta_p)$</td>
<td>diagonal matrix with diagonal elements $\theta_1, \ldots, \theta_p$</td>
</tr>
<tr>
<td>$\text{diag}(A_1, \ldots, A_k)$</td>
<td>block-diagonal matrix with elements $A_1, \ldots, A_k$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
</tr>
<tr>
<td>$\text{tr} A$</td>
<td>trace of a square matrix $A$</td>
</tr>
<tr>
<td>$A \otimes B$</td>
<td>direct product of matrices $A$ and $B$</td>
</tr>
<tr>
<td>$0$</td>
<td>zero matrix consisting of $0$s</td>
</tr>
<tr>
<td>$A &gt; B$</td>
<td>the matrix $A - B$ is positive definite</td>
</tr>
<tr>
<td>$A \geq B$</td>
<td>the matrix $A - B$ is positive semi-definite</td>
</tr>
<tr>
<td>d.f.</td>
<td>degrees of freedom</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>$X, Y, \ldots$</td>
<td>random variable (r.v.) — italic</td>
</tr>
<tr>
<td>$\mathbf{X}, \mathbf{Y}, \ldots$</td>
<td>random vector (r.vec.) — slanted boldface</td>
</tr>
<tr>
<td>$A, B, C, \ldots$</td>
<td>constant matrix — upright roman</td>
</tr>
<tr>
<td>$\mathbf{X}, \mathbf{Y}, \mathbf{S}, \ldots$</td>
<td>random matrix — wide boldface</td>
</tr>
<tr>
<td>$\text{vec}(\mathbf{X})$</td>
<td>$np$-vector $(\mathbf{X}'<em>{(1)}, \ldots, \mathbf{X}'</em>{(p)})'$ obtained from the matrix $\mathbf{X} = (\mathbf{X}<em>{(1)}, \ldots, \mathbf{X}</em>{(p)})$</td>
</tr>
<tr>
<td>$E(\mathbf{X}), \text{Var}(\mathbf{X})$</td>
<td>expectation and variance of random variable $\mathbf{X}$</td>
</tr>
<tr>
<td>$E(\mathbf{X}), \text{Var}(\mathbf{X})$</td>
<td>expectation and covariance matrix of random vector $\mathbf{X}$</td>
</tr>
<tr>
<td>$E(\mathbf{X}), \text{Var}(\mathbf{X})$</td>
<td>expectation and covariance matrix of random matrices $\mathbf{X}$</td>
</tr>
<tr>
<td>$\text{Cov}(\mathbf{X}, \mathbf{Y})$</td>
<td>covariance matrix of random matrices $\mathbf{X}$ and $\mathbf{Y}$</td>
</tr>
<tr>
<td>$\text{pdf}$</td>
<td>probability density function</td>
</tr>
<tr>
<td>$\text{cdf}$</td>
<td>cumulative distribution function</td>
</tr>
<tr>
<td>$N(\mu, \sigma^2)$</td>
<td>the normal distribution with mean $\mu$ and variance $\sigma^2$</td>
</tr>
<tr>
<td>$\varphi(x)$</td>
<td>pdf of the standard normal distribution $N(0, 1)$</td>
</tr>
<tr>
<td>$\Phi(x)$</td>
<td>cdf of the standard normal distribution $N(0, 1)$</td>
</tr>
</tbody>
</table>
G(σ, λ)  the gamma distribution with scale and shape parameters σ and λ
Be(p, q)  the beta distribution
t(n)    the t-distribution with n d.f.
χ²(p)   the chi-square distribution with p d.f.
χ²(p)   the chi-square distribution with p d.f.
χ²(p; τ²) the noncentral chi-square distribution with p d.f. and noncentrality parameter τ²
F(m, n) the F-distribution with (m, n) d.f.
F(m, n; τ²) the noncentral F-distribution with (m, n) d.f. and noncentrality parameter τ²
Λ_p(m, n) the distribution of Wilks' lambda criterion of dimension p with (m, n) d.f.
N_p(μ, Σ) p-variate normal distribution with mean vector μ and covariance matrix Σ. (N(μ, σ²) = N(μ, σ²))
W_p(Σ, n) Wishart distribution with n degrees of freedom and covariance matrix Σ
ch_i(A) the ith largest characteristic root of A
ch.r. characteristic root
LR likelihood ratio, likelihood ratio test
ML, MLE maximum likelihood, maximum likelihood estimator
PM probability of misdiscrimination
MANOVA multivariate analysis of variance
||x|| the Euclidean norm of the vector x = (x₁, ..., x_p)'
||X|| maximum absolute value of latent roots of matrix X
||f(x)||₁ the L₁-norm of real function f on R^p: \( \int_{R^p} |f(x)| \, dx \)