

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



WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2010 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Fujikoshi, Yasunori, 1942–

Multivariate statistics : high-dimensional and large-sample approximations / Yasunori

Fujikoshi, Vladimir V. Ulyanov, Ryoichi Shimizu.

p. cm. — (Wiley series in probability and statistics)

Includes bibliographical references and index.

ISBN 978-0-470-41169-8 (cloth)

1. Multivariate analysis. 2. Approximation theory. I. Ulyanov, Vladimir V., 1953– II.

Shimizu, Ryoichi, 1931– III. Title.

QA278.F84 2009

519.5'35—dc 22

2009017248

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

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

3.3.1	Decomposition of the Lambda Statistic	61
	Problems	64
4	Correlation Coefficients	69
4.1	Ordinary Correlation Coefficients	69
4.1.1	Population Correlation	69
4.1.2	Sample Correlation	71
4.2	Multiple Correlation Coefficient	75
4.2.1	Population Multiple Correlation	75
4.2.2	Sample Multiple Correlation	77
4.3	Partial Correlation	80
4.3.1	Population Partial Correlation	80
4.3.2	Sample Partial Correlation	82
4.3.3	Covariance Selection Model	83
	Problems	87
5	Asymptotic Expansions for Multivariate Basic Statistics	91
5.1	Edgeworth Expansion and its Validity	91
5.2	Sample Mean Vector and Covariance Matrix	98
5.3	T^2 Statistic	104
5.3.1	Outlines of Two Methods	104
5.3.2	Multivariate t -Statistic	107
5.3.3	Asymptotic Expansions	109
5.4	Statistics with a Class of Moments	111
5.4.1	Large-Sample Expansions	111
5.4.2	High-Dimensional Expansions	117
5.5	Perturbation Method	120
5.6	Cornish-Fisher Expansions	125
5.6.1	Expansion Formulas	125
5.6.2	Validity of Cornish-Fisher Expansions	129
5.7	Transformations for Improved Approximations	132
5.8	Bootstrap Approximations	135
5.9	High-Dimensional Approximations	138
5.9.1	Limiting Spectral Distribution	138
5.9.2	Central Limit Theorem	140
5.9.3	Martingale Limit Theorem	143
5.9.4	Geometric Representation	144
	Problems	145
6	MANOVA Models	149
6.1	Multivariate One-Way Analysis of Variance	149
6.2	Multivariate Two-Way Analysis of Variance	152
6.3	MANOVA Tests	157

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	C_p 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

8.3.1	Likelihood Ratio Tests and Moments	231
8.3.2	Asymptotic Expansions	233
8.3.3	Numerical Accuracy	235
8.4	Test for Independence	236
8.4.1	Likelihood Ratio Tests and Moments	236
8.4.2	Asymptotic Expansions	238
8.4.3	High-Dimensional Tests	239
8.5	Tests for Equality of Covariance Matrices	241
8.5.1	Likelihood Ratio Test and Moments	241
8.5.2	Asymptotic Expansions	243
8.5.3	High-Dimensional Tests	244
	Problems	245
9	Discriminant Analysis	249
9.1	Classification Rules for Known Distributions	249
9.2	Sample Classification Rules for Normal Populations	256
9.2.1	Two Normal Populations with $\Sigma_1 = \Sigma_2$	256
9.2.2	Case of Several Normal Populations	258
9.3	Probability of Misclassifications	258
9.3.1	W-Rule	259
9.3.2	Z-Rule	261
9.3.3	High-Dimensional Asymptotic Results	263
9.4	Canonical Discriminant Analysis	265
9.4.1	Canonical discriminant Method	265
9.4.2	Test for Additional Information	267
9.4.3	Selection of Variables	270
9.4.4	Estimation of Dimensionality	273
9.5	Regression Approach	276
9.6	High-Dimensional Approach	278
9.6.1	Penalized Discriminant Analysis	278
9.6.2	Other Approaches	278
	Problems	280
10	Principal Component Analysis	283
10.1	Definition of Principal Components	283
10.2	Optimality of Principal Components	286
10.3	Sample Principal Components	288
10.4	MLEs of the Characteristic Roots and Vectors	291
10.5	Distributions of the Characteristic Roots	292
10.5.1	Exact Distribution	293
10.5.2	Large-sample Case	294
10.5.3	High-dimensional Case	301

10.6	Model Selection Approach for Covariance Structures	302
10.6.1	General Approach	302
10.6.2	Models for Equality of the Smaller Roots	305
10.6.3	Selecting a Subset of Original Variables	306
10.7	Methods Related to Principal Components	308
10.7.1	Fixed-Effect Principal Component Model	308
10.7.2	Random-Effect Principal Components Model	310
	Problems	311
11	Canonical Correlation Analysis	317
11.1	Definition of Population Canonical Correlations and Variables	317
11.2	Sample Canonical Correlations	322
11.3	Distributions of Canonical Correlations	324
11.3.1	Distributional Reduction	324
11.3.2	Large-Sample Asymptotic Distributions	326
11.3.3	High-Dimensional Asymptotic Distributions	327
11.3.4	Fisher's z-Transformation	333
11.4	Inference for Dimensionality	335
11.4.1	Test of Dimensionality	335
11.4.2	Estimation of Dimensionality	337
11.5	Selection of Variables	338
11.5.1	Test for Redundancy	338
11.5.2	Selection of Variables	342
	Problems	345
12	Growth Curve Analysis	349
12.1	Growth Curve Model	349
12.2	Statistical Inference: One Group	352
12.2.1	Test for Adequacy	352
12.2.2	Estimation and Test	354
12.2.3	Confidence Intervals	357
12.3	Statistical Methods: Several Groups	359
12.4	Derivation of Statistical Inference	365
12.4.1	General Multivariate Linear Model	365
12.4.2	Estimation	366
12.4.3	LR Tests for General Linear Hypotheses	368
12.4.4	Confidence Intervals	369
12.5	Model Selection	370
12.5.1	AIC and CAIC	370
12.5.2	Derivation of CAIC	371
12.5.3	Extended Growth Curve Model	373

Problems	376
13 Approximation to the Scale-Mixed Distributions	379
13.1 Introduction	379
13.1.1 Simple Example: Student's t-Distribution	379
13.1.2 Improving the Approximation	381
13.2 Error Bounds evaluated in sup-Norm	384
13.2.1 General Theory	384
13.2.2 Scale-Mixed Normal	388
13.2.3 Scale-Mixed Gamma	390
13.3 Error Bounds evaluated in L_1 -Norm	395
13.3.1 Some Basic Results	395
13.3.2 Scale-Mixed Normal Density	397
13.3.3 Scale-Mixed Gamma Density	399
13.3.4 Scale-Mixed Chi-square Density	402
13.4 Multivariate Scale Mixtures	404
13.4.1 General Theory	404
13.4.2 Normal Case	410
13.4.3 Gamma Case	415
Problems	418
14 Approximation to Some Related Distributions	423
14.1 Location and Scale Mixtures	423
14.2 Maximum of Multivariate Variables	426
14.2.1 Distribution of the Maximum Component of a Multi- variate Variable	426
14.2.2 Multivariate t-Distribution	427
14.2.3 Multivariate F-Distribution	429
14.3 Scale Mixtures of the F -Distribution	430
14.4 Nonuniform Error Bounds	433
14.5 Method of Characteristic Functions	436
Problems	439
15 Error Bounds for Approximations of Multivariate Tests	441
15.1 Multivariate Scale Mixture and MANOVA Tests	441
15.2 Function of a Multivariate Scale Mixture	443
15.3 Hotelling's T_0^2 Statistic	445
15.4 Wilk's Lambda Distribution	448
15.4.1 Univariate Case	448
15.4.2 Multivariate Case	456
Problems	465
16 Error Bounds for Approximations to Some Other Statistics	467

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 $\hat{\gamma}$	481
16.2.3	Confidence Interval for γ	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