Monographs
on Statistics and
Applied Probability 89

Algebraic Statistics

Computational Commutative Algebra in Statistics

Giovanni Pistone
Eva Riccomagno
and Henry P. Wynn

CHAPMAN & HALL/CRC

Algebraic Statistics

Computational Commutative Algebra in Statistics

GIOVANNI PISTONE
EVA RICCOMAGNO
HENRY P. WYNN

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

1 Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2 Queues *D.R. Cox and W.L. Smith* (1961)
3 Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4 The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5 Population Genetics *W.J. Ewens* (1969)
6 Probability, Statistics and Time *M.S. Barlett* (1975)
7 Statistical Inference *S.D. Silvey* (1975)
8 The Analysis of Contingency Tables *B.S. Everitt* (1977)
9 Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10 Stochastic Abundance Models *S. Engen* (1978)
11 Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12 Point Processes *D.R. Cox and V. Isham* (1980)
13 Identification of Outliers *D.M. Hawkins* (1980)
14 Optimal Design *S.D. Silvey* (1980)
15 Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16 Classification *A.D. Gordon* (1981)
17 Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18 Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19 Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20 Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21 Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22 An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23 Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24 Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25 The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26 Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27 Regression Analysis with Applications *G.B. Wetherill* (1986)
28 Sequential Methods in Statistics, 3rd edition
G.B. Wetherill and K.D. Glazebrook (1986)
29 Tensor Methods in Statistics *P. McCullagh* (1987)
30 Transformation and Weighting in Regression
R.J. Carroll and D. Ruppert (1988)
31 Asymptotic Techniques for Use in Statistics
O.E. Bandorff-Nielsen and D.R. Cox (1989)
32 Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)

33 Analysis of Infectious Disease Data *N.G. Becker* (1989)

34 Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)

35 Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)

36 Symmetric Multivariate and Related Distributions
K.T. Fang, S. Kotz and K.W. Ng (1990)

37 Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)

38 Cyclic and Computer Generated Designs, 2nd edition
J.A. John and E.R. Williams (1995)

39 Analog Estimation Methods in Econometrics *C.F. Manski* (1988)

40 Subset Selection in Regression *A.J. Miller* (1990)

41 Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)

42 Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)

43 Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)

44 Inspection Errors for Attributes in Quality Control
N.L. Johnson, S. Kotz and X, Wu (1991)

45 The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46 The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)

47 Longitudinal Data with Serial Correlation—A state-space approach
R.H. Jones (1993)

48 Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)

49 Markov Models and Optimization *M.H.A. Davis* (1993)

50 Networks and Chaos—Statistical and probabilistic aspects
O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (1993)

51 Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)

52 Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)

53 Practical Risk Theory for Actuaries
C.D. Daykin, T. Pentikäinen and M. Pesonen (1994)

54 Biplots *J.C. Gower and D.J. Hand* (1996)

55 Predictive Inference—An introduction *S. Geisser* (1993)

56 Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)

57 An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)

58 Nonparametric Regression and Generalized Linear Models
P.J. Green and B.W. Silverman (1994)

59 Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)

60 Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)

61 Statistics for Long Memory Processes *J. Beran* (1995)

62 Nonlinear Models for Repeated Measurement Data
M. Davidian and D.M. Giltinan (1995)

63 Measurement Error in Nonlinear Models
R.J. Carroll, D. Rupert and L.A. Stefanski (1995)

64 Analyzing and Modeling Rank Data *J.J. Marden* (1995)

65 Time Series Models—In econometrics, finance and other fields
D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (1996)

66 Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)

67 Multivariate Dependencies—Models, analysis and interpretation
D.R. Cox and N. Wermuth (1996)

68 Statistical Inference—Based on the likelihood *A. Azzalini* (1996)

69 Bayes and Empirical Bayes Methods for Data Analysis
B.P. Carlin and T.A Louis (1996)

70 Hidden Markov and Other Models for Discrete-Valued Time Series
I.L. Macdonald and W. Zucchini (1997)

71 Statistical Evidence—A likelihood paradigm *R. Royall* (1997)

72 Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)

73 Multivariate Models and Dependence Concepts *H. Joe* (1997)

74 Theory of Sample Surveys *M.E. Thompson* (1997)

75 Retrial Queues *G. Falin and J.G.C. Templeton* (1997)

76 Theory of Dispersion Models *B. Jørgensen* (1997)

77 Mixed Poisson Processes *J. Grandell* (1997)

78 Variance Components Estimation—Mixed models, methodologies and applications
P.S.R.S. Rao (1997)

79 Bayesian Methods for Finite Population Sampling
G. Meeden and M. Ghosh (1997)

80 Stochastic Geometry—Likelihood and computation
O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (1998)

81 Computer-Assisted Analysis of Mixtures and Applications—
Meta-analysis, disease mapping and others *D. Böhning* (1999)

82 Classification, 2nd edition *A.D. Gordon* (1999)

83 Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)

84 Statistical Aspects of BSE and vCJD—Models for Epidemics
C.A. Donnelly and N.M. Ferguson (1999)

85 Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)

86 The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)

87 Complex Stochastic Systems
O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg (2001)

88 Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)

89 Algebraic Statistics—Computational Commutative Algebra in Statistics,
G. Pistone, E. Riccomagno and H.P. Wynn (2001)

Algebraic Statistics

Computational Commutative Algebra in Statistics

GIOVANNI PISTONE
EVA RICCOMAGNO
HENRY P. WYNN

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Pistone, Giovanni.
Algebraic statistics / Giovanni Pistone, Eva Riccomagno, Henry P. Wynn.
p. cm.-- (Monographs on statistics and applied probability ; 89)
Includes bibliographical references and index.
ISBN 1-58488-204-2 (alk. paper)
1. Mathematical statistics. 2. Algebra. I. Riccomagno, Eva. II. Wynn, Henry P. III. Title. IV. Series.

QA276 .P53 2000
519.5—dc21 00-047448

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

International Standard Book Number 1-58488-204-2
Library of Congress Card Number 00-047448
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Contents

List of figures

List of tables

Preface

About thirty-five years ago there was an awakening of interest of researchers in commutative algebra to the algorithmic and computational aspects of their field, marked by the publication of Buckberger's thesis in 1966. His work became the starting point of a new research field, called Computational Commutative Algebra. Currently, computer programs implementing versions of his and related algorithms are readily available both as commercial products and academic prototypes. These are of growing importance in almost every field of applied mathematics because they deal with very basic problems related to systems of polynomial equations. Statisticians, too, should find many useful tools in computational commutative algebra, together with interesting and enriching new perspectives. Just as the introduction of vectors and matrices has greatly improved the mathematics of statistics, these new tools provide a further step forward by offering a constructive methodology for a basic mathematical tool in statistics and probability, that is to say a ring. The mathematical structure of real random variables is precisely a ring, and other rings and ideals appear naturally in distribution theory and modeling. However, the ring of random variables is a ring with lattice operations which are not fully incorporated into the theory we present, at least not yet.

The authors' attention was drawn to the relevance of Gröbner basis theory by a paper on contingency tables by Sturmfels and Diaconis circulated as a manuscript in 1993. With initial help provided by Professor Teo Mora (University of Genova), a first application to design of experiments was published by G. Pistone and H. Wynn in 1996 (*Biometrika*) and this field of application was more fully developed by E. Riccomagno in her Ph.D. thesis work during 1996-97 at the University of Warwick. Subsequent papers in the same direction were published by the authors and a number of coauthors. We are pleased to acknowledge (in alphabetic order) Ron Bates, Massimo Caboara, Roberto Fontana, Beatrice Giglio, Tim Holliday, Maria-Piera Rogantin.

During the few years this monograph was in the making, we have benefitted from many contributions by others, and further related work is in progress. Some of the contents of this book was first exposed at the series of four GROSTAT workshops, which took place in successive years, starting in 1997 at the University of Warwick (UK), the IUT-STID in Nice-Côte d'Azur in Menton (France), EURANDOM in Eindhoven (NL), and again, in 2000, in Menton. We must thank all the participants and these institutions for their support, in particular Professor Annie Cavarero, director of IUT-STID.

We found keen collaborators at the University of Genova. We should at least mention, together with those above, Professor Lorenzo Robbiano (who also supported GROSTAT IV) and the CoCoA team who have had a major influence on the algebraic and computational aspects of the field. We are very grateful to them all for the early and generous access to their research, for the high level of illumination it provided on the mathematical foundations and the very fast computer code developed under the wings of CoCoA.

We are grateful for many discussions with colleagues and coworkers. A minimal list includes Wilf Kendall, Thomas Richardson, Raffaella Settimi and Jim Smith, in Warwick, and Alessandro Di Bucchianico and Arjeh Cohen, in Eindhoven. Special thanks to Dan Naiman of The Johns Hopkins University for allowing us to draw on recent joint work on tube theory in Chapter 4. Ian Dinwoodie, from Tulane University, helped to strengthen our understanding of the work of Diaconis and Sturmfels on toric ideals, which we reach in the final sections of the book, from our own particular direction. Because a considerable volume of the monograph is based on work in progress, we have, on a few occasions, had to refer to unpublished, although available, technical reports. We thank all the colleagues who helped us by reading different versions of this work, some of them already mentioned, and also Neil Parkin for careful reading of the whole book. We also thank our publishers for their help and considerable patience.

A cocktail of different grants and institutions has funded this research. We should thank the UK Engineering and Physical Sciences Research Council, the Italian Consiglio Nazionale delle Richerche, EURANDOM, and, last but not least, IRMA and the University L. Pasteur of Strasbourg, and Professor Dominique Collombier, who has hosted us during the final revision of the book.

This book is dedicated to our families, with apologies to all for the absences that a triple collaboration must entail.

Giovanni Pistone
Eva Riccomagno
Henry Wynn

Strasbourg, France, October 2000

Notation

Common symbols

$\mathbb{N}$	positive integer numbers
$\mathbb{Z}$	integer numbers
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers
S^*	* excludes the 0 from the set S
S_+	non-negative entries of the set of numbers S: for example $\mathbb{Z}_+ = \{a \in \mathbb{Z} : a \geq 0\} = \{0\} \cup \mathbb{N}$
d superscript	dimension of the cartesian product for example, $\mathbb{Z}^d$ stands for $\underbrace{\mathbb{Z} \times \cdots \times \mathbb{Z}}_{d \text{ times}}$
$\{a\}$	1. component-wise fractional part operator, $a \in \mathbb{R}^d$ 2. the set whose element is a
$\#A$	number of elements in the set A
$[p]$	vector or list p as a column vector
$[a_1 \cdots a_n]$	matrix with the vectors a_i, $i = 1, \ldots, n$ as columns
$[[\ldots], \ldots, [\ldots]]$	matrix as a list of rows
A^t	transpose of A where A is a matrix or a vector
I	identity matrix
$x_1, \ldots, x_d$	factors, variables, indeterminates
d	1. number of independent factors 2. number of variables 3. number of indeterminates
s	number of x_i's if the algebra is emphasised
N	1. sample size 2. number of design points 3. number of support points
k, $\mathcal{K}$	fields of coefficients for example, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Q}(\theta)$, transcendental extension, $\mathbb{Q}(\sqrt{2})$, algebraic extension

Notation for Gröbner bases

$k[x_1, \ldots, x_s]$	ring of polynomials in $x_1, \ldots, x_s$ and with coefficients in k
$x^\alpha = x_1^{\alpha_1} \ldots x_s^{\alpha_s}$	monomial in $k[x_1, \ldots, x_s]$
$p(x_1, \ldots, x_s)$	polynomial in $k[x_1, \ldots, x_s]$
$\tau, \succ, \succ_\tau$	term-ordering
$x_{i_1} \succ \ldots \succ x_{i_s}$	initial ordering on the indeterminates
$\tau(x_{i_1} \succ \ldots \succ x_{i_s})$	emphasis on the initial ordering
$\mathrm{LT}_\tau(p(x))$	leading term of the polynomial p with respect to the term-ordering τ
$\mathrm{Ideal}\,(g_1, \ldots, g_h)$ $\langle g_1, \ldots, g_h \rangle$	ideal of $k[x_1, \ldots, x_s]$ generated by $g_1, \ldots, g_h$
$\mathrm{Variety}(I)$	set of zeros of all polynomials in I
$\mathrm{Ideal}(V)$	set of all polynomials vanishing at V
$\mathrm{Variety}(f_1, \ldots, f_l)$	set of common roots of f_i, $i = 1, \ldots, l$
$\mathrm{Rem}(f), \mathrm{Rem}(f, G)$	1. normal form of f with respect to the Gröbner basis G 2. remainder of the division of f with respect to the set of polynomials G

Notation for experimental design

D, D_N	1. experimental design 2. support for a discrete distribution
a, x	design point
$x(i)$, $(x(i)_1, \ldots, x(i)_d)$	ith design point for $i = 1, \ldots, N$
$\mathcal{X}$	design region
$\mathrm{Est}_\tau(D)$	estimable terms with respect to τ and D
$\mathcal{F}$	polynomial regression vector
$Z = [f(x)]_{x \in D, f \in \mathcal{F}}$	design matrix for a model with support $\mathcal{F}$ and a design D; the orderings on D and $\mathcal{F}$ carry over to Z
$Z^t Z$	information matrix
$y = (y_1, \ldots, y_N)$	responses, values at the support points
θ, c, b, a	parameters or coefficients
$k[x_1, \ldots, x_d]/\,\mathrm{Ideal}(D)$ $k[x]/\,\mathrm{Ideal}(D)$	quotient ring
L	list of exponents of a vector space basis of $k[x_1, \ldots, x_d]/\mathrm{Ideal}(D)$
L_0	$L \setminus \{(0, \ldots, 0)\}$
L'	$L' \subseteq L$