19th International Conference on Computational Statistics

Paris - France, August 22-27, 2010

COMPSTAT'2010
Book of Abstracts

Conservatoire National des Arts et Métiers (CNAM)

and

the French National Institute for Research in Computer Science and Control (INRIA)

Title COMPSTAT'2010: Book of Abstracts

Cover design:

Authors: Several

Number of copies: 600

Preface

The 19th Conference of IASC-ERS, COMPSTAT'2010, is held in Paris, France, from August 22nd to August 27th 2010, locally organised by the Conservatoire National des Arts et Métiers (CNAM) and the French National Institute for Research in Computer Science and Control (INRIA).
COMPSTAT is an initiative of the European Regional Section of the International Association for Statistical Computing (IASC-ERS), a section of the International Statistical Institute (ISI).

Keynote lectures are addressed by **Luc Devroye** (School of Computer Science, McGill University, Montreal), **Lutz Edler** (Division of Biostatistics, German Cancer Research Center, Heidelberg) and **David Hand** (Statistics section, Imperial College, London). Each COMPSTAT meeting is organised with a number of topics highlighted, which lead to Invited Sessions. The Conference program includes also contributed sessions and short communications (both oral communications and posters).

The Conference Scientific Program Committee chaired by Gilbert Saporta, CNAM, includes:

Ana Maria Aguilera, Universidad Granada
Avner Bar-Hen, Université René Descartes, Paris
Maria Paula Brito, University of Porto
Christophe Croux, Katholieke Universiteit Leuven
Michel Denuit, Université Catholique de Louvain
Gejza Dohnal, Technical University, Prag
Patrick J. F. Groenen, Erasmus University, Rotterdam
Georges Hébrail, TELECOM ParisTech, Paris
Henk Kiers, University of Groningen
Erricos Kontoghiorghes, University of Cyprus
Martina Mittlböck, Medical University of Vienna
Christian P. Robert, Université Paris-Dauphine, Paris
Maurizio Vichi, Universita La Sapienza, Roma
Peter Winker, Universität Giessen
Moon Yul Huh, SungKyunKwan University, Seoul, Korea
Djamel Zighed, Université Lumière, Lyon

who were responsible for the Conference Scientific Program, and whom the organisers wish to thank for their invaluable cooperation and permanent avaibility.

This book contains the abstracts corresponding to all the presentations at the conference.

The organisers would like to express their gratitude to all people from CNAM and INRIA who contributed to the success of COMPSTAT'2010, and worked actively for its organisation. We are very grateful to all our sponsors, for their generous support. Finally, we thank all authors and participants, without whom the conference would not have been possible.

The organisers of COMPSTAT'2010 wish the best success to Erricos Kontoghiorghes, Chairman of the 20th edition of COMPSTAT, which will be held in Cyprus in Summer 2012. See you there!

Paris, August 2010
The Local Committee Organisers

Gilbert Saporta
Yves Lechevallier

Stéphanie Aubin
Gérard Biau
Stéphanie Chaix
Marc Christine
Laurence de Crémiers
Séverine Demeyer
Thierry Despeyroux
Christian Derquenne
Vincenzo Esposito Vinzi
Ali Gannoun
Jean-Pierre Gauchi
Chantal Girodon
Pierre-Louis Gonzalez
Luan Jaupi
Ludovic Lebart
Ndeye Niang
Françoise Potier
Giorgio Russolillo
Julie Séguéla

Acknowledgements

The Editors are extremely grateful to the reviewers, whose work was determinant for the scientific quality of these proceeding. They were, in alphatbetical order:

Hervé Abdi
Ana Maria Aguilera
Massimo Aria
Josef Arlt
Avner Bar-Hen
Jean-Patrick Baudry
Younès Bennani
Petr Berka
Patrice Bertrand
Pierre Bertrand
Gerard Biau
Christophe Biernacki
Lynne Billard
Hans-Hermann Bock
Frank Bretz
Henri Briand
Maria Paula Brito
Edgar Brunner
Stephane Canu
Gilles Celeux
Andrea Cerioli
Roy Cerqueti
Ka Chun Cheung
Marc Christine
Guillaume Cleuziou
Claudio Conversano
Christophe Croux
Francisco de Assis De Carvalho
Michel Denuit
Christian Derquenne
Thierry Despeyroux
Gejza Dohnal
Antonio D'Ambrosio
Manuel Escabias
Vincenzo Esposito Vinzi
Christian Francq
Giuliano Galimberti
Ali Gannoun
Bernard Garel
Cristian Gatu
Jean-Pierre Gauchi
Pierre-Louis Gonzalez
Gérard Govaert
Patrick Groenen
Nistor Grozavu
Fabrice Guillet
Frederic Guilloux
Anne Gégout-Petit
Hakim Hacid
Peter Hall
André Hardy
Georges Hébrail
Harald Heinzl
Marc Hoffman
Moon Yul Huh
Alfonso Iodice d'Enza
Antonio Irpino
Junling Ji
François-Xavier Jollois
Henk A.L. Kiers
Dong Kim
Christine Kiss
Erricos Kontoghiorghes
Labiod Lazhar
Ludovic Lebart
Mustapha Lebbah
Yves Lechevallier
Seung Lee
Guodong Li
Olivier Lopez
Maria Laura Maag
Jean-Michel Marin
Claudia Marinica
Roland Marion-Gallois
Geoffrey McLachlan
Bertrand Michel

Martina Mittlboeck
Angela Montanari
Irini Moustaki
Shu Ng
Ndeye Niang
Monique Noirhomme
Francisco A. Ocaña
Matej Oresic
Chongsun Park
Franceso Palumbo
Fabien Picarougne
Jean-Michel Poggi
Christian Preda
Tommaso Proietti
Pierre Pudlo
Jan Rauch
Marco Riani
Christian Robert
Nicoleta Rogovschi
Rosaria Romano
Fabrice Rossi
Anthony Rossini
Judith Rousseau
Laurent Rouviere
Maurice Roux
Giorgio Russolillo
Lorenza Saitta
Ryan Skraba
Gilbert Saporta
Seisho Sato
Roberta Siciliano
Francoise Soulie Fogelman
Matthias Studer
Laura Trinchera
Brigitte Trousse
Mariano J. Valderrama
Stefan Van Aelst
Gilles Venturini
Rosanna Verde
Maurizio Vichi
Emmanuel Viennet
Cinzia Viroli
Michal Vrabec
Franois Wahl
William Wieczorek
Peter Winker
Jingyun Yang
In-Kwon Yeo
Kam Yuen
Daniela Zaharie
Djamel A. Zighed
Lihong Zhang
Xinyuan Zhao

Sponsors

We are extremely grateful to the following institutions whose support contributes to the success of COMPSTAT'2010:

- Conseil Régional Ile de France
- Mairie de Paris
- Société Française de Statistique
- Association EGC (Extraction et Gestion des Connaissances)
- Société Francophone de Classification
- Electricité de France
- Institut National de la Recherche Agronomique
- Institut National de Recherche sur les Transports et leur Sécurit
- Institut National de la Statistique et des Etudes Economiques
- IPSOS
- Orange Labs
- SAS-Institute

Programme

With the help of the Scientific Programme Committee and many reviewers, we have prepared a scientific programme which we hope very attractive, as well as the social programme.

Besides the 3 keynote lectures and the 39 invited communications, we received about 450 submissions. After the reviewing process we retained 127 'long papers', 92 'short papers' and 144 posters.

The authors come from 52 countries.

country	authors	country	authors	country	authors
Spain	115	Brazil	11	Luxembourg	2
France	94	Iran	10	Mexico	2
Japan	83	Hong Kong	8	Morocco	2
Italy	76	Korea Republic	8	Norway	2
Germany	52	China	5	Romania	2
United States	51	New Zealand	5	Slovenia	2
Belgium	39	Russian Federation	4	Bahrain	1
Czech Republic	38	Slovakia	4	Bangladesh	1
Turkey	30	Estonia	3	Colombia	1
United Kingdom	24	Finland	3	Georgia	1
Portugal	19	India	3	Greece	1
Austria	18	Netherlands	3	Hungary	1
Australia	16	Singapore	3	Indonesia	1
Taiwan	16	Sweden	3	Malaysia	1
Poland	15	Algeria	2	Niger	1
Switzerland	14	Argentina	2	Senegal	1
Canada	12	Bulgaria	2		
Tunisia	12	Lithuania	2		

Scientific Programme by Day

This book contains the abstracts corresponding to all the presentations at the conference. It is organized by day, according to the type of session: 3 keynote sessions, 14 invited sessions, 23 contributed sessions, 8 sessions of short papers and 2 posters sessions.

Monday, August 23

Time	Session	Title	Pages
9h30-10h30	KN1	David Hand	3
11h00-12h30	IP1	Algorithms for Robust Statistics	4-6
	IP2	Functional Data Analysis	7-9
14h00-16h00	CP1	Algorithm for Robust Statistics & Robustness	10-15
	CP2	Categorical Data Analysis	16-21
	CP3	Computational Bayesian Methods	22-27
	CP4	Multivariate Data Analysis 1	28-33
	CP5	Time Series Analysis & Signal Processing 1	34-38
	SP1	Biostatistics	39-50
16h30-18h30	CP6	Biostatistics & Bio-Computing 1	51-58
	CP7	Clustering & Classification 1	59-65
	CP8	Functional Data	66-72
	CP9	Multivariate Data Analysis 2	73-78
	CP10	Time Series Analysis & Signal Processing 2	79-85
	SP2	Nonparametric Statistics	86-96

Tuesday, August 24

Time	Session	Title	Pages
9h00-10h30	IP3	Brain Imaging	99-101
	IP4	Computer-Intensive Actuarial Methods	102-104
11h00-12h30	CP11	Biostatistics & Bio-Computing 2	105-109
	CP12	Clustering & Classification 2	110-114
	CP13	Nonparametric Statistics & Smoothing	115-119
	CP14	Optimization Heuristics	120-123
	SP3	Econometrics & Finance	124-133
	SP4	Multivariate Analysis 1 & Spatial Statistics	134-145
14h00-16h00	CP15	Multivariate Data Analysis 3	146-150
14h00-15h00	CP16	Data Visualization	151-153
15h00-16h00	CP17	Sampling Methods	154-156
14h00-16h00	CP18	Spatial Statistics	157-160
	CP19	Time Series Analysis & Signal Processing 3	161-166
	SP5	Classification	167-178
	SP6	Data Analysis	179-190
16h30-18h30	PS1	poster session 1	191-260

Wednesday , august 25

Time	Session	Title	Pages
9h00-10h30	IP5	Computational Econometrics	263-265
	IP6	Optimization Heuristics in Statistical Modelling	266-268
11h30- 12h30	KN2	Luc Devroye	269
14h00-15h30	IP7	Data Stream Mining	270-271
	IP8	ARS Session (Financial) Time Series	272-274

Thursday, August 26

Time	Session	Title	Pages
9h00-10h30	IP9	Spatial Statistics / Spatial Epidemiology	277-279
	IP10	KDD Session: Topological Learning	280-281
11h00-12h30	PS2	poster session 2	282-356
14h00-15h30	IP11	ABC Methods for Genetic Data	357-359
	IP12	IFCS Session	360-362
16h30-18h30	CP20	Computational Econometrics & Finance	363-368
	CP21	Machine Learning	369-374
	CP22	Numerical Methods	375-380
	CP23	Symbolic Data Analysis	381-386
	SP7	Multivariate Analysis 2	387-397
	SP8	Time Series & Numerical Methods	398-409

Friday, August 27

Time	Session	Title	Pages
9h00-10h30	IP13	Kernel Methods	413-414
	IP14	Monte Carlo Methods	415-417
11h00-12h00	KN3	Lutz Edler	418

Contents

SP1: Biostatistics

CP6: Biostatistics & Bio-Computing 1

CP7: Clustering & Classification 1

CP8: Functional Data

CP13: Nonparametric Statistics & Smoothing

CP14: Optimization heuristics in Statistical Modelling

SP3: Econometrics & Finance

SP4: Multivariate Analysis 1 & Spatial Statistics

CP15: Multivariate Data Analysis 3

SP6: Data Analysis

PS1: Poster Session 1

IP6: Optimization heuristics in statistical modelling

KN2: Keynote Speaker 2

IP7: Data Stream Mining

IP8: ARS Session (Financial) Time Series

Part IV. Thursday August 26

IP9: Spatial Statistics / Spatial Epidemiology

IP10: KDD Session: Topological Learning

PS2: Poster Session 2

SP8: Time Series & Numerical Methods

IP14: Monte Carlo Methods

KN3: Kenote Speaker 3

High-Dimensional Classification in the Presence of Correlation: A Factor Model Approach

A. Pedro Duarte Silva

Faculdade de Economia e Gestão & CEGE, Univ. Católica Portuguesa at Porto, Rua Diogo Botelho, 1327, 4169-005 Porto, Portugal; *psilva@porto.ucp.pt*

Abstract. A class of linear discrimination rules, designed for problems with many correlated variables, is proposed. These rules try to incorporate the most important patterns revealed by the empirical correlations and accurately approximate the optimal Bayes rule as the number of variables increases without limit. In order to achieve this goal, the rules rely on covariance matrix estimates derived from Gaussian factor models with small intrinsic dimensionality.

Asymptotic results, based on an analysis that allows the number of variables to grow faster than the number of observations, show that the worst possible expected error rate of the proposed rules converges to the error of the optimal Bayes rule when the postulated model is true, and to a slightly larger constant when this model is a close approximation to the data generating process.

Simulation results using real micro-array data and a new variable selection scheme based on Donoho and Jin's Higher Criticism statistic, suggest that under the conditions they were designed for, rules derived from covariance one-factor models perform equally well or better than the most successful extant alternatives.

Keywords: Discriminant Analysis, High Dimensionality, Expected Misclassification Rate, MicroArray Classification.

References

BICKEL, P.J and LEVINA, E. (2004): Some theory for Fisher's Linear Discriminant Function, Naive Bayes and some alternatives when there are many more variables than observations. *Bernoulli 10 (6), 989-1010.*

DONOHO, D. and JIN, J. (2004): Higher Criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics 32 (3), 962-994.*

DONOHO, D. and JIN, J. (2008): Higher criticism thresholding. Optimal feature selection when useful features are rare and weak. In: *Proc. Natl. Acad. Sci.*, 105: 14790-14795.