

ABEL SYMPOSIA

Edited by the Norwegian Mathematical Society

More information about this series at <http://www.springer.com/series/7462>



Participants at the 2014 Abel Symposium at Nyvågar.
Photo taken by Mette Langaas

Arnoldo Frigessi • Peter Bühlmann •
Ingrid K. Glad • Mette Langaas •
Sylvia Richardson • Marina Vannucci
Editors

Statistical Analysis for High-Dimensional Data

The Abel Symposium 2014



ABEL
PRISEN

 Springer

Editors

Arnoldo Frigessi
Oslo Centre for Biostatistics and
Epidemiology
University of Oslo
Oslo, Norway

Peter Bühlmann
Seminar for Statistics
ETH Zürich
Zürich, Switzerland

Ingrid K. Glad
Department of Mathematics
University of Oslo
Oslo, Norway

Mette Langaas
Department of Mathematical Sciences
Norwegian University of Science
and Technology
Trondheim, Norway

Sylvia Richardson
MRC Biostatistics Unit
Cambridge Institute of Public Health
University of Cambridge
Cambridge, UK

Marina Vannucci
Department of Statistics
Rice University
Houston
Texas, USA

ISSN 2193-2808

Abel Symposia

ISBN 978-3-319-27097-5

DOI 10.1007/978-3-319-27099-9

ISSN 2197-8549 (electronic)

ISBN 978-3-319-27099-9 (eBook)

Library of Congress Control Number: 2016931284

Mathematics Subject Classification (2010): 62F15, 62H30, 62P10, 62F99, 62H99, 62J99

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Foreword

The Norwegian government established the Abel Prize in mathematics in 2002, and the first prize was awarded in 2003. In addition to honoring the great Norwegian mathematician Niels Henrik Abel by awarding an international prize for outstanding scientific work in the field of mathematics, the prize shall contribute toward raising the status of mathematics in society and stimulate the interest for science among school children and students. In keeping with this objective, the Niels Henrik Abel Board has decided to finance annual Abel Symposia. The topic of the symposia may be selected broadly in the area of pure and applied mathematics. The symposia should be at the highest international level and serve to build bridges between the national and international research communities. The Norwegian Mathematical Society is responsible for the events. It has also been decided that the contributions from these symposia should be presented in a series of proceedings, and Springer Verlag has enthusiastically agreed to publish the series. The Niels Henrik Abel Board is confident that the series will be a valuable contribution to the mathematical literature.

Chair of the Niels Henrik Abel Board

Helge Holden

Preface

The 11th Abel Symposium on “Statistical Analysis for High-Dimensional Data” was held at Nyvågar Rorbuhotell, near the village of Kabelvåg, in the Lofoten archipelago, Norway, from Monday, May 5, to Friday, May 9, 2014. This Abel Symposium Proceedings contains a selection of papers within this field.

There were 53 participants, and participation was by invitation only. The scientific committee for the symposium consisted of the editors of this Abel Symposium volume: Peter Bühlmann, Arnoldo Frigessi, Ingrid K. Glad, Mette Langaas, Sylvia Richardson, and Marina Vannucci.

The symposium had 23 invited talks:

Yoav Benjamini	Addressing statistical woes affecting replicability
Peter J. Bickel	Asymptotic analysis of unlabeled networks and an application
Leonardo Bottolo	Bayesian hierarchical mixture models
Tony Cai	Sparse PCA: optimal rates, adaptive estimation, and computational challenges
Mathias Drton	Ising model selection with Bayesian information criteria
Edward I. George	Spike-and-slab variable selection
Jelle J. Goeman	A global \times global test for association between two high-dimensional data sets
Chris Hans	From a conditional Lindley’s paradox to poly-hyper- g priors in Bayesian regression
Trevor Hastie	Sparse models and coordinate descent
Nicolai Meinshausen	Maximin effects for heterogeneous large-scale data
Xiao-Li Meng	The potential and perils of preprocessing: building new foundations
Sach Mukherjee	High-dimensional approaches for investigating disease heterogeneity

Peter Müller	A Bayesian feature allocation model for tumor heterogeneity
Susan A. Murphy	Micro-randomized trials and off-policy reinforcement learning
Judith Rousseau	Asymptotic properties of empirical Bayes procedures in parametric and nonparametric models
Richard Samworth	Independent component analysis via nonparametric maximum likelihood estimation
Terence P. Speed	Removing unwanted variation for prediction
Sara van de Geer	Confidence sets for the graphical lasso
Martin J. Wainwright	Randomized sketches of high-dimensional convex programs: practical schemes and sharp rates
Mike West	Dynamic sparsity modeling
Bin Yu	Relating developmental transcription factors (TFs) based on <i>Drosophila</i> embryonic gene expression images

The participants not giving invited talks had the possibility to give a short presentation of their work. Contributed talks were given by Magne T. Aldrin, Bettina Kulle Andreassen, Elja Arjas, Lamiae Azizi, Linn Cecilie Bergersen, Peter Bühlmann, Jan Christoph Ernest, Arnaldo Frigessi, Ingrid K. Glad, Clara-Cecilie Günther, Kari K. Halle, Kristoffer H. Hellton, Eric B. Laber, Mette Langaas, Tonje G. Lien, Ole Christian Lingjærde, Keli Liu, Edgar A. Muro Jimenez, Christian M. Page, Andrea I. Riebler, Sylvia Richardson, Daniel P. Simpson, Nicolas Städler, Øystein Sørensen, Marina Vannucci, Valeria Vitelli, Verena Zuber, and Erik W. van Zwet.

The full list of participants follows:

Aldrin, Magne T.	Norwegian Computing Center
Andreassen, Bettina Kulle	University of Oslo
Arjas, Elja	University of Oslo
Azizi, Lamiae	University of Cambridge
Benjamini, Yoav	Tel Aviv University
Bergersen, Linn Cecilie	University of Oslo
Bickel, Peter J.	University of California at Berkeley
Bottolo, Leonardo	Imperial College London
Bühlmann, Peter	ETH Zürich
Cai, Tony	University of Pennsylvania
Drton, Mathias	University of Washington
Ernest, Jan Christoph	ETH Zürich
Frigessi, Arnaldo	University of Oslo
George, Edward I.	University of Pennsylvania
Glad, Ingrid K.	University of Oslo

Goeman, Jelle J.	Leiden University
Günther, Clara-Cecilie	Norwegian Computing Center
Halle, Kari K.	Norwegian University of Science and Technology
Hans, Chris	Ohio State University
Hastie, Trevor	Stanford University
Hellton, Kristoffer H.	University of Oslo
Hjort, Nils Lid	University of Oslo
Laber, Eric B.	North Carolina State University
Langaas, Mette	Norwegian University of Science and Technology
Lien, Tonje G.	University of Oslo
Lingjærde, Ole Christian	University of Oslo
Liu, Keli	Harvard University
Low, Mark G.	University of Pennsylvania
Meinshausen, Nicolai	ETH Zürich
Meng, Xiao-Li	Harvard University
Mukherjee, Sach	Netherlands Cancer Institute
Müller, Peter	University of Texas, Austin
Muro Jimenez, Edgar A.	ETH Zürich
Murphy, Susan A.	University of Michigan
Page, Christian M.	University of Oslo
Richardson, Sylvia	University of Cambridge
Riebler, Andrea I.	Norwegian University of Science and Technology
Rousseau, Judith	Université Paris Dauphine
Rue, Håvard	Norwegian University of Science and Technology
Salomond, Jean-Bernard	Université Paris Dauphine
Samworth, Richard	University of Cambridge
Simpson, Daniel P.	Norwegian University of Science and Technology
Sørensen, Øystein	University of Oslo
Speed, Terence P.	Walter and Eliza Hall Institute of Medical Research
Städler, Nicolas	The Netherlands Cancer Institute
Thoresen, Magne	University of Oslo
van de Geer, Sara	ETH Zürich
van Zwet, Erik W.	Leiden University
Vannucci, Marina	Rice University
Vitelli, Valeria	University of Oslo
Wainwright, Martin J.	University of California, Berkeley
West, Mike	Duke University
Yu, Bin	University of California, Berkeley
Zuber, Verena	University of Oslo

We thank the Norwegian Mathematical Society for nominating our application and the Board for the Niels Henrik Abel Memorial Fund for deciding to support it.

We also thank the organizing committee, consisting of the participants Magne T. Aldrin, Bettina Kulle Andreassen, Arnaldo Frigessi, Ingrid K. Glad, Clara-Cecilie Günther, Kristoffer H. Hellton, Mette Langaas, Tonje G. Lien, Christian M. Page, Øystein Sørensen, and in addition Elisabeth H. Seland. Further; we would like to acknowledge the excellent service we received from the staff at the Nyvågar Rorbuhotell during our stay.

Special thanks goes to the contributors to this volume and to the participants and the speakers for their part toward the success of the symposium. Finally, we would like to thank Springer and Ruth Allewelt, for great patience and service during the preparation of this volume.

Oslo, Norway
Zürich, Switzerland
Oslo, Norway,
Trondheim, Norway
Cambridge, UK
Houston, TX, USA
October 2015

Arnaldo Frigessi
Peter Bühlmann
Ingrid K. Glad
Mette Langaas
Sylvia Richardson
Marina Vannucci

Contents

Some Themes in High-Dimensional Statistics	1
Arnoldo Frigessi, Peter Bühlmann, Ingrid K. Glad, Sylvia Richardson, and Marina Vannucci	
Laplace Approximation in High-Dimensional Bayesian Regression	15
Rina Foygel Barber, Mathias Drton, and Kean Ming Tan	
Preselection in Lasso-Type Analysis for Ultra-High Dimensional Genomic Exploration	37
Linn Cecilie Bergersen, Ismail Ahmed, Arnoldo Frigessi, Ingrid K. Glad, and Sylvia Richardson	
Spectral Clustering and Block Models: A Review and a New Algorithm	67
Sharmodeep Bhattacharyya and Peter J. Bickel	
Bayesian Hierarchical Mixture Models	91
Leonardo Bottolo and Petros Dellaportas	
iBATCGH: Integrative Bayesian Analysis of Transcriptomic and CGH Data	105
Alberto Cassese, Michele Guindani, and Marina Vannucci	
Models of Random Sparse Eigenmatrices and Bayesian Analysis of Multivariate Structure	125
Andrew Cron and Mike West	
Combining Single and Paired End RNA-seq Data for Differential Expression Analyses	155
Zhi-Ping Feng, Francois Collin, and Terence P. Speed	
An Imputation Method for Estimating the Learning Curve in Classification Problems	189
Eric B. Laber, Kerby Shedden, and Yang Yang	

Bayesian Feature Allocation Models for Tumor Heterogeneity 211
Juhee Lee, Peter Müller, Subhajit Sengupta, Kamalakar Gulukota,
and Yuan Ji

Bayesian Penalty Mixing: The Case of a Non-separable Penalty 233
Veronika Ročková and Edward I. George

**Confidence Intervals for Maximin Effects in Inhomogeneous
Large-Scale Data** 255
Dominik Rothenhäusler, Nicolai Meinshausen,
and Peter Bühlmann

χ^2 -Confidence Sets in High-Dimensional Regression 279
Sara van de Geer and Benjamin Stucky