Lecture Notes in Mathematics 2159

Saint-Flour Probability Summer School

The Saint-Flour volumes are reflections of the courses given at the Saint-Flour Probability Summer School. Founded in 1971, this school is organised every year by the Laboratoire de Mathématiques (CNRS and Université Blaise Pascal, Clermont-Ferrand, France). It is intended for PhD students, teachers and researchers who are interested in probability theory, statistics, and in their applications.

The duration of each school is 13 days (it was 17 days up to 2005), and up to 70 participants can attend it. The aim is to provide, in three high-level courses, a comprehensive study of some fields in probability theory or Statistics. The lecturers are chosen by an international scientific board. The participants themselves also have the opportunity to give short lectures about their research work.

Participants are lodged and work in the same building, a former seminary built in the 18th century in the city of Saint-Flour, at an altitude of 900 m. The pleasant surroundings facilitate scientific discussion and exchange.

The Saint-Flour Probability Summer School is supported by:
- Université Blaise Pascal
- Centre National de la Recherche Scientifique (C.N.R.S.)
- Ministère délégué à l'Enseignement supérieur et à la Recherche

For more information, see

http://recherche.math.univ-bpclermont.fr/stflour/stflour-en.php

Christophe Bahadoran
bahadora@math.univ-bpclermont.fr

Arnaud Guillin
Arnaud.Guillin@math.univ-bpclermont.fr

Laurent Serlet
Laurent.Serlet@math.univ-bpclermont.fr

Université Blaise Pascal – Aubière cedex, France

More information about this series at http://www.springer.com/series/304

Sara van de Geer

Estimation and Testing Under Sparsity

École d'Été de Probabilités de Saint-Flour XLV – 2015

Springer

Sara van de Geer
Seminar für Statistik HGG 24.1
ETH Zentrum
Zürich, Switzerland

ISSN 0075-8434 ISSN 1617-9692 (electronic)
Lecture Notes in Mathematics
ISBN 978-3-319-32773-0 ISBN 978-3-319-32774-7 (eBook)
DOI 10.1007/978-3-319-32774-7

Library of Congress Control Number: 2016943427

Mathematics Subject Classification (2010): 62-XX; 60-XX, 68Q87

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To my mother

Preface

These lecture notes were written for the 45th École d'Été de Probabilités de Saint-Flour in 2015. It was a great honour and a pleasure for me to be invited there and enjoy the *Grand Séminaire*, the beautiful surroundings and, most of all, the wonderful participants. The summer was very warm, with temperatures rising up to 40°. However, the old building kept us cool and ready to digest quite a few lectures each day. Indeed, the amount of mathematical activity was impressive, with a scientific programme in both mornings and afternoons and many excellent talks presented by my fellow participants. But there were also numerous other activities: chatting, hiking, dining, ping-pong, and looking after the small kids.

The notes aim to provide an overview of the techniques used to obtain theoretical results for models of high-dimensional data. High-dimensional models have more unknown parameters p than observations n. Sparsity means that the number of relevant—or active—parameters is actually much smaller than n. However, which parameters these are, or how many there are, is not known beforehand. The first goal in these notes is to study methods that perform almost as well as in the hypothetical case with a known active set. The next goal is then to zoom in on certain parameters of interest and test their significance.

An important technique in high-dimensional regression is the Lasso method. It is taken here as a prototype for understanding other methods, such as those inducing structured sparsity or low rank or those based on more general loss functions. The common features are highlighted so that—hopefully—they will serve as a good starting point for the theory of new methods, not treated in this book.

I am very grateful to the Scientific Board for having given me the opportunity to lecture at the Saint-Flour Summer School. I thank all participants and am greatly indebted to Claire Boyer, Yohann De Castro and Joseph Salmon, who spontaneously did a careful reading of the version of the lecture notes available at the time. The mistakes in the current version have entered *after* their proofreading. Special thanks goes to the organisers on the spot, Christophe Bahadoran and Laurent Serlet. I also thank the two other main lecturers, Sourav Chatterjee and Lorenzo Zambotti for our

very pleasant time together and for their inspiring courses which opened up new windows with magnificent views.

I thank my colleagues at the Seminar for Statistics in Zürich for their support, for their motivating interest and for providing me with an ideal research environment.

Finally, a very special thanks is extended to my family.

Zürich, Switzerland
March 2016

Sara van de Geer

Contents