Lectures on Algebraic Statistics

Bearbeitet von
Mathias Drton, Bernd Sturmfels, Seth Sullivant

Weitere Fachgebiete > Mathematik > Stochastik > Mathematische Statistik

schnell und portofrei erhältlich bei

beck-shop.de
DIE FACHBUCHHANDLUNG

Preface

Algebraic statistics is concerned with the development of techniques in algebraic geometry, commutative algebra, and combinatorics, to address problems in statistics and its applications. On the one hand, algebra provides a powerful tool set for addressing statistical problems. On the other hand, it is rarely the case that algebraic techniques are ready-made to address statistical challenges, and usually new algebraic results need to be developed. This way the dialogue between algebra and statistics benefits both disciplines.

Algebraic statistics is a relatively new field that has developed and changed rather rapidly over the last fifteen years. One of the first pieces of work in this area was the paper of Diaconis and the second author [33], which introduced the notion of a Markov basis for log-linear statistical models and showed its connection to commutative algebra. From there, the algebra/statistics connection spread to a number of different areas including the design of experiments (highlighted in the monograph [74]), graphical models, phylogenetic invariants, parametric inference, algebraic tools for maximum likelihood estimation, and disclosure limitation, to name just a few. References to this literature are surveyed in the editorial [47] and the two review articles [4, 41] in a special issue of the journal *Statistica Sinica*. An area where there has been particularly strong activity is in applications to computational biology, which is highlighted in the book *Algebraic Statistics for Computational Biology* of Lior Pachter and the second author [73]. We will sometimes refer to that book as the "ASCB book."

These lecture notes arose out of a five-day Oberwolfach Seminar, given at the *Mathematisches Forschungsinstitut Oberwolfach* (MFO), in Germany's Black Forest, over the days May 12–16, 2008. The seminar lectures provided an introduction to some of the fundamental notions in algebraic statistics, as well as a snapshot of some of the current research directions. Given such a short timeframe, we were forced to pick and choose topics to present, and many areas of active research in algebraic statistics have been left out. Still, we hope that these notes give an overview of some of the main ideas in the area and directions for future research.

The lecture notes are an expanded version of the thirteen lectures we gave throughout the week, with many more examples and background material than we could fit into our hour-long lectures. The first five chapters cover the material

in those thirteen lectures and roughly correspond to the five days of the workshop. Chapter 1 reviews statistical tests for contingency table analysis and explains the notion of a Markov basis for a log-linear model. We connect this notion to commutative algebra, and give some of the most important structural theorems about Markov bases. Chapter 2 is concerned with likelihood inference in algebraic statistical models. We introduce these models for discrete and normal random variables, explain how to solve the likelihood equations parametrically and implicitly, and show how model geometry connects to asymptotics of likelihood ratio statistics. Chapter 3 is an algebraic study of conditional independence structures. We introduce these generally, and then focus in on the special class of graphical models. Chapter 4 is an introduction to hidden variable models. From the algebraic point of view, these models often give rise to secant varieties. Finally, Chapter 5 concerns Bayesian integrals, both from an asymptotic large-sample perspective and from the standpoint of exact evaluation for small samples.

During our week in Oberwolfach, we held several student problem sessions to complement our lectures. We created eight problems highlighting material from the different lectures and assigned the students into groups to work on these problems. The exercises presented a range of computational and theoretical challenges. After daily and sometimes late-night problem solving sessions, the students wrote up solutions, which appear in Chapter 6. On the closing day of the workshop, we held an open problem session, where we and the participants presented open research problems related to algebraic statistics. These appear in Chapter 7.