

# Abstract

## On Boundaries of Statistical Models

Thomas Kahle

In the thesis “On Boundaries of Statistical Models” problems related to a description of probability distributions with zeros, lying in the boundary of a statistical model, are treated. The distributions considered are joint distributions of finite collections of finite discrete random variables. Owing to this restriction, statistical models are subsets of  $\mathbb{R}^m$ . The support set problem for exponential families, the main class of models considered in the thesis, is to characterize the possible supports of distributions in the boundaries of these statistical models. It is shown that this problem is equivalent to a characterization of the face lattice of a convex polytope, called the convex support. The main tool for treating questions related to the boundary are implicit representations. Exponential families are shown to be sets of solutions of binomial equations, connected to an underlying combinatorial structure, called oriented matroid. Under an additional assumption these equations are polynomial and one is placed in the setting of commutative algebra and algebraic geometry. In this case one recovers results from algebraic statistics. The combinatorial theory of exponential families using oriented matroids makes the established connection between an exponential family and its convex support completely natural: Both are derived from the same oriented matroid.

The second part of the thesis deals with hierarchical models, which are a special class of exponential families constructed from simplicial complexes. The main technical tool for their treatment in this thesis are so called elementary circuits. After their introduction, they are used to derive properties of the implicit representations of hierarchical models. Each elementary circuit gives an equation holding on the hierarchical model, and these equations are shown to be the “simplest”, in the sense that the smallest degree among the equations corresponding to elementary circuits gives a lower bound on the degree of all equations characterizing the model. Translating this result back to polyhedral geometry yields a neighborliness property of marginal polytopes, the convex supports of hierarchical models. Elementary circuits of small support are related to independence statements holding between the random variables whose joint distributions the hierarchical model describes. Models for which the complete set of circuits consists of elementary circuits are shown to be described by totally unimodular matrices. The thesis also contains an analysis of the case of binary random variables. In this special situation, marginal polytopes can be represented as the convex hulls of linear codes. Among the results here is a classification of full-dimensional linear code polytopes in terms of their subgroups.

If represented by polynomial equations, exponential families are the varieties of binomial prime ideals. The third part of the thesis describes tools to treat models defined by not necessarily prime binomial ideals. It follows from Eisenbud and Sturmfels’ results on binomial ideals that these models are unions of exponential families, and apart from solving the support set problem for each of these, one is faced with finding the decomposition. The thesis discusses algorithms for specialized treatment of binomial ideals, exploiting their combinatorial nature. The provided software package *Binomials* is shown to be able to compute very large primary decompositions, yielding a counterexample to a recent conjecture in algebraic statistics.