博士論文 (要約)

Algebraic Statistical Methods for Conditional Inference of Discrete Statistical Models

(離散統計モデルの条件付き推測問題に対する代数統計的手法)

小川 光紀

Testing and parameter estimation are fundamental topics in statistical inference. Traditionally these problems have mainly been solved by large sample approximation techniques. However, such problems sometimes require conditional methods instead of the usual statistical methods. In this thesis, therefore, we develop some algebraic statistical methods for conditional inferences. This thesis consists of six chapters. Chapter 1 is the introduction and it gives a general background of the themes which are dealt in this thesis. In Chapters 2 and 3 we study the structure of the Markov bases and the Graver bases for some concrete statistical models. In Chapter 2 we deal with the models for two-way contingency tables considering the effects of subtables. We derive the explicit forms of the corresponding Markov bases and apply them to numerical experiments with real data sets. Chapter 3 is devoted to construct the Markov chain Monte Carlo (MCMC) method for sampling from the conditional distribution of the beta model. We provide the explicit description of the Graver basis for an undirected graph and apply it to the MCMC procedure. In Chapter 4, a new procedure for generating a series of configurations from an original configuration is defined, and we study the Markov degree of the resulting configurations. In Chapter 5 we discuss the application of A-hypergeometric system for computation of the normalizing constant of the conditional likelihood. The objective of this application is the conditional maximum likelihood estimation, which is important approach in parameter estimation in the presence of many nuisance parameters. Chapter 6 is the concluding remark of this thesis.

Algebraic statistics is a rapidly developing area that exploits algebraic geometry and related techniques to solve problems in statistics. There are two seminal papers: Diaconis and Sturmfels [7] and Pistone and Wynn [37]. The former developed an algebraic sampling algorithm for the conditional distributions of discrete exponential families, whereas the latter developed the experimental design theory in algebraic statistics. Both of them applied Gröbner basis theory in polynomial rings to statistical problems. This connection between statistics and computational algebra has inspired the numerous developments and successful interactions between the two areas.

Goodness-of-fit tests for statistical models are usually performed by a large sample approximation to the null distribution of a test statistic. However, as shown in Haberman [15], the large sample approximation may not be appropriate when the expected frequencies are not large enough. In such cases it is desirable to use a conditional test based on the exact distribution of the test statistic.

Diaconis and Sturmfels [7] developed an algebraic sampling algorithm for conditional distributions based on the MCMC method. The set of moves, known as the Markov basis, which guarantees the connectivity for every fiber, plays an essential role in their algorithm. They established the equivalence between a Markov basis and a binomial generator for the toric ideal arising from a statistical model of discrete exponential families. Thanks to their algorithm, once we have a Markov basis for a given statistical model, we can perform a conditional test for that model via the MCMC method. For the general background on conditional tests and Markov bases, see Drton et al. [10].

Due to the equivalence of the Markov basis and the binomial generator for the toric ideal, we can use the theory and algorithms for the Gröbner bases to calculate a Markov basis. There exist algebraic softwares for computing Gröbner bases such as `4ti2` [1]. Since the number of elements of Markov bases is usually too large, the computation of Markov bases is often difficult in a practical amount of time. Furthermore, even if we obtain all elements of a Markov basis, we need a large amount of memory to hold the elements of them. Thus, it is desirable to derive the explicit forms of Markov bases and perform an adaptive algorithm for the conditional test.

The structures of Markov bases for statistical models of contingency tables have been studied by many researchers in algebraic statistics (e.g., Dobra and Sullivant [8], Aoki and

Takemura [2], Rapallo [38], Hara et al. [19]). However, the structures of Markov bases are complicated in general and the results on the structures of Markov bases are known only for limited statistical models.

The first aim of this thesis is to derive the Markov bases for some statistical models of two-way contingency tables considering the effect of subtables. It is a well-known fact that the set of square-free moves of degree two (basic moves) forms the minimal Markov basis for the complete independence model of two-way contingency tables. On the other hand, when a subtable effect is added to the model, the set of basic moves does not necessarily form a Markov basis. This problem is called the two-way subtable sum problem ([18]).

Hara et al. [17, 18] discussed Markov bases for the configuration for one subtable effect. In [18] it is shown that the set of basic moves is a Markov basis if and only if the subtable is a 2×2-diagonal or triangular block. Ohsugi and Hibi [32] discussed the same problem from an algebraic viewpoint. In this thesis we consider some statistical models with typical block-wise subtable effects. For the model which has a quadratic Markov basis, we also discuss the algebraic properties of the configuration arising from the model.

The second aim of this thesis is to construct the MCMC method with Graver bases for conditional testing of the beta model of random graphs. Random graphs and their applications to the statistical modeling of complex networks have been attracting much interest in many fields, including statistical mechanics, ecology, biology and sociology (e.g., Newman [27], Goldenberg et al. [13]). Statistical models for random graphs have been studied since Solomonoff and Rapoport [46] and Erdős and Rényi [11] introduced the Bernoulli random graph model. The beta model generalizes the Bernoulli model to a discrete exponential family with vertex degrees as sufficient statistics. The beta model was discussed by Holland and Leinhardt [22] in the directed case and by Park and Newman [35], Blitzstein and Diaconis [3] and Chatterjee et al. [5] in the undirected case. The Rasch model [39], which is a standard model in the item response theory, is also interpreted as a beta model for undirected complete bipartite graphs. In this thesis we discuss the random sampling of graphs from the conditional distribution in the beta model when the vertex degrees are fixed.

In the context of a social network the vertices of the graph represent individuals and their edges represent relationships between individuals. In the undirected case the graphs are sometimes restricted to be simple, i.e., they contain no loops or multiple edges. The sample size for such cases is at most the number of edges of the graph and is often small. Thus, for the beta model with the restriction that graphs are simple, it is desirable to perform a conditional testing procedure.

Random sampling of graphs with a given vertex degree sequence enables us to numerically evaluate the exact distribution of a test statistic for the beta model. Blitzstein and Diaconis [3] developed a sequential importance sampling algorithm for simple graphs that generates graphs through operations on vertex degree sequences. In this thesis we construct a Markov chain Monte Carlo algorithm for sampling graphs by using the Graver basis for the toric ideal arising from the underlying graph of the beta model.

A set of graphs with a given vertex degree sequence is called a fiber for the underlying graph of the beta model. A Markov basis for the underlying graph of the beta model is also considered as a set of Markov transition operators connecting all elements of every fiber. Petrović et al. [36] discussed some properties of the toric ideal arising from the model of [22] and provided Markov bases of the model for small directed graphs. Properties of toric ideals arising from undirected graphs have been studied in a series of papers by Ohsugi and Hibi [29, 30, 31] and more recently by Hibi et al. [21].

The Graver basis is the set of primitive binomials of the toric ideal. Applications of the Graver basis to integer programming are discussed in Onn [34]. Since the Graver basis is a superset of any minimal Markov basis, the Graver basis is also a Markov basis and therefore connects every fiber. When the graph is restricted to be simple, however, a Markov basis does not necessarily connect all elements of every fiber. A recent result by Hara and Takemura [16] implies that the set of square-free elements of the Graver basis connects all elements of every fiber of simple graphs with a given vertex degree sequence. Thus if we have the Graver basis, we can sample graphs from any fiber, with or without the restriction that graphs are simple, in such a way that every graph in the fiber is generated with positive probability.

In the sequential importance sampling algorithm of [3], the underlying graph for the model was assumed to be complete. In our approach we can allow that some edges are absent from the beginning (structural zero edges in the terminology of contingency table analysis), such as the bipartite graph for the case of the Rasch model. Moreover, our algorithm can be applied not only for sampling simple graphs but also for sampling general undirected graphs without substantial adjustment. These are the advantages of the Graver basis.

The Graver basis for small graphs can be computed by a computer algebra system such as `4ti2`. For even moderate-sized graphs, however, it is difficult to compute the Graver basis via `4ti2` in a practical amount of time. In this thesis we provide a complete description of the Graver basis for an undirected graph. In general, the number of elements of the Graver basis is too large. We therefore construct an adaptive algorithm for sampling elements from the Graver basis, which is sufficient for constructing a connected Markov chain over any fiber. The recent paper of Reyes et al. [40] discusses the Graver basis for an undirected graph and gives a characterization of the Graver basis. In this thesis we give a different description of the Graver basis, which is more suitable for sampling elements from the Graver basis.

As stated above, Markov basis theory for the concrete statistical model has been developing very rapidly. When we study Markov bases for a specific problem, usually we are not faced with a single configuration, but rather with a series of configurations, possibly parameterized by a few parameters. For example, Markov bases associated with complete bipartite graphs $K_{I,J}$ (in statistical terms, the independence model of $I \times J$ two-way contingency tables) are parameterized by I and J. In this case, Markov bases consist of moves of degree two irrespective of I and J. In more general cases, some measure of complexity of Markov bases grows with the parameter and we are interested in bounding the growth.

There are some procedures to generate a series of configurations based on a given set of configurations. One of the most important constructions is the higher Lawrence lifting of a configuration, for which Santos and Sturmfels [43] described the growth by the notion of Graver complexity. Another important construction is the nested configuration of Ohsugi and Hibi [33], where a generated series of configurations basically inherit the desirable properties of the original configurations. In this thesis we define a new procedure to generate a series of configurations using fibers of a given configuration, which we call the base configuration. This construction is closely related to the higher Lawrence lifting of the base configuration and, exploiting this fact, we prove that the Markov degree of the configurations is bounded from above by the Markov complexity of the base configuration.

There are some intriguing problems, such as the complete bipartite graphs, in which moves of degree two form a Markov basis. When a minimal Markov basis contains a move of degree three or higher, it is usually very hard to control the measures of complexity of the Markov bases. A notable exception is the conjecture by Diaconis and Eriksson [6] that the Markov degree associated with the Birkhoff polytope is three, i.e., the toric ideal associated with the Birkhoff polytope is generated by binomials of degree at most three. We positively proved this conjecture in a previous work [47]. In view of [14] and [47], Christian Haase (personal communication, 2013) suggested that the Markov degree associated with two-way transportation polytopes and flow polytopes is three. Very recently, Domokos and Joó [9] gave a proof of this general conjecture. Adapting the arguments in [47], we provide a proof that the Markov degree associated with two-way transportation polytopes is three. Two-way transportation polytopes are important examples in our framework, since they are fibers of the incidence matrix of a complete bipartite graph.

Maximum likelihood estimation (MLE) is one of the fundamental parameter estimation methods in statistical inference. However, as in Kleinbaum and Klein [23] and Breslow and Day [4], the (unconditional) MLE has some bias for the matched data analysis. In such cases, the conditional MLE (CMLE) is desirable for an appropriate estimation. Traditionally the CMLE is used for estimation of the odds ratio in the common odds ratio model.

When we apply the CMLE for the generalized setting of the common odds ratio model, the conditional likelihood becomes a difficult problem due to the computational cost of the normalizing constant of conditional likelihood. Although there have been many previous studies on computation of the p-value for conditional testing, only a few results for CMLE are known, except for the case of estimating the common odds ratio for 2×2 tables. In this thesis we discuss the computation of the CMLE for the statistical models whose normalizing constant of the conditional likelihood can be dealt with the A-hypergeometric function, which is a solution of A-hypergeometric system.

The A-hypergeometric system is one of the most important classes of the system of linear partial differential equations introduced by Gel′fand et al. [12]. Since the A-hypergeometric system has fruitful mathematical properties, many researchers have studied it from analytical, combinatorial, and algorithmic viewpoints (see e.g. [42]). A difference version of the A-hypergeometric system was introduced by Ohara and Takayama [28].

Nakayama et al. [26] proposed the holonomic gradient descend (HGD) as an application of the Gröbner basis theory in D-modules to a numerical optimization problem. In directional statistics, computation of the normalizing constant can sometimes be difficult, since this constant has no closed form and requires numerical integration in general. In [26], Nakayama et al. derived a system of linear partial differential equations satisfied by the normalizing constant of the Fisher–Bingham distribution and applied the Gröbner basis technique in the ring of differential operators to evaluate the normalizing constant.

Since the motivation of [26] was the MLE for the Fisher–Bingham distribution, they name the method holonomic gradient descent (HGD). The key part of the HGD is numerical evaluation of the normalizing constant, and we can apply the framework called the holonomic gradient method (HGM) to various computational problems in statistics. Fortunately, because the class of holonomic functions covers a large class of normalizing constants of important statistical models, the theory and application of the HGM have been rapidly developed, especially for solving the computational problems in statistics (see [24, 25, 20, 45, 44]).

Once we have a holonomic system for computing the normalizing constant, we can obtain a Pfaffian system by using algebraic computational software such as `Risa/Asir` ([41]) via Gröbner basis computation, at least in principle. However, since the computational cost is huge in general and the singular locus of the Pfaffian system causes difficulties in the numerical computation, the theoretical study of Pfaffian systems and appropriate numerical implementation are important. Thus we mainly concentrate on the specific settings of the A-hypergeometric system derived from the two-way contingency tables. Our model is based on the classical independence model of two-way contingency tables. In this case the corresponding A-hypergeometric system is known as the Aomoto–Gel′fand system in the theory of hypergeometric functions.

Bibliography

[1] 4ti2 team. 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.

[2] S. Aoki and A. Takemura. Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *J. Stat. Comput. Simul.*, 75(10):787–812, 2005.

[3] J. Blitzstein and P. Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.*, 6(4):489–522, 2010.

[4] N. E. Breslow and N. E. Day, editors. *Statistical Methods in Cancer Research. Volume I - The Analysis of Case-control Studies.* International Agency for Research on Cancer (IARC Scientific Publication No. 32), Lyon, 1980.

[5] S. Chatterjee, P. Diaconis, and A. Sly. Random graphs with a given degree sequence. *Ann. Appl. Probab.*, 21(4):1400–1435, 2011.

[6] P. Diaconis and N. Eriksson. Markov bases for noncommutative Fourier analysis of ranked data. *J. Symbolic Comput.*, 41(2):182–195, 2006.

[7] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998.

[8] A. Dobra and S. Sullivant. A divide-and-conquer algorithm for generating Markov bases of multi-way tables. *Comput. Statist.*, 19(3):347–366, 2004.

[9] M. Domokos and D. Joó. On the equations and classification of toric quiver varieties. arXiv:1402.5096v1, 2014.

[10] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel, 2009.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[12] I. M. Gel′fand, A. V. Zelevinskiĭ, and M. M. Kapranov. Hypergeometric functions and toric varieties. *Funktsional. Anal. i Prilozhen.*, 23(2):12–26, 1989.

[13] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2:129–233, 2009.

[14] C. Haase and A. Paffenholz. Quadratic Gröbner bases for smooth 3×3 transportation polytopes. *J. Algebraic Combin.*, 30(4):477–489, 2009.

[15] S. J. Haberman. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Amer. Statist. Assoc.*, 83(402):555–560, 1988.

[16] H. Hara and A. Takemura. Connecting tables with zero-one entries by a subset of a markov basis. In *Algebraic Methods in Statistics and Probability II*, volume 516 of *Contemp. Math.*, pages 199–213. Amer. Math. Soc., Providence, RI, 2010.

[17] H. Hara, A. Takemura, and R. Yoshida. A Markov basis for conditional test of common diagonal effect in quasi-independence model for square contingency tables. *Comput. Statist. Data Anal.*, 53(4):1006–1014, 2009.

[18] H. Hara, A. Takemura, and R. Yoshida. Markov bases for two-way subtable sum problems. *J. Pure Appl. Algebra*, 213(8):1507–1521, 2009.

[19] H. Hara, S. Aoki, and A. Takemura. Minimal and minimal invariant Markov bases of decomposable models for contingency tables. *Bernoulli*, 16(1):208–233, 2010.

[20] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura. The holonomic gradient

method for the distribution function of the largest root of a Wishart matrix. *J. Multivariate Anal.*, 117:296–312, 2013.

[21] T. Hibi, K. Nishiyama, H. Ohsugi, and A. Shikama. Many toric ideals generated by quadratic binomials possess no quadratic Gröbner bases. *J. Algebra*, 408:138–146, 2014.

[22] P. Holland and S. Leinhardt. An exponential family of probability distribution for directed graphs. *J. Amer. Statist. Soc.*, 76(373):33–50, 1981.

[23] D. G. Kleinbaum and M. Klein. *Survival Analysis*. Statistics for Biology and Health. Springer, New York, third edition, 2012. A self-learning text.

[24] T. Koyama. A holonomic ideal which annihilates the Fisher-Bingham integral. *Funkcial. Ekvac.*, 56(1):51–61, 2013.

[25] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama. The holonomic rank of the Fisher-Bingham system of differential equations. *J. Pure Appl. Algebra*, 218(11): 2060–2071, 2014.

[26] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura. Holonomic gradient descent and its application to the Fisher–Bingham integral. *Adv. in Appl. Math.*, 47(3):639–658, 2011.

[27] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45 (2):167–256 (electronic), 2003.

[28] K. Ohara and N. Takayama. Holonomic rank of A-hypergeometric differential-difference equations. *J. Pure Appl. Algebra*, 213(8):1536–1544, 2009.

[29] H. Ohsugi and T. Hibi. Koszul bipartite graphs. *Adv. in Appl. Math.*, 22(1):25–28, 1999.

[30] H. Ohsugi and T. Hibi. Toric ideals generated by quadratic binomials. *J. Algebra*, 218(2):509–527, 1999.

[31] H. Ohsugi and T. Hibi. Indispensable binomials of finite graphs. *J. Algebra Appl.*, 4 (4):421–434, 2005.

[32] H. Ohsugi and T. Hibi. Two way subtable sum problems and quadratic Gröbner bases. *Proc. Amer. Math. Soc.*, 137(5):1539–1542, 2009.

[33] H. Ohsugi and T. Hibi. Toric rings and ideals of nested configurations. *J. Commut. Algebra*, 2(2):187–208, 2010.

[34] S. Onn. *Nonlinear Discrete Optimization*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2010. An algorithmic theory.

[35] J. Park and M. E. J. Newman. Statistical mechanics of networks. *Phys. Rev. E (3)*, 70(6):066117, 2004.

[36] S. Petrović, A. Rinaldo, and S. E. Fienberg. Algebraic statistics for a directed random graph model with reciprocation. In *Algebraic Methods in Statistics and Probability II*, volume 516 of *Contemp. Math.*, pages 261–283. Amer. Math. Soc., Providence, RI, 2010.

[37] G. Pistone and H. P. Wynn. Generalised confounding with Gröbner bases. *Biometrika*, 83(3):653–666, 1996.

[38] F. Rapallo. Markov bases and structural zeros. *J. Symbolic Comput.*, 41(2):164–172, 2006.

[39] G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press, Chicago, 1980.

[40] E. Reyes, C. Tatakis, and A. Thoma. Minimal generators of toric ideals of graphs. *Adv. in Appl. Math.*, 48(1):64–78, 2012.

[41] RisaAsir developing team. Risa/asir, a computer algebra system. Available at http://www.math.kobe-u.ac.jp/Asir/.

[42] M. Saito, B. Sturmfels, and N. Takayama. *Gröbner Deformations of Hypergeometric Differential Equations*, volume 6 of *Algorithms and Computation in Mathematics*.

Springer-Verlag, Berlin, 2000.

[43] F. Santos and B. Sturmfels. Higher Lawrence configurations. *J. Combin. Theory Ser. A*, 103(1):151–164, 2003.

[44] T. Sei and A. Kume. Calculating the normalising constant of the bingham distribution on the sphere using the holonomic gradient method. *Statistics and Computing*, pages 1–12, 2013.

[45] T. Sei, H. Shibata, A. Takemura, K. Ohara, and N. Takayama. Properties and applications of Fisher distribution on the rotation group. *J. Multivariate Anal.*, 116: 440–455, 2013.

[46] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys.*, 13:107–117, 1951.

[47] T. Yamaguchi, M. Ogawa, and A. Takemura. Markov degree of the Birkhoff model. *J. Algebraic Combin.*, 40(1):293–311, 2014.