**JOURNAL OF ALGEBRAIC STATISTICS**
**Issue in honor of Stephen E. Fienberg**
Volume 10, No. 1, 2019, pp. i-vi
https://jalgstat.library.iit.edu/
ISSN 1309-3452

# "Old Wine in New Bottles," and Some More New Wine – Stephen Fienberg's Influence on Algebraic Statistics

Sonja Petrović[1], Aleksandra Slavkovic[2], Ruriko Yoshida[3]

[1] *Managing Editor, Illinois Institute of Technology*
[2] *Guest Editor for this issue, Pennsylvania State University*
[3] *Former Managing Editor and Guest Editor for this issue, Naval Postgraduate School*

**Abstract.** Stephen (Steve) E. Fienberg (1942-2016) was an eminent statistician, whose impact on research, education and the practice of statistics, and many other fields is astonishing in its breadth. He was a visionary when it came to linking many different areas to address real scientific issues. He professed the importance of statistics in many disciplines, but recognized that true interdisciplinary work requires joining of the expertise across different areas, and it is in this spirit that he helped steer algebraic statistics toward becoming a thriving subject. Many of his favorite topics in the area are covered in this special issue. We are grateful to all authors for contributing to this volume to honor him and his influence on the field.

During the preparation of this issue, we learned about the tragic killing of his widow, Joyce Fienberg, during the Tree of Life Synagogue massacre in Pittsburgh, PA on October 27, 2018. *This issue is dedicated to their memory.*

Steve Fienberg published seven books and over 340 papers in leading journals in statistics, sociology, and machine learning, and wrote hundreds of editorials and reviews. He has made fundamental contributions to the foundation of statistical inference and Bayesian analysis, including analysis of categorical data, causal inference, mixed membership models, networks, forensic science, data privacy and disclosure limitation, statistics and the law, surveys, problems related to the federal statistics, and the list goes on. Geometric reasoning played an important role in his work, and he took joy in rediscovering old concepts from new points of view that gave them new interpretations and wider applicability, and this is how he came to avidly support establishment and growth of algebraic statistics.

His dissertation work on log-linear models used geometric arguments, and his first technical paper that appeared in the Annals of Mathematical Statistics in 1968 was on the geometry of an $r \times c$ contingency table [8]. Throughout his career, he continued to work on this subject, on the geometry of exponential families and shining new light on the potential

*Email addresses:* sonja.petrovic@iit.edu, sesa@psu.edu, ryoshida@nps.edu

of algebraic statistics, including problems on sampling from fibers based on models for contingency tables, exact inference using Markov bases, existence of maximum likelihood estimates in log-linear and network models, and others; see for example [2, 4, 3, 5], [6], [7], [9, 10, 11], [12], [13], [14, 15, 16], [18, 17, 19], [20], [21], [23], [24, 25], [26, 27], [28]. The classic book titled *Discrete Multivariate Analysis: Theory and Practice* that he co-authored with Y. Bishop and P. Holland in 1975 [1], contains some of those early arguments that many of us still go back to re-address problems with new tools from algebraic geometry and computational algebra.

Steve became more actively involved with algebraic statistics in the early 2000s through collaborations with Bernd Sturmfels, and supported student exchanges and collaborations, and organization of numerous workshops in the U.S.; up to that point the primary workshop for the field was GROSTAT (Gröbner bases in Statistics) organized by colleagues in Europe, who also produced the earliest book on algebraic statistics [22]. In 2007, the Institute of Mathematics and its Applications (IMA) in Minneapolis hosted a workshop on algebraic statistics as part of a year-long thematic program on applications of algebraic geometry, and Steve was to give a closing talk. Steve had spent the week browsing students' and postdocs' posters, interacting with mathematicians and statisticians, learning about their work. On the last day, his talk titled *'Algebraic Statistics and the Analysis of Contingency Tables: Old Wine in New Bottles?'* gave context to all of the work he saw during the workshop, put it in perspective using previous statistics literature, and connected it to other work in algebraic geometry. In essence, he recognized that algebraic statistics can make a difference in statistical methodology, and helped many young mathematicians in attendance see these important links to previous literature in statistics, and the present statisticians, the potential usefulness of algebraic tools. He taught many of us about *relevance* of our mathematical work in statistical theory and applications.

Over the years, he continued to support the field in many ways: attending dedicated conferences; helping in advising a program organization; collaborating with many of us on algebraic and geometric problems in statistics; relentlessly sending us references to statistics papers; and even advising some of the editors of this very journal during its early days.

Steve's last slide in that 2007 talk said the following: *"This is not the same old wine. There is an exciting new vintage about to be bottled! And it may last for years."* His contributions, guidance, and tireless mentoring continues to inspire many of us and we are thankful for the difference his presence has made in our lives, and the papers in this issue are reflective of that.

## Articles in This Issue

A number of the articles in this issue tackle problems that Steve has worked on, along with his students and collaborators (e.g., existence of maximum likelihood estimates (MLEs), model specification, and sampling from model fibers for testing fit), and some point to exciting directions into which the field is starting to branch out.

We open the technical part of the issue with "Generalized Fréchet Bounds for Cell

Entries in Multidimensional Contingency Tables" by Uhler and Richards, which offers a new solution to an old problem on computing bounds on the cells of contingency tables — Steve's favorite data summary object. Log-linear models capture associations and independence relations between categorical data, and in "Exact Tests to Compare Contingency Tables Under Quasi-independence and Quasi-symmetry", Bocci and Rapallo study the problem of model specification for a specific sub-class of log-linear models, and the question of testing their fit using Markov bases, which they derive, theoretically characterize, and suggest some open problems. Pham and Kateri in "Inference for Ordinal Log-Linear Models Based on Algebraic Statistics" consider another special case of log-linear models that account for ordinal nature of variables, and describe and implement use of algebraic methods for model fitting and testing within the context of two-way tables.

Both the latent class models and the network models are fundamentally tied to discrete data and graphs, and something else that Steve has worked on with vigor. In "Maximum Likelihood Estimation of the Latent Class Model Through Model Boundary Decomposition", Allman, Banos, Evans, Hoşten, Kubjas, Lemke, Rhodes, and Zwiernik address one of Steve's favorite problems on the issue of the MLE existence, and they use recent results from algebraic geometry to give a characterization of the boundary stratification of binary latent class models with a binary hidden variable, and address an issue raised by Steve back in the 70s on reliability of the expectation maximization (EM) algorithm. Lauritzen, Rinaldo, and Sadeghi, as the title of their paper "On Exchangeability in Network Models" suggests, derive representation theorems for exchangeable distributions on finite and infinite graphs using elementary arguments based on geometric and graph-theoretic concepts, illuminating some key differences for statistical models on graphs of a given size and those that define a consistent sequence of probability distributions on graphs of increasing size.

Fassino, Riccomagno, and Rogantin in "Cubature Rules and Expected Value of Some Complex Functions" unveil a novel connection between cubature rules and the algebraic statistics theory of fractional factorial design of experiments, similar in spirit to earlier connections found between Markov bases for contingency tables and design of experiments. They use this to compute the expected value of some complex valued random vectors, and thus offer extension to Gaussian case beyond discrete data. We close the issue with a question of complexity of Markov bases. While an avid supporter of this methodology, Steve was always also cautious, pointing out that algebra—being blind to observed data—produces many bases elements that are not applicable in practice. In "Strongly robust toric ideals in codimension 2", Sullivant answers an open problem about move complexity needed for tables that have sampling constraints.

## References

[1] Yvonne M. Bishop, Stephen E. Fienberg, and Paul W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Springer, 1974.

[2] Adrian Dobra and Stephen E. Fienberg. Bounds for cell entries in contingency tables

given marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences*, 97(22):11885–11892, 2000.

[3] Adrian Dobra and Stephen E. Fienberg. Bounds for cell entries in contingency tables induced by fixed marginal totals with applications to disclosure limitation. *Statistical Journal of the United Nations Economic Commission for Europe*, 18(4):363–371, 2001.

[4] Adrian Dobra and Stephen E. Fienberg. *Bounding entries in multi-way contingency tables given a set of marginal totals*, pages 3–16. Physica, Heidelberg, 2003.

[5] Adrian Dobra and Stephen E. Fienberg. The generalized shuttle algorithm. In *Algebraic and geometric methods in statistics*, pages 135–156. Cambridge University Press, 2010.

[6] Adrian Dobra, Stephen E. Fienberg, Alessandro Rinaldo, Aleksandra Slavković, and Yi Zhou. Algebraic statistics and contingency table problems: Log-linear models, likelihood estimation and disclosure limitation. In *In IMA Volumes in Mathematics and its Applications: Emerging Applications of Algebraic Geometry*, pages 63–88. Springer Science+Business Media, Inc, 2008.

[7] Nicholas Eriksson, Stephen E. Fienberg, Alessandro Rinaldo, and Seth Sullivant. Polyhedral conditions for the nonexistence of the mle for hierarchical log-linear models. *Journal of Symbolic Computation*, 41(2):222–233, 2006.

[8] Stephen E Fienberg. The geometry of an r× c contingency table. *The Annals of Mathematical Statistics*, 39(4):1186–1190, 1968.

[9] Stephen E. Fienberg. An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics*, 41(3):907–917, 1970.

[10] Stephen E. Fienberg. Fréchet and bonferroni bounds for multi-way tables of counts with applications to disclosure limitation. In *Statistical Data Protection (SDP'98) Proceedings*, pages 115–129, 1999.

[11] Stephen E. Fienberg. *The analysis of cross-classified categorical data*. Springer Science & Business Media, 2007.

[12] Stephen E. Fienberg, Patricia Hersh, Alessandro Rinaldo, and Yi Zhou. Maximum likelihood estimation in latent class models for contingency table data. In Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin, and Henry P. Wynn, editors, *Algebraic and Geometric Methods in Statistics*. Cambridge University Press, 2007.

[13] Stephen E. Fienberg, Sonja Petrović, and Alessandro Rinaldo. Algebraic statistics for $p_1$ random graph models: Markov bases and their uses. In S Sinharay and N. J. Dorans, editors, *Papers in Honor of Paul W. Holland, ETS*. Springer, 2010.

[14] Stephen E. Fienberg and Alessandro Rinaldo. Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, 137(11):3430–3445, 2007.

[15] Stephen E. Fienberg and Alessandro Rinaldo. Maximum likelihood estimation in log-linear models: Theory and algorithms. arXiv: 1104.3618, 2011.

[16] Stephen E. Fienberg and Alessandro Rinaldo. Maximum likelihood estimation in log-linear models. *The Annals of Statistics*, 40(2):996–1023, 2012.

[17] Stephen E. Fienberg and Aleksandra B. Slavković. Making the release of confidential data from multi-way tables count. *Chance*, 17(3):5–10, 2004.

[18] Stephen E. Fienberg and Aleksandra B. Slavković. Preserving the confidentiality of categorical statistical data bases when releasing information for association rules. *Data Mining and Knowledge Discovery*, 11(2):155–180, 2005.

[19] Stephen E. Fienberg and Aleksandra B. Slavković. A survey of statistical approaches to preserving confidentiality of contingency table entries. *Privacy-Preserving Data Mining*, pages 291–312, 2008.

[20] Steffen Lauritzen and Russell J. Steele. Disclosure limitation using perturbation and related methods for categorical data. *Journal of Official Statistics*, 14(1):485, 1998.

[21] Sonja Petrović, Alessandro Rinaldo, and Stephen E. Fienberg. Algebraic statistics for a directed random graph model with reciprocation. In M. Viana and H. Wynn, editors, *Algebraic Methods in Statistics and Probability II*, volume 516 of *Contemporary Mathematics*, pages 261–283. American Mathematical Society, Providence RI, 2010.

[22] Giovanni Pistone, Eva Riccomagno, and Henry Wynn. *Algebraic Statistics*, volume 89 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[23] Alessandro Rinaldo, Stephen E. Fienberg, and Yi Zhou. On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, 3:446–484, 2009.

[24] Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg. On the existence of the MLE for a directed random graph network model with reciprocation. Technical report, 2010. http://arxiv.org/abs/1010.0745.

[25] Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg. Maximum likelihood estimation in the Beta model. *Annals of Statistics*, 41(3):1085–1110, 2013.

[26] Aleksandra B. Slavković and Stephen E. Fienberg. Bounds for cell entries in two-way tables given conditional relative frequencies. *International Workshop on Privacy in Statistical Databases*, pages 30–43, 2004.

[27] Aleksandra B. Slavković and Stephen E. Fienberg. Algebraic geometry of $2 \times 2$ contingency tables. In *Algebraic and geometric methods in statistics*, pages 63–82. Cambridge University Press, 2009.

[28] Despina Stasi, Kayvan Sadeghi, Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg. Beta models for random hypergraphs with a given degree sequence. In *Proceedings of 21st International Conference on Computational Statistics*, 2014.