On the geometry of discrete exponential families with application to exponential random graph models

Alessandro Rinaldo, *Carnegie Mellon University*
Stephen Elliott Fienberg, *Carnegie Mellon University*
Yi Zhou, *Carnegie Mellon University*

Abstract

There has been an explosion of interest in statistical models for analyzing network data, and considerable interest in the class of exponential random graph (ERG) models, especially in connection with difficulties in computing maximum likelihood estimates. The issues associated with these difficulties relate to the broader structure of discrete exponential families. This paper re-examines the issues in two parts. First we consider the closure of k-dimensional exponential families of distribution with discrete base measure and polyhedral convex support $mathrm {P}$. We show that the normal fan of $mathrm{P}$ is a geometric object that plays a fundamental role in deriving the statistical and geometric properties of the corresponding extended exponential families. We discuss its relevance to maximum likelihood estimation, both from a theoretical and computational standpoint. Second, we apply our results to the analysis of ERG models. By means of a detailed example, we provide some characterization of the properties of ERG models, and, in particular, of certain behaviors of ERG models known as degeneracy.

AMS 2000 subject classifications: Primary 62F99; secondary 62F99.

Keywords: Exponential families, normal cone, maximum likelihood estimates, exponential random graphs.

Full Text: PDF

Rinaldo, Alessandro, Fienberg, Stephen Elliott, Zhou, Yi, On the geometry of discrete exponential families with application to exponential random graph models, Electronic Journal of Statistics, 3, (2009), 446-484 (electronic). DOI: 10.1214/08-EJS350.

References

Bardorff-Nielsen, O. (1978). Information and Exponential Families in Statistical Theory, New York: John Wiley & Sons, New York. MR0489333

Bardorff-Nielsen, O. (2006). Koopman–Darmois–Pitman Families, Encycploedia of Statistical Sciences.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA, 1975. Reprinted by Springer-Verlag, New York, 2007. MR2344876

Brown, L. (1986). Fundamentals of Statistical Exponential Families, IMS Lecture Notes-Monograph Series, Vol.9, Hayward, CA.

Csiszár, I. and Matúš, F. (2008). Generalized maximum likelihood estimates for exponential families, Probability Theory and Related Fields, 141, 213–246. MR2372970

Csiszár, I. and Matúš, F. (2005). Closure of exponential families, Annals of Probability, 33 (2), 582–600.

Csiszár, I. and Matúš, F. (2003). Information projection revisited, IEEE Transaction of Information Theory, 49 (6), 1474–1490.

Csiszár, I. and Matúš, F. (2001). Convex cores of measures, Studia Scientiarum Mathematicarum Hungarica, 38, 177–190. MR1877777

Eriksson, N., Fienberg, S.E., Rinaldo, A. and Sullivant S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models, Journal of Symbolic Computation, 41, 222–233. MR2197157

Fienberg, S.E. and Wasserman, S. (1981). An exponential family of probability distributions for directed graphs: comment, Journal of the American Statistical Association, 76(373), 54–57. MR0608176

Frank, O. and Strauss, D. (1986). Markov Graphs, Journal of the American Statistical Association, 81, 832–842. MR0860518

Geiger, A., Meek, C. and Sturmfels, B. (2006). On the toric algebra of graphical models, Annals of Statistics, 34(3), 1463–1492. MR2278364

Geyer, C. (2008). Likelihood Inference in Exponential Families and Directions of Recession, Technical Report TR 672, Department of Statistics, University of Minnesota.

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood calculations (with discussion), Journal of the Royal Statistical Society, Series B, 54, 657–699. MR1185217

Goldenberg, A., Zheng, A.X., Fienberg, S.E., and Airoldi, E.M. (2009). A survey of statistical network models, Foundations and Trends in Machine Learning, to appear.

Goodreau, S. (2007). Advances in exponential random graph (p^*) models applied to a large social network, Social Networks, 29, 231–248.

Haberman, S.J. (1981). An exponential family of probability distributions for directed graphs: comment, Journal of the American Statistical Association, 76(373), 60–61. MR0608176

Handcock, M.S., Hunter, D., Butts, C., Goodreau, S. and Morris, M. (2006). Statnet: An R Package for the Statistical Analysis and Simulation of Social Networks. Manual. University of Washington, http://www.csde. washington.edu/statnet.

Handcock, M.S. (2003). Assessing degeneracy in statistical models for social networks. Working paper 39, Center for Statistics and the Social Sciences, University of Washington.

Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76 (373), 33–65. MR0608176

Hunter, D.R., Goodreau, S.M. and Handcock, M.S. (2008). Goodness of fit of social network models, Journal of the American Statistical Association, 103(481), 248–258. MR2394635

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion), Journal of the Royal Statistical Society, Series B, 58, 619–678. MR1410182

Letac, G. (1992). Lectures on Natural Exponential Families and Their Variance Functions, Monografias de Matemática 50. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil. MR1182991

Moreno, J.L. (1934). Who Shall Survive? Nervous and Mental Disease Publishing Company, Washington, DC.

Rinaldo, A. (2006a). On maximum likelihood estimation in log-linear models, Technical Report 833, Department of Statistics, Carnegie Mellon University.

Rinaldo, A. (2006b). Computing Maximum Likelihood Estimates in Log-Linear Models, Technical Report 835, Department of Statistics, Carnegie Mellon University.

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007). An introduction to exponential random graph (p^*) models for social networks, Social Networks, 29, 171–191.

Robins, G, Snijers, T., Wang, P., Handcock, M. and Pattison, P. (2007). Recent developments in exponential random graph (p^*) models for social networks, Social Networks, 29, 192–215.

Rockafellar, R.T. (1970). Convex Analysis, Princeton University Press, Princeton. MR0274683

Schrijver, A. (1998). Theory of Linear and Integer Programming, Wiley & Sons, New York. MR0874114

Snijders, T.A.B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models, Journal of Social Structure, 3.

Strauss, D. and M. Ikeda (1990). Pseudolikelihood estimation for social networks, Journal of the American Statistical Association, 85, 204–212. MR1137368

Sturmfels, B. (1995). Gröbner Bases and Convex Polytopes, American Mathematical Society, Providence, RI.

Wasserman, S. and Pattison, P.E. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*, Psychometrika 61, 401–425. MR1424909

Wasserman, S.S. and Robins, G.L. (2004). An Introduction to Random Graphs, Dependence Graphs, and p^*, in Models and Methods in Social Network Analysis, eds. P. Carrington, J.S. and Wasserman, S.S., Cambridge: Cambridge University Press, New York.

Ziegler, G.M. (2001). Lectures on 0/1 polytopes, in Polytopes—Combinatorics and Computations, DMS Seminar, Band 29, G. Kalai and G.M. Ziegler editors, Birkhäuser. MR1785291

Ziegler, G.M. (1995). Lectures on Polytopes, Springer-Verlag, New York. MR1311028

Electronic Journal of Statistics. ISSN: 1935-7524