Electronic Journal of Statistics > Vol. 2 (2008) open journal systems

Estimation of Gaussian graphs by model selection

Christophe Giraud, *Université de Nice and INRA Jouy-en-Josas*

Abstract

We investigate in this paper the estimation of Gaussian graphs by model selection from a non-asymptotic point of view. We start from a n-sample of a Gaussian law $mathbb{P}_C$ in $mathbb{R}^p$ and focus on the disadvantageous case where n is smaller than p. To estimate the graph of conditional dependences of $mathbb{P}_C$, we introduce a collection of candidate graphs and then select one of them by minimizing a penalized empirical risk. Our main result assesses the performance of the procedure in a non-asymptotic setting. We pay a special attention to the maximal degree D of the graphs that we can handle, which turns to be roughly $n/(2 log p)$.

AMS 2000 subject classifications: Primary 62G08; secondary 15A52, 62J05.

Keywords: Gaussian graphical model, Random matrices, Model selection, Penalized empirical risk.

Full Text: PDF

Giraud, Christophe, Estimation of Gaussian graphs by model selection, Electronic Journal of Statistics, 2, (2008), 542-563 (electronic). DOI: 10.1214/08-EJS228.

References

[1] O. Banerjee, L.E. Ghaoui and A. d'Aspremont. Model selection through sparse maximum likelihood estimation. J. Machine Learning Research 9 (2008), 485–516.

[2] R. Baraniuk, M. Davenport, R. De Vore and M. Wakin. A simple proof of the restricted isometry property for random matrices. To appear in Constructive Approximation (2007)

[3] Y. Baraud. Model selection for regression on a random design. ESAIM Probab. Statist. 6 (2002), 127–146 (electronic). MR1918295

[4] Y. Baraud, C. Giraud and S. Huet. Gaussian model selection with unknown variance. To appear in the Annals of Statistics. http://arxiv.org/abs/math/0701250v1

[5] E. Candès and T. Tao. Decoding by linear programing. IEEE Trans. Inf. Theory 51 (2005) no. 12, 4203–4215. MR2243152

[6] A. Cohen, W. Dahmen and R. De Vore. Compressed sensing and the best k-term approximation. Preprint (2006) http://www.math.sc.edu/~devore/publications/CDDSensing_6.pdf

[7] K.R. Davidson and S.J. Szarek. Local operator theory, random matrices and Banach spaces. Handbook in Banach Spaces Vol I, ed. W. B. Johnson, J. Lindenstrauss, Elsevier (2001), 317–366. MR1863696

[8] M. Drton and M. Perlman. A sinful approach to Gaussian graphical model selection. Tech. Rep. 457 (2004), Dept. of Statistics, University of Washington, Seattle. http://www.stat.washington.edu/www/research/ reports/2004/tr457.pdf

[9] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, and M. West. Sparse graphical models for exploring gene expression data. J. Multivariate Analysis 90 (2004),196–212. MR2064941

[10] M. Drton and M. Perlman. Multiple testing and error control in Gaussian Graphical model selection. Statist. Sci. 22 (2007) no. 3, 430–449.

[11] J. Friedman, T. Hastie, R. Tibshirani. Sparse inverse covariance estimation with the lasso. Biostatistics 9 (2008) no. 3, 432–441.

[12] C. Giraud, S. Huet and N. Verzelen. In preparation.

[13] J.Z. Huang, N. Liu, M. Pourahmadi and L. Liu. Covariance matrix selection and estimation via penalised normal likelihood. Biometrika 93 no 1, (2006), 85–98 MR2277742

[14] H. Kishino and P.J. Waddell. Correspondence analysis of genes and tissue types and finding genetic links from microarray data. Genome Informatics 11 (2000), 83–95.

[15] N. Meinshausen and P. Bühlmann. High dimensional graphs and variable selection with the lasso. Annals of Statistics 34 (2006), 1436–1462. MR2278363

[16] J. Schäfer and K. Strimmer. An empirical bayes approach to inferring large-scale gene association networks. Bioinformatics 21 (2005), 754–764.

[17] N. Verzelen and F. Villers. Test of neighborhood for Gaussian graphical models. To appear in the Annals of Statistics.

[18] F. Villers, B. Schaeffer, C. Bertin, and S. Huet. Assessing the validity domains of graphical Gaussian models in order to infer relationships among components of complex biological systems. Technical Report, INRA (2008).

[19] A. Wille and P. Bühlmann. Low-order conditional independence graphs for inferring genetic networks. Stat. Appl. Genet. Mol. Biol. 5 (2006). MR2221304

[20] W. Wu and Y. Ye. Exploring gene causal interactions using an enhanced constraint-based method. Pattern Recognition 39 (2006) 2439–2449.

[21] M. Yuan and Y. Lin Model selection and estimation in the Gaussian graphical model. Biometrika 94 (2007), 19–35. MR2367824