STATISTICAL SCIENCE

Volume 34, Number 3 August 2019

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 34, Number 3, August 2019. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Statistical Science Volume 34, Number 3 (361–521) August 2019

Volume 34 **Number 3** **August 2019**

Statistical Science
2019, Vol. 34, No. 3, 361–390
https://doi.org/10.1214/19-STS697

ROS Regression: Integrating Regularization with Optimal Scaling Regression

Jacqueline J. Meulman, Anita J. van der Kooij and Kevin L. W. Duisters

Abstract. We present a methodology for multiple regression analysis that deals with categorical variables (possibly mixed with continuous ones), in combination with regularization, variable selection and high-dimensional data ($P \gg N$). Regularization and optimal scaling (OS) are two important extensions of ordinary least squares regression (OLS) that will be combined in this paper. There are two data analytic situations for which optimal scaling was developed. One is the analysis of categorical data, and the other the need for transformations because of nonlinear relationships between predictors and outcome. Optimal scaling of categorical data finds quantifications for the categories, both for the predictors and for the outcome variables, that are optimal for the regression model in the sense that they maximize the multiple correlation. When nonlinear relationships exist, nonlinear transformation of predictors and outcome maximize the multiple correlation in the same way. We will consider a variety of transformation types; typically we use step functions for categorical variables, and smooth (spline) functions for continuous variables. Both types of functions can be restricted to be monotonic, preserving the ordinal information in the data. In combination with optimal scaling, three popular regularization methods will be considered: Ridge regression, the Lasso and the Elastic Net. The resulting method will be called ROS Regression (Regularized Optimal Scaling Regression). The OS algorithm provides straightforward and efficient estimation of the regularized regression coefficients, automatically gives the Group Lasso and Blockwise Sparse Regression, and extends them by the possibility to maintain ordinal properties in the data. Extended examples are provided.

Key words and phrases: Lasso and Elastic Net regularization for nominal and ordinal data, monotonic group Lasso, regularization for categorical high-dimensional data, optimal scaling, linearization of nonlinear relationships, monotonic step functions and splines.

REFERENCES

ANGOFF, C. and MENCKEN, H. L. (1931). The worst American state. *American Mercury* **24** 1–16. 175–188, 355–31.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*: *The Theory and Application of Isotonic Regression*. Wiley, New York.

BOCK, R. D. (1960). *Methods and Applications of Optimal Scaling*. *Report* 25. L. L. Thurstone Lab, Univ. North Carolina, Chapel Hill.

Jacqueline J. Meulman is Professor, Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands (e-mail: jmeulman@math.leidenuniv.nl) and Adjunct Professor, Department of Statistics, Stanford University, Stanford, Californoa 94305, USA (e-mail: jmeulman@stanford.edu). Anita J. van der Kooij is Senior Researcher, Institute of Psychology, Division of Methodology and Statistics, Wassenaarseweg 52, 2323 CA Leiden, The Netherlands (e-mail: kooij@leidenuniv.nl). Kevin L.W. Duisters is Ph.D. candidate at the Mathematical Institute, Leiden University, Leiden, The Netherlands (e-mail: k.l.w.duisters@math.leidenuniv.nl).

BREIMAN, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37** 373–384. MR1365720

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619. MR0803258

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth, Belmont, CA. MR0726392

BUJA, A. (1990). Remarks on functional canonical variates, alternating least squares methods and ACE. *Ann. Statist.* **18** 1032–1069. MR1062698

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249

CAI, T. T. and ZHANG, L. (2018). High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Statist. Sinica* **28** 963–993. MR3791096

CÉA, J. and GLOWINSKI, R. (1973). Sur des méthodes d'optimisation par relaxation. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge* **7** 5–31. MR0367765

CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094

CHOULDECHOVA, A. and HASTIE, T. J. (2015). Generalized Additive Model Selection. Available at arXiv:1506.03850 [stat.ML].

DAUBECHIES, I., DEFRISE, M. and DE MOL, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* **57** 1413–1457. MR2077704

DE LEEUW, J., YOUNG, F. W. and TAKANE, Y. (1976). Additive structure in qualitative data. *Psychometrika* **41** 471–503.

DETTE, H., VAN HECKE, R. and VOLGUSHEV, S. (2014). Some comments on copula-based regression. *J. Amer. Statist. Assoc.* **109** 1319–1324. MR3265699

DHILLON, I. S. (2008). The log-determinant divergence and its applications. Paper presented at the Householder Symposium XVII, Zeuthen, Germany.

EFRON, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Amer. Statist. Assoc.* **78** 316–331. MR0711106

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

FRANK, I. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–148.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.* **19** 1–141. MR1091842

FRIEDMAN, J. HASTIE, T. J. and TIBSHIRANI, R. J. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **22** 1548–7660.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2012). glmnet: Lasso and elastic-net regularized generalized linear models. Available at http://CRAN.R-project.org/package=glmnet. R package version 1.9-5.

FRIEDMAN, J. H. and MEULMAN, J. J. (2003). Prediction with multiple additive regression trees with application in epidemiology. *Stat. Med.* **22** 1365–1381.

FRIEDMAN, J. H. and POPESCU, B. E. (2004). Gradient directed regularization for linear regression and classification Technical report, Dept. Statistics, Stanford Univ., Stanford, CA.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. MR0650892

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. MR2415737

FU, W. J. (1998). Penalized regressions: The Bridge versus the Lasso. *J. Comput. Graph. Statist.* **7** 397–416. MR1646710

GENEST, C. and NEŠLEHOVÁ, J. (2007). A primer on copulas for count data. *Astin Bull.* **37** 475–515. MR2422797

GIFI, A. (1981). Nonlinear multivariate analysis. Unpublished Manuscript. Department of Data Theory, Leiden University, Leiden.

GIFI, A. (1990). *Nonlinear Multivariate Analysis*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR1076188

GOLUB, G. H., HEATH, M. and WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–223. MR0533250

GROENEN, P. J. F., VAN OS, B.-J. and MEULMAN, J. J. (2000). Optimal scaling by alternating length-constrained nonnegative least squares, with application to distance-based analysis. *Psychometrika* **65** 511–524. MR1849280

GURIN, L. G., POLJAK, B. T. and RAĬK, È. V. (1967). Projection methods for finding a common point of convex sets. *Zh. Vychisl. Mat. Mat. Fiz.* **7** 1211–1228. MR0232225

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HASTIE, T., TIBSHIRANI, R. and BUJA, A. (1994). Flexible discriminant analysis by optimal scoring. *J. Amer. Statist. Assoc.* **89** 1255–1270. MR1310220

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294

HAYASHI, C. (1952). On the prediction of phenomena from qualitive data and the quantification of qualitative data from the mathematico-statistical point of view. *Ann. Inst. Statist. Math.* **1952** 93–96.

HOERL, A. E. and KENNARD, R. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

HOERL, A. E. and KENNARD, R. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics* **12** 69–82.

HOFF, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.* **1** 265–283. MR2393851

IBM CORP. (2010). *IBM SPSS Statistics* 19.0 *Algorithms*. IBM Corp., Amonk, NY.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

KIM, Y., KIM, J. and KIM, Y. (2006). Blockwise sparse regression. *Statist. Sinica* **16** 375–390. MR2267240

KOLEV, N. and PAIVA, D. (2009). Copula-based regression models: A survey. *J. Statist. Plann. Inference* **139** 3847–3856. MR2553771

KRUSKAL, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29** 115–129. MR169713

KRUSKAL, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *J. Roy. Statist. Soc. Ser. B* **27** 251–263. MR0195212

MASAROTTO, G. and VARIN, C. (2012). Gaussian copula marginal regression. *Electron. J. Stat.* **6** 1517–1549. MR2988457

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769

MEULMAN, J. J. (1986). *A Distance Approach to Nonlinear Multivariate Analysis*. DSWO Press, Leiden.

MEULMAN, J. J., HEISER, W. J. and SPSS (1998). SPSS Categories 8.0, Chicago, IL SPSS Inc.

MEULMAN, J. J., HEISER, W. J. and SPSS INC. (2010). *IBM SPSS Categories* 19. IBM Corp., Amonk, NY.

MEULMAN, J. J., ZEPPA, P., BOON, M. E. and RIETVELD, W. J. (1992). Prediction of various grades of cervical neoplasia on plastic-embedded cytobrush samples. Discriminant analysis with qualitative and quantitative predictors. *Anal. Quant. Cytol. Histol.* **14** 60–72.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications. Mathematical Expositions* **24**. Univ. Toronto Press, Toronto. MR0600656

NISHISATO, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Lawrence Erlbaum, Hillsdale, NJ.

NOH, H., EL GHOUCH, A. and BOUEZMARNI, T. (2013). Copula-based regression estimation and inference. *J. Amer. Statist. Assoc.* **108** 676–688. MR3174651

OBERHOFER, W. and KMENTA, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica* **42** 579–590. MR0440805

OSBORNE, M. R., PRESNELL, B. and TURLACH, B. A. (2000a). A new approach to variable selection in least squares problems. *IMA J. Numer. Anal.* **20** 389–403. MR1773265

OSBORNE, M. R., PRESNELL, B. and TURLACH, B. A. (2000b). On the LASSO and its dual. *J. Comput. Graph. Statist.* **9** 319–337. MR1822089

PARSA, R. A. and KLUGMAN, S. A. (2011). Copula regression. *Proc. Casualty Actuar. Soc.* **5** 45–54.

PERKINS, S., LACKER, K. and THEILER, J. (2003). Grafting: Fast, incremental feature selection by gradient descent in function space. *J. Mach. Learn. Res.* **3** 1333–1356. MR2020763

PITT, M., CHAN, D. and KOHN, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* **93** 537–554. MR2261441

R CORE TEAM (2017). R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

RAMSAY, J. O. (1988). Monotone regression splines in action (with discussion). *Statist. Sci.* **4** 425–441.

SAS/STAT (1990). *User's Guide, Version* 6 **2**. SAS Institute Inc., Cary NC.

SKLAR, M. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

SONG, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scand. J. Stat.* **27** 305–320. MR1777506

TIBSHIRANI, R. (1988). Estimating transformations for regression via additivity and variance stabilization. *J. Amer. Statist. Assoc.* **83** 394–405. MR0971365

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TIBSHIRANI, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 273–282. MR2815776

TIKHONOV, A. N. (1943). On the stability of inverse problems. *C. R. (Dokl.) Acad. Sci. URSS* **39** 176–179. MR0009685

TRIVEDI, P. K. and ZIMMER, D. M. (2005). Copula modeling: An introduction for practitioners. *Found. Trends Econom.* **1** 1–111.

TSENG, P. (1988). Coordinate Ascent for Maximizing Nondifferentiable Concave Functions. Technical Report LIDS-P, 1840, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA.

TSENG, P. (2001). Convergence of a block method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** 475–494. MR1835069

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285

VAN DER KOOIJ, A. J. (2007). Prediction accuracy and stability of regression with optimal scaling transformations. Thesis, Leiden Univ. Available at https://openaccess.leidenuniv.nl/handle/1887/12096.

WAINER, H. and THISSEN, D. (1981). Graphical data analysis. *Annu. Rev. Psychol.* **32** 191–241.

WALBERG, H. J. and RASHER, S. P. (1977). The ways schooling makes a difference. *Phi Delta Kappan* **58** 703–707.

WINSBERG, S. and RAMSAY, J. O. (1980). Monotonic transformations to additivity using splines. *Biometrika* **67** 669–674. MR0601105

WU, T. T. and LANGE, K. (2008). Coordinate descent algorithms for lasso penalized regression. *Ann. Appl. Stat.* **2** 224–244. MR2415601

YOUNG, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika* **46** 357–388. MR0668307

YOUNG, F. W., DE LEEUW, J. and TAKANE, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with Optimal Scaling features. *Psychometrika* **41** 505–529.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574

ZANGWILL, W. I. (1969/70). Convergence conditions for nonlinear programming algorithms. *Manage. Sci.* **16** 1–13. MR0302199

ZHAO, P. and YU, B. (2007). Stagewise lasso. *J. Mach. Learn. Res.* **8** 2701–2726. MR2383572

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

Statistical Science
2019, Vol. 34, No. 3, 391–404
https://doi.org/10.1214/19-STS698

An Overview of Semiparametric Extensions of Finite Mixture Models

Sijia Xiang, Weixin Yao and Guangren Yang

Abstract. Finite mixture models have offered a very important tool for exploring complex data structures in many scientific areas, such as economics, epidemiology and finance. Semiparametric mixture models, which were introduced into traditional finite mixture models in the past decade, have brought forth exciting developments in their methodologies, theories, and applications. In this article, we not only provide a selective overview of the newly-developed semiparametric mixture models, but also discuss their estimation methodologies, theoretical properties if applicable, and some open questions. Recent developments are also discussed.

Key words and phrases: EM algorithm, mixture models, mixture regression models, semiparametric mixture models.

REFERENCES

AL MOHAMAD, D. and BOUMAHDAF, A. (2018). Semiparametric two-component mixture models when one component is defined through linear constraints. *IEEE Trans. Inform. Theory* **64** 795–830. MR3762592

ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554

BALABDAOUI, F. (2017). Revisiting the Hodges–Lehmann estimator in a location mixture model: Is asymptotic normality good enough? *Electron. J. Stat.* **11** 4563–4595. MR3724489

BALABDAOUI, F. and DOSS, C. R. (2018). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* **24** 1053–1071. MR3706787

BENAGLIA, T., CHAUVEAU, D. and HUNTER, D. R. (2009). An EM-like algorithm for semi- and nonparametric estimation in multivariate mixtures. *J. Comput. Graph. Statist.* **18** 505–526. MR2749842

BORDES, L., CHAUVEAU, D. and VANDEKERKHOVE, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Statist. Data Anal.* **51** 5429–5443. MR2370882

BORDES, L., DELMAS, C. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scand. J. Stat.* **33** 733–752. MR2300913

BORDES, L., KOJADINOVIC, I. and VANDEKERKHOVE, P. (2013). Semiparametric estimation of a two-component mixture of linear regressions in which one component is known. *Electron. J. Stat.* **7** 2603–2644. MR3121625

BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356

BORDES, L. and VANDEKERKHOVE, P. (2010). Semiparametric two-component mixture model with a known component: An asymptotically normal estimator. *Math. Methods Statist.* **19** 22–41. MR2682853

BUTUCEA, C., NGUEYEP TZOUMPE, R. and VANDEKERKHOVE, P. (2017). Semiparametric topographical mixture models with symmetric errors. *Bernoulli* **23** 825–862. MR3606752

BUTUCEA, C. and VANDEKERKHOVE, P. (2014). Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.* **41** 227–239. MR3181141

CAO, J. and YAO, W. (2012). Semiparametric mixture of binomial regression with a degenerate component. *Statist. Sinica* **22** 27–46. MR2933166

CHANG, G. T. and WALTHER, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Statist. Data Anal.* **51** 6242–6251. MR2408591

CHAUVEAU, D., HUNTER, D. R. and LEVINEZ, M. (2015). Estimation for conditional independence multivariate finite mixture models. *Stat. Surv.* **9** 1–31.

CHEE, C.-S. and WANG, Y. (2013). Estimation of finite mixtures with symmetric components. *Stat. Comput.* **23** 233–249. MR3016941

CHEN, H., CHEN, J. and KALBFLEISCH, J. D. (2004). Testing for a finite mixture model with two components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 95–115. MR2035761

Sijia Xiang is Associate Professor, School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang, 310018, P.R. China (e-mail: sjxiang@zufe.edu.cn). Weixin Yao is Associate Professor, Department of Statistics, University of California, Riverside, California, USA (e-mail: weixin.yao@ucr.edu). Guangren Yang is Associate Professor, Department of Statistics, School of Economics, Jinan University, Guangzhou, 510632, China (e-mail: tygr@jnu.edu.cn).

CHEN, J. and LI, P. (2009). Hypothesis test for normal mixture models: The EM approach. *Ann. Statist.* **37** 2523–2542. MR2543701

DACUNHA-CASTELLE, D. and GASSIAT, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *Ann. Statist.* **27** 1178–1209. MR1740115

DANNEMANN, J., HOLZMANN, H. and LEISTER, A. (2014). Semiparametric hidden Markov models: Identifiability and estimation. *Comput. Statist.* **6** 418–425.

DE CASTRO, Y., GASSIAT, É. and LE CORFF, S. (2017). Consistent estimation of the filtering and marginal smoothing distributions in nonparametric hidden Markov models. *IEEE Trans. Inform. Theory* **63** 4758–4777. MR3683535

DZIAK, J. J., LI, R., TAN, X., SHIFFMAN, S. and SHIYKO, M. P. (2015). Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. *Psychol. Methods* **20** 444–469.

FAICEL, C. (2016). Unsupervised learning of regression mixture models with unknown number of components. *J. Stat. Comput. Simul.* **86** 2308–2334. MR3502164

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

GASSIAT, E. (2017). Mixtures of nonparametric components and hidden Markov models. In *Handbook of Mixture Analysis* (S. Fruhwirth-Schnatter, G. Celeux and C. P. Robert, eds.) 343–360. CRC Press, Boca Raton, FL. MR3889699

GASSIAT, E., CLEYNEN, A. and ROBIN, S. (2016). Inference in finite state space non parametric hidden Markov models and applications. *Stat. Comput.* **26** 61–71. MR3439359

GASSIAT, E. and ROUSSEAU, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli* **22** 193–212. MR3449780

GASSIAT, E., ROUSSEAU, J. and VERNET, E. (2018). Efficient semiparametric estimation and model selection for multidimensional mixtures. *Electron. J. Stat.* **12** 703–740. MR3769193

HALL, P. and ZHOU, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31** 201–224. MR1962504

HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171

HOHMANN, D. and HOLZMANN, H. (2013). Semiparametric location mixtures with distinct components. *Statistics* **47** 348–362. MR3043704

HU, H., WU, Y. and YAO, W. (2016). Maximum likelihood estimation of the mixture of log-concave densities. *Comput. Statist. Data Anal.* **101** 137–147. MR3504841

HU, H., YAO, W. and WU, Y. (2017). The robust EM-type algorithms for log-concave mixtures of regression models. *Comput. Statist. Data Anal.* **111** 14–26. MR3630215

HUANG, M., LI, R. and WANG, S. (2013). Nonparametric mixture of regression models. *J. Amer. Statist. Assoc.* **108** 929–941. MR3174674

HUANG, M. and YAO, W. (2012). Mixture of regression models with varying mixing proportions: A semiparametric approach. *J. Amer. Statist. Assoc.* **107** 711–724. MR2980079

HUANG, M., LI, R., WANG, H. and YAO, W. (2014). Estimating mixture of Gaussian processes by kernel smoothing. *J. Bus. Econom. Statist.* **32** 259–270. MR3207838

HUANG, M., WANG, S., WANG, H. and JIN, T. (2018a). Maximum smoothed likelihood estimation for a class of semiparametric Pareto mixture densities. *Stat. Interface* **11** 31–40. MR3690796

HUANG, M., WANG, S., YAO, W. and CHEN, Y. (2018b). Statistical inference and applications of mixture of varying coefficient models. *Scand. J. Stat.* **45** 618–643. MR3858949

HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251. MR2332275

HUNTER, D. R. and YOUNG, D. S. (2012). Semiparametric mixtures of regressions. *J. Nonparametr. Stat.* **24** 19–38. MR2885823

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981

JACOBS, R. A., PENG, F. and TANNER, M. A. (1997). A Bayesian approach to model selection in hierarchical mixtures-of-experts architectures. *Neural Netw.* **10** 231–241.

LEMDANI, M. and PONS, O. (1999). Likelihood ratio tests in contamination models. *Bernoulli* **5** 705–719. MR1704563

LEROUX, B. G. (1992). Consistent estimation of a mixing distribution. *Ann. Statist.* **20** 1350–1360. MR1186253

LEVINE, M., HUNTER, D. R. and CHAUVEAU, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika* **98** 403–416. MR2806437

LINDSAY, B. G. (1983). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94. MR0684866

MA, Y. and YAO, W. (2015). Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electron. J. Stat.* **9** 444–474. MR3326131

MA, Y., WANG, S., XU, L. and YAO, W. (2018). Semiparametric mixture regression with unspecified error distributions. Available at arXiv:1811.01117.

MAIBORODA, R. and SUGAKOVA, O. (2011). Generalized estimating equations for symmetric distributions observed with admixture. *Comm. Statist. Theory Methods* **40** 96–116. MR2747201

MCLACHLAN, G. J., BEAN, R. W. and JONES, L. B.-T. (2006). A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics* **22** 1608–1615.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley Interscience, New York. MR1789474

MONTUELLE, L. and LE PENNEC, E. (2014). Mixture of Gaussian regressions model with logistic weights, a penalized maximum likelihood approach. *Electron. J. Stat.* **8** 1661–1695. MR3263134

NGUYEN, V. H. and MATIAS, C. (2014). On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. *Scand. J. Stat.* **41** 1167–1194. MR3277044

PATRA, R. K. and SEN, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 869–893. MR3534354

POMMERET, D. and VANDEKERKHOVE, P. (2018). Semiparametric false discovery rate model Gaussianity test. Available at https://hal.archives-ouvertes.fr/hal-01868272.

ROGER, G. and POL, S. (1991). *Stochastic Finite Elements*: *A Special Approach*. Springer, Berlin.

RUFIBACH, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *J*. *Stat*. *Comput*. *Simul*. **77** 561–574. MR2407642

SONG, S., NICOLAE, D. L. and SONG, J. (2010). Estimating the mixing proportion in a semiparametric mixture model. *Comput*. *Statist*. *Data Anal*. **54** 2276–2283. MR2720488

TAN, X., SHIYKO, M. P., LI, R., LI, Y. and DIERKER, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychol*. *Methods* **17** 61–77.

VANDEKERKHOVE, P. (2013). Estimation of a semiparametric mixture of regressions model. *J*. *Nonparametr*. *Stat*. **25** 181–208. MR3039977

VON NEUMANN, J. (1931). Die Eindeutigkeit der Schrödingerschen Operatoren. *Math*. *Ann*. **104** 570–578. MR1512685

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J*. *Amer*. *Statist*. *Assoc*. **97** 508–513. MR1941467

WANG, Y. (2010). Maximum likelihood computation for fitting semiparametric mixture models. *Stat*. *Comput*. **20** 75–86. MR2578078

WANG, S., YAO, W. and HUANG, M. (2014). A note on the identifiability of nonparametric and semiparametric mixtures of GLMs. *Statist*. *Probab*. *Lett*. **93** 41–45. MR3244553

WANG, S., HUANG, M., WU, X. and YAO, W. (2016). Mixture of functional linear models and its application to CO_2-GDP functional data. *Comput*. *Statist*. *Data Anal*. **97** 1–15. MR3447032

WU, Q. and YAO, W. (2016). Mixtures of quantile regressions. *Comput*. *Statist*. *Data Anal*. **93** 162–176. MR3406203

WU, J., YAO, W. and XIANG, S. (2017). Computation of an efficient and robust estimator in a semiparametric mixture model. *J*. *Stat*. *Comput*. *Simul*. **87** 2128–2137. MR3656095

XIANG, S. and YAO, W. (2017). Semiparametric mixtures of regressions with single-index for model based clustering. Available at arXiv:1708.04142.

XIANG, S. and YAO, W. (2018). Semiparametric mixtures of nonparametric regressions. *Ann*. *Inst*. *Statist*. *Math*. **70** 131–154. MR3742821

XIANG, S., YAO, W. and SEO, B. (2016). Semiparametric mixture: Continuous scale mixture approach. *Comput*. *Statist*. *Data Anal*. **103** 413–425. MR3522641

XIANG, S., YAO, W. and WU, J. (2014). Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canad*. *J*. *Statist*. **42** 246–267. MR3208338

YAO, F., FU, Y. and LEE, T. C. M. (2011). Functional mixture regression. *Biostatistics* **12** 341–353.

YOUNG, D. S. (2014). Mixtures of regressions with changepoints. *Stat*. *Comput*. **24** 265–281. MR3165553

YOUNG, D. S. and HUNTER, D. R. (2010). Mixtures of regressions with predictor-dependent mixing proportions. *Comput*. *Statist*. *Data Anal*. **54** 2253–2266. MR2720486

Statistical Science
2019, Vol. 34, No. 3, 405–427
https://doi.org/10.1214/19-STS700

Lasso Meets Horseshoe: A Survey

Anindya Bhadra, Jyotishka Datta, Nicholas G. Polson and Brandon Willard

Abstract. The goal of this paper is to contrast and survey the major advances in two of the most commonly used high-dimensional techniques, namely, the Lasso and horseshoe regularization. Lasso is a gold standard for predictor selection while horseshoe is a state-of-the-art Bayesian estimator for sparse signals. Lasso is fast and scalable and uses convex optimization whilst the horseshoe is nonconvex. Our novel perspective focuses on three aspects: (i) theoretical optimality in high-dimensional inference for the Gaussian sparse model and beyond, (ii) efficiency and scalability of computation and (iii) methodological development and performance.

Key words and phrases: Global-local priors, horseshoe, horseshoe+, hyperparameter tuning, Lasso, regression, regularization, sparsity.

REFERENCES

ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B* **36** 99–102. MR0359122

ARMAGAN, A., CLYDE, M. and DUNSON, D. B. (2011). Generalized beta mixtures of Gaussians. In *Advances in Neural Information Processing Systems* 523–531.

ARMAGAN, A., DUNSON, D. B. and LEE, J. (2013). Generalized double Pareto shrinkage. *Statist. Sinica* **23** 119–143. MR3076161

ARMAGAN, A., DUNSON, D. B., LEE, J., BAJWA, W. U. and STRAWN, N. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika* **100** 1011–1018. MR3142348

BAI, R. and GHOSH, M. (2017). The inverse gamma–gamma prior for optimal posterior contraction and multiple hypothesis testing. Available at arXiv:1710.04369.

BELITSER, E. and NURUSHEV, N. (2015). Needles and straw in a haystack: Robust confidence for possibly sparse sequences. Available at arXiv:1511.01803.

BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324

BERZUINI, C., GUO, H., BURGESS, S. and BERNARDINELLI, L. (2016). Mendelian randomization with poor instruments: A Bayesian approach. Available at arXiv:1608.02990 [math, stat], Aug. arXiv:1608.02990.

BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2016a). Global–local mixtures. Avalable at arXiv:1604.07487.

BHADRA, A., DATTA, J., LI, Y., POLSON, N. G. and WILLARD, B. (2016b). Prediction risk for the horseshoe regression. Available at arXiv:1605.04796.

BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2016c). Default Bayesian analysis with global–local shrinkage priors. *Biometrika* **103** 955–969. MR3620450

BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2017a). Horseshoe regularization for feature subset selection. Available at arXiv:1702.07400.

BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2017b). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.* **12** 1105–1131. MR3724980

BHATTACHARYA, A., CHAKRABORTY, A. and MALLICK, B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika* **103** 985–991. MR3620452

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *J. Amer. Statist. Assoc.* **110** 1479–1490. MR3449048

BIEN, J., TAYLOR, J. and TIBSHIRANI, R. (2013). A LASSO for hierarchical interactions. *Ann. Statist.* **41** 1111–1141. MR3113805

BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. MR2850212

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. MR2807761

Anindya Bhadra is Associate Professor, Department of Statistics, Purdue University, 250 N. University St., West Lafayette, Indiana 47907, USA (e-mail: bhadra@purdue.edu). Jyotishka Datta is Assistant Professor, Department of Mathematical Sciences, University of Arkansas, 1 University of Arkansas, Fayetteville, Arkansas 72704, USA (e-mail: jd033@uark.edu). Nicholas G. Polson is Professor, Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, Illinois 60637, USA (e-mail: ngp@chicagobooth.edu). Brandon Willard is Researcher, Booth School of Business, University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, Illinois 60637, USA (e-mail: bwillard@uchicago.edu).

CANDÈS, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris* **346** 589–592. MR2412803

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **56** 2053–2080. MR2723472

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2009). Handling sparsity via the horseshoe. *J. Mach. Learn. Res.* **5** 73–80.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751

CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874

CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077

CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. MR2847974

CHERNOZHUKOV, V., HANSEN, C. and LIAO, Y. (2017). A lava attack on the recovery of sums of dense and sparse signals. *Ann. Statist.* **45** 39–76. MR3611486

CUTILLO, L., JUNG, Y. Y., RUGGERI, F. and VIDAKOVIC, B. (2008). Larger posterior mode wavelet thresholding and applications. *J. Statist. Plann. Inference* **138** 3758–3773. MR2455964

DATTA, J. and DUNSON, D. B. (2016). Bayesian inference on quasi-sparse count data. *Biometrika* **103** 971–983. MR3620451

DATTA, J. and GHOSH, J. K. (2013). Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Anal.* **8** 111–131. MR3036256

DATTA, J. and GHOSH, J. K. (2015). In search of optimal objective priors for model selection and estimation. In *Current Trends in Bayesian Methodology with Applications* 225–243. CRC Press, Boca Raton, FL. MR3644673

DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* **52** 1289–1306. MR2241189

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464

DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. MR1157714

EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

FAULKNER, J. R. and MININ, V. N. (2015). Bayesian trend filtering: Adaptive temporal smoothing with shrinkage priors.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. MR2415737

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24** 997–1016. MR3253850

GEORGE, E. I. (2000). The variable selection problem. *J. Amer. Statist. Assoc.* **95** 1304–1308. MR1825282

GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747. MR1813972

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

GHOSH, P. and CHAKRABARTI, A. (2017). Asymptotic optimality of one-group shrinkage priors in sparse high-dimensional problems. *Bayesian Anal.* **12** 1133–1161. MR3724981

GHOSH, P., TANG, X., GHOSH, M. and CHAKRABARTI, A. (2016). Asymptotic properties of Bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Anal.* **11** 753–796. MR3498045

GORDY, M. B. (1998). Computationally convenient distributional assumptions for common-value auctions. *Comput. Econ.* **12** 61–78.

GRAMACY, R. B. and PANTALEO, E. (2010). Shrinkage regression for multivariate inference with missing data, and an application to portfolio balancing. *Bayesian Anal.* **5** 237–262. MR2719652

GRIFFIN, J. E. and BROWN, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Anal.* **5** 171–188. MR2596440

HAHN, P. R., HE, J. and LOPES, H. (2016). Elliptical slice sampling for Bayesian shrinkage regression with applications to causal inference. Technical report.

HAHN, P. R., HE, J. and LOPES, H. (2018). Bayesian factor model shrinkage for linear IV regression with many instruments. *J. Bus. Econom. Statist.* **36** 278–287. MR3790214

HAHN, P. R., HE, J. and LOPES, H. F. (2019). Efficient sampling for Gaussian linear regression with arbitrary priors. *J. Comput. Graph. Statist.* **28** 142–154. MR3939378

HANS, C. (2011). Elastic net regression modeling with the orthant normal prior. *J. Amer. Statist. Assoc.* **106** 1383–1393. MR2896843

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294

HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141

HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

INGRAHAM, J. B. and MARKS, D. S. (2016). Bayesian sparsity for intractable distributions. Available at arXiv:1602.03807.

ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. (2013). *An Introduction to Statistical Learning: With Applications in R. Springer Texts in Statistics* **103**. Springer, New York. MR3100153

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

JEFFREYS, H. and SWIRLES, B. (1972). *Methods of Mathematical Physics*, 3rd ed. Cambridge Univ. Press, Cambridge. MR1744997

JOHNDROW, J. E. and ORENSTEIN, P. (2017). Scalable MCMC for Bayes shrinkage priors. Available at arXiv:1705.00841.

JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649. MR2089135

JOLLIFFE, I. T., TRENDAFILOV, N. T. and UDDIN, M. (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Statist.* **12** 531–547. MR2002634

KOWAL, D. R., MATTESON, D. S. and RUPPERT, D. (2017). Dynamic shrinkage processes. Available at arXiv:1707.00763.

LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135

LI, Y., CRAIG, B. A. and BHADRA, A. (2017). The graphical horseshoe estimator for inverse covariance matrices. Available at arXiv:1707.06661.

LIU, H. and YU, B. (2013). Asymptotic properties of Lasso + mLS and Lasso + Ridge in sparse high-dimensional linear regression. *Electron. J. Stat.* **7** 3124–3169. MR3151764

LOUIZOS, C., ULLRICH, K. and WELLING, M. (2017). Bayesian compression for deep learning. In *Advances in Neural Information Processing Systems* 3290–3300.

MAGNUSSON, M., JONSSON, L. and VILLANI, M. (2016). DOLDA—A regularized supervised topic model for high-dimensional multi-class regression. Available at arXiv:1602.00260, Jan.

MAKALIC, E. and SCHMIDT, D. F. (2016). High-dimensional Bayesian regularised regression with the BayesReg package. Available at arXiv:1611.06649.

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769

MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322. MR2719857

MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1036. MR0997578

NALENZ, M. and VILLANI, M. (2018). Tree ensembles with rule structured horseshoe regularization. *Ann. Appl. Stat.* **12** 2379–2408. MR3875705

NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450

PELTOLA, T., HAVULINNA, A. S., SALOMAA, V. and VEHTARI, A. (2014). Hierarchical Bayesian survival analysis and projective covariate selection in cardiovascular event risk prediction. In *Proceedings of the Eleventh UAI Conference on Bayesian Modeling Applications Workshop, Vol.* 1218 79–88 CEUR-WS.org.

PIIRONEN, J. and VEHTARI, A. (2015). Projection predictive variable selection using Stan + R. Available at arXiv:1508.02502, Aug.

PIIRONEN, J. and VEHTARI, A. (2017a). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In *Artificial Intelligence and Statistics* 905–913.

PIIRONEN, J. and VEHTARI, A. (2017b). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11** 5018–5051. MR3738204

POLSON, N. G. and SCOTT, J. G. (2010). Large-scale simultaneous testing with hypergeometric inverted-beta priors. Available at arXiv:1010.5223.

POLSON, N. G. and SCOTT, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics* 9 501–538. Oxford Univ. Press, Oxford. MR3204017

POLSON, N. G. and SCOTT, J. G. (2012a). Local shrinkage rules, Lévy processes and regularized regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 287–311. MR2899864

POLSON, N. G. and SCOTT, J. G. (2012b). On the half-Cauchy prior for a global scale parameter. *Bayesian Anal.* **7** 887–902. MR3000018

POLSON, N. G., SCOTT, J. G. and WILLARD, B. T. (2015). Proximal algorithms in statistics and machine learning. *Statist. Sci.* **30** 559–581. MR3432841

ROČKOVÁ, V. and GEORGE, E. I. (2018). The spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. MR3803476

SCOTT, J. G. (2010). Parameter expansion in local-shrinkage models. Available at arXiv:1010.5265.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley, CA. MR0084922

STEPHENS, M. and BALDING, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* **10** 681–690.

SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166

TANG, X., GHOSH, M., XU, X. and GHOSH, P. (2018). Bayesian variable selection and estimation based on global–local shrinkage priors. *Sankhya A* **80** 215–246. MR3850065

TERENIN, A., DONG, S. and DRAPER, D. (2019). GPU-accelerated Gibbs sampling: A case study of the horseshoe probit model. *Stat. Comput.* **29** 301–310. MR3914622

TIAO, G. C. and TAN, W. Y. (1966). Bayesian analysis of random-effect models in the analysis of variance. II. Effect of autocorrelated errors. *Biometrika* **53** 477–495. MR0214218

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TIBSHIRANI, R. J. (2014). In praise of sparsity and convexity. In *Past, Present, and Future of Statistical Science* 497–505.

TIBSHIRANI, R. J., HOEFLING, H. and TIBSHIRANI, R. (2011). Nearly-isotonic regression. *Technometrics* **53** 54–61. MR2791946

TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *Ann. Statist.* **39** 1335–1371. MR2850205

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. MR2136641

TIKHONOV, A. (1963). Solution of incorrectly formulated problems and the regularization method. *Sov. Math., Dokl.* **4** 1035–1038.

TSENG, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** 475–494. MR1835069

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285

VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Stat.* **8** 2585–2618. MR3285877

VAN DER PAS, S. L., SALOMOND, J.-B. and SCHMIDT-HIEBER, J. (2016). Conditions for posterior contraction in the sparse normal means problem. *Electron. J. Stat.* **10** 976–1000. MR3486423

VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2016). How many needles in the haystack? Adaptive inference and uncertainty quantification for the horseshoe. Available at arXiv:1607.01892.

VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Adaptive posterior contraction rates for the horseshoe. *Electron. J. Stat.* **11** 3196–3225. MR3705450

VAN DER PAS, S., SCOTT, J., CHAKRABORTY, A. and BHATTACHARYA, A. (2016). horseshoe: Implementation of the horseshoe prior. R package version 0.1.0.

WANG, H. and PILLAI, N. S. (2013). On a class of shrinkage priors for covariance matrix estimation. *J. Comput. Graph. Statist.* **22** 689–707. MR3173737

WEI, R. (2017). *Bayesian Variable Selection Using Continuous Shrinkage Priors for Nonparametric Models and Non-Gaussian Data*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.), North Carolina State Univ. MR3797433

WITTEN, D. M., TIBSHIRANI, R. and HASTIE, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10** 515–534.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

ZHANG, Y., REICH, B. J. and BONDELL, H. D. (2016). High dimensional linear regression via the R2–D2 shrinkage prior. Available at arXiv:1609.00046.

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940

ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

Statistical Science
2019, Vol. 34, No. 3, 428–453
https://doi.org/10.1214/19-STS702

The Geometry of Continuous Latent Space Models for Network Data

Anna L. Smith, Dena M. Asta and Catherine A. Calder

Abstract. We review the class of continuous latent space (statistical) models for network data, paying particular attention to the role of the geometry of the latent space. In these models, the presence/absence of network dyadic ties are assumed to be conditionally independent given the dyads' unobserved positions in a latent space. In this way, these models provide a probabilistic framework for embedding network nodes in a continuous space equipped with a geometry that facilitates the description of dependence between random dyadic ties. Specifically, these models naturally capture homophilous tendencies and triadic clustering, among other common properties of observed networks. In addition to reviewing the literature on continuous latent space models from a geometric perspective, we highlight the important role the geometry of the latent space plays on properties of networks arising from these models via intuition and simulation. Finally, we discuss results from spectral graph theory that allow us to explore the role of the geometry of the latent space, independent of network size. We conclude with conjectures about how these results might be used to infer the appropriate latent space geometry from observed networks.

Key words and phrases: Geometric curvature, graph Laplacian, latent variable, network model.

REFERENCES

ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. MR3447993

ABU-ATA, M. and DRAGAN, F. F. (2016). Metric tree-like structures in real-world networks: An empirical study. *Networks* **67** 49–68. MR3439658

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E., GOLDENBERG, A., XING, E. P. and ZHENG, A. X. (2008a). *Statistical Network Analysis: Models, Issues, and New Directions: ICML* 2006 *Workshop on Statistical Network Analysis, Pittsburgh, PA, USA, June* 29, 2006, *Revised Selected Papers* **4503**. Springer, Berlin.

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008b). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

AITCHISON, J., BARCELÓ-VIDAL, C., MARTÍN-FERNÁNDEZ, J. and PAWLOWSKY-GLAHN, V. (2000). Logratio analysis and compositional distance. *Mathematical Geology* **32** 271–275.

ALDECOA, R., ORSINI, C. and KRIOUKOV, D. (2015). Hyperbolic graph generator. *Comput. Phys. Commun.* **196** 492–496.

AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

ASTA, D. and SHALIZI, C. R. (2014). Geometric network comparison. Under review. Preprint. Available at arXiv:1411.1350.

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BARTHOLOMEW, D., KNOTT, M. and MOUSTAKI, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2849614

BELKIN, M. and NIYOGI, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *Learning Theory. Lecture Notes in Computer Science* **3559** 486–500. Springer, Berlin. MR2203282

Anna L. Smith is Assistant Professor, Department of Statistics, University of Kentucky, 725 Rose Street, Lexington, Kentucky 40536, USA (e-mail: anna.l.smith@uky.edu). Dena M. Asta is Assistant Professor, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210, USA (e-mail: dasta@stat.osu.edu). Catherine A. Calder is Professor, Department of Statistics and Data Sciences, The University of Texas at Austin, 2317 Speedway, Stop D9800, Austin, Texas 78712, USA (e-mail: calder@austin.utexas.edu).

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*. **106** 21068–21073.

BUTTS, C. T. (2008). Network: A package for managing relational data in R. *J. Stat. Softw.* **24**.

BUTTS, C. T. (2016). sna: Tools for Social Network Analysis. R package version 2.4. https://CRAN.R-project.org/package=sna.

CELINSKA, D. and KOPCZYNSKI, E. (2017). Programming languages in GitHub: A visualization in hyperbolic plane. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (*ICWSM* 2017) 727–728. Association for the Advancement of Artificial Intelligence. Menlo Park, CA.

CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461

CLAUSET, A., MOORE, C. and NEWMAN, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature* **453** 98–101.

CLAUSET, A., NEWMAN, M. E. and MOORE, C. (2004). Finding community structure in very large networks. *Phys. Rev. E* **70** 066111.

COLEMAN, J. S. (1964). *Introduction to Mathematical Sociology*. Free Press, New York.

CSARDI, G. and NEPUSZ, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems* 1695.

DAVIS, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *Am. Sociol. Rev.* 843–851.

DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

DOVETON, H. (1998). Beyond the perfect martini: Teaching the mathematics of petrological logs. In *Proceedings of IAMG*98, *The Fourth Annual Conference of the International Association for Mathematical Geology* 71–75. ACM Press/Addison-Wesley Co., De Frede, Naples.

EASH, R., CHON, K., LEE, Y. and BOYCE, D. (1979). Equilibrium traffic assignment on an aggregated highway network for sketch planning. *Transportation Research* **13** 243–257.

ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

FIENBERG, S. E. and WASSERMAN, S. S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **12** 156–192.

FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606

GILBERT, E. N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839

GOODREAU, S. M., KITTS, J. A. and MORRIS, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* **46** 103–125.

HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *J. Roy. Statist. Soc. Ser. A* **170** 301–354. MR2364300

HEIN, M., AUDIBERT, J.-Y. and VON LUXBURG, U. (2007). Graph Laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.* **8** 1325–1368. MR2332434

HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.* **100** 286–295. MR2156838

HOFF, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* 657–664.

HOFF, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory* **15** 261–272.

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

HOLLAND, P. W. and LEINHARDT, S. (1970). A method for detecting structure in sociometric data. *Amer. J. Sociol.* 492–513.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

HOLLY, J. E. (2001). Pictures of ultrametric spaces, the p-adic numbers, and valued fields. *Amer. Math. Monthly* **108** 721–728. MR1865659

IBRAGIMOV, Z. (2014). A hyperbolic filling for ultrametric spaces. *Comput. Methods Funct. Theory* **14** 315–329. MR3265364

IVRIĬ, V. JA. (1980). The second term of the spectral asymptotics for a Laplace–Beltrami operator on manifolds with boundary. *Funktsional. Anal. i Prilozhen.* **14** 25–34. MR0575202

JAMES, C. (1990). *Foundations of Social Theory*. Belknap, Cambridge, MA.

KOLACZYK, E. D. and CSÁRDI, G. (2014). *Statistical Analysis of Network Data with R. Use R!* Springer, New York. MR3288852

KRIOUKOV, D., PAPADOPOULOS, F., KITSAK, M., VAHDAT, A. and BOGUÑÁ, M. (2010). Hyperbolic geometry of complex networks. *Phys. Rev. E* (3) **82** 036106, 18. MR2787998

KRIVITSKY, P. N., HANDCOCK, M. S., RAFTERY, A. E. and HOFF, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc. Netw.* **31** 204–213.

KUNEGIS, J. (2013). *KONECT—The Koblenz Network Collection*.

LAMPING, J., RAO, R. and PIROLLI, P. (1995). A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 401–408. ACM Press/Addison-Wesley, New York.

LAZARSFELD, P. F., HENRY, N. W. and ANDERSON, T. W. (1968). *Latent Structure Analysis* **109**. Houghton Mifflin, Boston.

LICHNEROWICZ, A. (1958). Geometrie des groupes de transformations, Dunod, Paris, 1958. *Zentralblatt MATH* **96**.

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035

MADORE, D. Hyperbolic maze. Available at http://www.madore.org/~david/math/hyperbolic-maze.html. Hyperbolic maze.

MCCORMICK, T. H. and ZHENG, T. (2015). Latent surface models for networks using aggregated relational data. *J. Amer. Statist. Assoc.* **110** 1684–1695. MR3449064

MINHAS, S., HOFF, P. D. and WARD, M. D. (2016). Inferential approaches for network analyses: AMEN for latent factor models. Preprint. Available at arXiv:1611.00460.

MUNZNER, T. (1997). H3: Laying out large directed graphs in 3D hyperbolic space. In *IEEE Symposium on Information Visualization*, 1997. *Proceedings* 2–10. IEEE Press, New York.

NEWMAN, M. E. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69** 026113.

NICKEL, C. L. M. (2009). Random dot product graphs a model for social networks. Ph.D. dissertation. Johns Hopkins Univ., Baltimore, MD.

OPSAHL, T. (2009). Structure and evolution of weighted networks. Ph.D thesis, Queen Mary, Univ. London.

PADGETT, J. F. (1994). *Marriage and Elite Structure in Renaissance Florence*, 1282–1500. Social Science History Association.

PAO, H., COPPERSMITH, G. A. and PRIEBE, C. E. (2011). Statistical inference on random graphs: Comparative power analyses via Monte Carlo. *J. Comput. Graph. Statist.* **20** 395–416. MR2847801

PATTISON, P. and ROBINS, G. (2002). Neighborhood-based models for social networks. *Sociol. Method.* **32** 301–337.

R CORE TEAM (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAPOPORT, A. (1953). Spread of information through a population with socio-structural bias. I. Assumption of transitivity. *Bull. Math. Biophys.* **15** 523–533. MR0058955

ROBINS, G., PATTISON, P., KALISH, Y. and LUSHER, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29** 173–191.

ROGUE, Z. and ROGUE, T. (2011). HyperRogue. Available at http://www.roguetemple.com/z/hyper/.

SALDAÑA, D. F., YU, Y. and FENG, Y. (2017). How many communities are there? *J. Comput. Graph. Statist.* **26** 171–181. MR3610418

SARKAR, P. and MOORE, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems* 1145–1152.

SCHWEINBERGER, M. and SNIJDERS, T. A. (2003). Settings in social networks: A measurement model. *Sociol. Method.* **33** 307–341.

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061

SIMMEL, G. (1950). The sociology of Georg Simmel. In *Individual and Society*. 3–84. Free Press, New York.

SMITH, A. (2017). Statistical methodology for multiple networks. PhD thesis, The Ohio State Univ., Columbus, OH.

SNIJDERS, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure* **3** 1–40.

SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. MR1449742

SNIJDERS, T. A., PATTISON, P. E., ROBINS, G. L. and HANDCOCK, M. S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.

SOLÉ, R. V. and VALVERDE, S. (2004). Information theory of complex networks: On evolution and architectural constraints. In *Complex Networks. Lecture Notes in Physics* **650** 189–207. Springer, Berlin. MR2108978

SPEARMAN, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology* **15** 201–292.

TING, D., HUANG, L. and JORDAN, M. (2011). An analysis of the convergence of graph Laplacians. Preprint. Available at arXiv:1101.5435.

VAN DER LINDEN, W. J. and HAMBLETON, R. K., eds. (2013). *Handbook of Modern Item Response Theory*. Springer, New York.

VAN DUIJN, M. A. J., SNIJDERS, T. A. B. and ZIJLSTRA, B. J. H. (2004). p_2: A random effects model with covariates for directed graphs. *Stat. Neerl.* **58** 234–254. MR2064846

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

WATTS, D. J. (1999a). Networks, dynamics, and the small-world phenomenon. *Amer. J. Sociol.* **105** 493–527.

WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393** 440–442.

WEYL, H. (1911). Sur la distribution asymptotique des valeurs propres. *Nouvelles de la Société des Sciences sur Göttingen, Mathematical-Physical Class* 110–117.

WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at arXiv:1309.5936.

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912

ZACHARY, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* **33** 452–473.

Statistical Science
2019, Vol. 34, No. 3, 454–471
https://doi.org/10.1214/19-STS711

User-Friendly Covariance Estimation for Heavy-Tailed Distributions

Yuan Ke, Stanislav Minsker, Zhao Ren, Qiang Sun and Wen-Xin Zhou

Abstract. We provide a survey of recent results on covariance estimation for heavy-tailed distributions. By unifying ideas scattered in the literature, we propose user-friendly methods that facilitate practical implementation. Specifically, we introduce elementwise and spectrumwise truncation operators, as well as their M-estimator counterparts, to robustify the sample covariance matrix. Different from the classical notion of robustness that is characterized by the breakdown property, we focus on the tail robustness which is evidenced by the connection between nonasymptotic deviation and confidence level. The key insight is that estimators should adapt to the sample size, dimensionality and noise level to achieve optimal tradeoff between bias and robustness. Furthermore, to facilitate practical implementation, we propose data-driven procedures that automatically calibrate the tuning parameters. We demonstrate their applications to a series of structured models in high dimensions, including the bandable and low-rank covariance matrices and sparse precision matrices. Numerical studies lend strong support to the proposed methods.

Key words and phrases: Covariance estimation, heavy-tailed data, M-estimation, nonasymptotics, tail robustness, truncation.

REFERENCES

AVELLA-MEDINA, M., BATTEY, H. S., FAN, J. and LI, Q. (2018). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika* **105** 271–284. MR3804402

BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

BICKEL, P. J. and LEVINA, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. MR3405602

BUTLER, R. W., DAVIES, P. L. and JHUN, M. (1993). Asymptotics for the minimum covariance determinant estimator. *Ann. Statist.* **21** 1385–1400. MR1241271

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

CAI, T. T., REN, Z. and ZHOU, H. H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electron. J. Stat.* **10** 1–59. MR3466172

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407

CATONI, O. (2016). PAC-Bayesian bounds for the Gram matrix and least squares regression with a random design. Preprint. Available at arXiv:1603.05229.

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006

Yuan Ke is Assistant Professor, Department of Statistics, University of Georgia, Athens, Georgia 30602, USA (e-mail: yuan.ke@uga.edu). Stanislav Minsker is Assistant Professor, Department of Mathematics, University of Southern California, Los Angeles, California 90089, USA (e-mail: minsker@usc.edu). Zhao Ren is Assistant Professor, Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA (e-mail: zren@pitt.edu). Qiang Sun is Assistant Professor, Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada (e-mail: qsun@utstat.toronto.edu). Wen-Xin Zhou is Assistant Professor, Department of Mathematics, University of California, San Diego, La Jolla, California 92093, USA (e-mail: wez243@ucsd.edu).

CHEN, X. and ZHOU, W.-X. (2019). Robust inference via multiplier bootstrap. *Ann. Statist.* To appear. Available at arXiv:1903.07208.

CHERAPANAMJERI, Y., FLAMMARION, N. and BARTLETT, P. L. (2019). Fast mean estimation with sub-Gaussian rates. Preprint. Available at arXiv:1902.01998.

CONT, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* **1** 223–236.

DAVIES, L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843. MR1193314

DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558

EKLUND, A., NICHOLS, T. E. and KNUTSSON, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. USA* **113** 7900–7905.

FAN, J., LIAO, Y. and LIU, H. (2016). An overview of the estimation of large covariance and precision matrices. *Econom. J.* **19** C1–C32. MR3501529

FAN, J., LIU, H., SUN, Q. and ZHANG, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Ann. Statist.* **46** 814–841. MR3782385

FAN, J., SUN, Q., ZHOU, W.-X. and ZHU, Z. (2019). Principal component analysis for big data. *Wiley StatsRef: Statistics Reference Online*. To appear. DOI:10.1002/9781118445112.stat08122.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42** 1887–1896. MR0301858

HOPKINS, S. B. (2018). Mean estimation with sub-Gaussian rates in polynomial time. Preprint. Available at arXiv:1809.07425.

HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18, 40. MR3491112

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35** 73–101. MR0161415

HUBERT, M., ROUSSEEUW, P. J. and VAN AELST, S. (2008). High-breakdown robust multivariate methods. *Statist. Sci.* **23** 92–119. MR2431867

KE, Y., MINSKER, S., REN, Z., SUN, Q. and ZHOU, W.-X (2019). Supplement to "User-Friendly Covariance Estimation for Heavy-Tailed Distributions." DOI:10.1214/19-STS711SUPP.

LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. MR1604408

LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400

LIU, Y. and REN, Z. (2018). Minimax estimation of large precision matrices with bandable Cholesky factor. Preprint. Available at arXiv:1712.09483.

LIU, L., HAWKINS, D. M., GHOSH, S. and YOUNG, S. S. (2003). Robust singular value decomposition analysis of microarray data. *Proc. Natl. Acad. Sci. USA* **100** 13167–13172. MR2016727

LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437

LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67. MR0388656

MENDELSON, S. and ZHIVOTOVSKIY, N. (2018). Robust covariance estimation under L_4–L_2 norm equivalence. Preprint. Available at arXiv:1809.10462.

MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468

MINSKER, S. (2018). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Ann. Statist.* **46** 2871–2903. MR3851758

MINSKER, S. and STRAWN, N. (2017). Distributed statistical estimation and rates of convergence in normal approximation. Preprint. Available at arXiv:1704.02658.

MINSKER, S. and WEI, X. (2018). Robust modifications of U-statistics and applications to covariance estimation problems. Preprint. Available at arXiv:1801.05565.

MITRA, R. and ZHANG, C.-H. (2014). Multivariate analysis of nonparametric estimates of large correlation matrices. Preprint. Available at arXiv:1403.6195.

MIZERA, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736. MR1969447

NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. A Wiley-Interscience Publication*. Wiley, New York. MR0702836

PORTNOY, S. and HE, X. (2000). A robust journey in the new millennium. *J. Amer. Statist. Assoc.* **95** 1331–1335. MR1825288

PURDOM, E. and HOLMES, S. P. (2005). Error distribution for gene expression data. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 16. MR2170432

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230. MR0760684

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. MR0770281

ROUSSEEUW, P. and YOHAI, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis* (*Heidelberg*, 1983). *Lect. Notes Stat.* **26** 256–272. Springer, New York. MR0786313

SALIBIAN-BARRERA, M. and ZAMAR, R. H. (2002). Bootstrapping robust estimates of regression. *Ann. Statist.* **30** 556–582. MR1902899

SUN, Q., ZHOU, W.-X. and FAN, J. (2019). Adaptive Huber regression. *J. Amer. Statist. Assoc.* DOI:10.1080/01621459.2018.1543124.

SUN, Q., TAN, K. M., LIU, H. and ZHANG, T. (2018). Graphical nonconvex optimization via an adaptive convex relaxation. In *Proceedings of the* 35*th International Conference on Machine Learning* **80** 4810–4817.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, vol. 2.

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251. MR0885734

VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

WANG, L., ZHENG, C., ZHOU, W. and ZHOU, W.-X. (2018). A new principle for tuning-free Huber regression. Technical Report.

YOHAI, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642–656. MR0888431

ZHANG, T., CHENG, X. and SINGER, A. (2016). Marčenko–Pastur law for Tyler's M-estimator. *J. Multivariate Anal.* **149** 114–123. MR3507318

ZHANG, T. and ZOU, H. (2014). Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika* **101** 103–120. MR3180660

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005

Statistical Science
2019, Vol. 34, No. 3, 472–485
https://doi.org/10.1214/19-STS712

Conditionally Conjugate Mean-Field Variational Bayes for Logistic Models

Daniele Durante and Tommaso Rigon

Abstract. Variational Bayes (VB) is a common strategy for approximate Bayesian inference, but simple methods are only available for specific classes of models including, in particular, representations having conditionally conjugate constructions within an exponential family. Models with logit components are an apparently notable exception to this class, due to the absence of conjugacy among the logistic likelihood and the Gaussian priors for the coefficients in the linear predictor. To facilitate approximate inference within this widely used class of models, Jaakkola and Jordan (*Stat. Comput.* **10** (2000) 25–37) proposed a simple variational approach which relies on a family of tangent quadratic lower bounds of the logistic log-likelihood, thus restoring conjugacy between these approximate bounds and the Gaussian priors. This strategy is still implemented successfully, but few attempts have been made to formally understand the reasons underlying its excellent performance. Following a review on VB for logistic models, we cover this gap by providing a formal connection between the above bound and a recent Pólya-gamma data augmentation for logistic regression. Such a result places the computational methods associated with the aforementioned bounds within the framework of variational inference for conditionally conjugate exponential family models, thereby allowing recent advances for this class to be inherited also by the methods relying on Jaakkola and Jordan (*Stat. Comput.* **10** (2000) 25–37).

Key words and phrases: EM, logistic regression, Pólya-gamma data augmentation, quadratic approximation, variational Bayes.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

BEAL, M. J. and GHAHRAMANI, Z. (2003). The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures. In *Bayesian Statistics*, 7 (*Tenerife*, 2002) 453–463. Oxford Univ. Press, New York. MR2003189

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York. MR2247587

BISHOP, C. M. and SVENSÉN, M. (2003). Bayesian hierarchical mixtures of experts. *Proc. Conf. Uncertain. Artif. Intell.* 57–64.

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776

BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.

BÖHNING, D. and LINDSAY, B. G. (1988). Monotonicity of quadratic-approximation algorithms. *Ann. Inst. Statist. Math.* **40** 641–663. MR0996690

BRAUN, M. and MCAULIFFE, J. (2010). Variational inference for large-scale models of discrete choice. *J. Amer. Statist. Assoc.* **105** 324–335. MR2757203

BROWNE, R. P. and MCNICHOLAS, P. D. (2015). Multivariate sharp quadratic bounds via Σ-strong convexity and the Fenchel connection. *Electron. J. Stat.* **9** 1913–1938. MR3391124

CARBONETTO, P. and STEPHENS, M. (2012). Scalable variational inference for Bayesian variable selection in regression, and its

Daniele Durante is Assistant Professor of Statistics, Department of Decision Sciences and affiliate to the Bocconi Institute for Data Science and Analytics, Bocconi University, Via Roentgen 1, Milan, Italy (e-mail: daniele.durante@unibocconi.it). Tommaso Rigon is Ph.D. Student in Statistics, Department of Decision Sciences, Bocconi University, Via Roentgen 1, Milan, Italy (e-mail: tommaso.rigon@phd.unibocconi.it).

accuracy in genetic association studies. *Bayesian Anal*. **7** 73–107. MR2896713

CHOI, H. M. and HOBERT, J. P. (2013). The Polya-gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electron*. *J*. *Stat*. **7** 2054–2064. MR3091616

DE LEEUW, J. and LANGE, K. (2009). Sharp quadratic majorization in one dimension. *Comput*. *Statist*. *Data Anal*. **53** 2471–2484. MR2665900

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **39** 1–38. MR0501537

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J*. *Amer*. *Statist*. *Assoc*. **85** 398–409. MR1141740

GIORDANO, R. J., BRODERICK, T. and JORDAN, M. I. (2015). Linear response methods for accurate covariance estimates from mean field variational Bayes. *Adv*. *Neural Inf*. *Process*. *Syst*. 1441–1449.

HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J*. *Mach*. *Learn*. *Res*. **14** 1303–1347. MR3081926

HUNTER, D. R. and LANGE, K. (2004). A tutorial on MM algorithms. *Amer*. *Statist*. **58** 30–37. MR2055509

JAAKKOLA, T. S. and JORDAN, M. I. (2000). Bayesian parameter estimation via variational methods. *Stat*. *Comput*. **10** 25–37.

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach*. *Learn*. **37** 183–233.

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann*. *Math*. *Stat*. **22** 79–86. MR0039968

LEE, S., HUANG, J. Z. and HU, J. (2010). Sparse logistic principal components analysis for binary data. *Ann*. *Appl*. *Stat*. **4** 1579–1601. MR2758342

MCLACHLAN, G. J. and KRISHNAN, T. (1997). *The EM Algorithm and Extensions*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1417721

ORMEROD, J. T. and WAND, M. P. (2010). Explaining variational approximations. *Amer*. *Statist*. **64** 140–153. MR2757005

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *J*. *Amer*. *Statist*. *Assoc*. **108** 1339–1349. MR3174712

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

REN, L., DU, L., CARIN, L. and DUNSON, D. B. (2011). Logistic stick-breaking process. *J*. *Mach*. *Learn*. *Res*. **12** 203–239. MR2773552

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann*. *Math*. *Stat*. **22** 400–407. MR0042668

SCOTT, J. G. and SUN, L. (2013). Expectation-maximization for logistic regression. Available at arXiv:1306.0040.

SPALL, J. C. (2003). *Introduction to Stochastic Search and Optimization*: *Estimation*, *Simulation*, *and Control*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, Hoboken, NJ. MR1968388

TANG, Y., BROWNE, R. P. and MCNICHOLAS, P. D. (2015). Model based clustering of high-dimensional binary data. *Comput*. *Statist*. *Data Anal*. **87** 84–101. MR3319809

WAND, M. P. (2017). Fast approximate inference for arbitrarily large semiparametric regression models via message passing. *J*. *Amer*. *Statist*. *Assoc*. **112** 137–156. MR3646558

WAND, M. P., ORMEROD, J. T., PADOAN, S. A. and FRÜHRWIRTH, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Anal*. **6** 847–900. MR2869967

WANG, C. and BLEI, D. M. (2013). Variational inference in nonconjugate models. *J*. *Mach*. *Learn*. *Res*. **14** 1005–1031. MR3063617

WANG, B. and TITTERINGTON, D. M. (2004). Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values. *Proc*. *Conf*. *Uncertain*. *Artif*. *Intell*. 577–584.

ZHU, L. (2012). New inequalities for hyperbolic functions and their applications. *J*. *Inequal*. *Appl*. **303** 1–29. MR3017334

Statistical Science
2019, Vol. 34, No. 3, 486–503
https://doi.org/10.1214/19-STS713

Assessing the Causal Effect of Binary Interventions from Observational Panel Data with Few Treated Units

Pantelis Samartsidis, Shaun R. Seaman, Anne M. Presanis, Matthew Hickman and Daniela De Angelis

Abstract. Researchers are often challenged with assessing the impact of an intervention on an outcome of interest in situations where the intervention is nonrandomised, the intervention is only applied to one or few units, the intervention is binary, and outcome measurements are available at multiple time points. In this paper, we review existing methods for causal inference in these situations. We detail the assumptions underlying each method, emphasize connections between the different approaches and provide guidelines regarding their practical implementation. Several open problems are identified thus highlighting the need for future research.

Key words and phrases: Causal impact, causal inference, difference-in-differences, intervention evaluation, latent factor models, panel data, synthetic controls.

REFERENCES

ABADIE, A. (2005). Semiparametric difference-in-differences estimators. *Rev. Econ. Stud.* **72** 1–19. MR2116973

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2011). Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software* **42**.

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2015). Comparative politics and the synthetic control method. *Amer. J. Polit. Sci.* **59** 495–510.

ABADIE, A. and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque country. *Am. Econ. Rev.* **93** 113–132.

ACEMOGLU, D., JOHNSON, S., KERMANI, A., KWAK, J. and MITTON, T. (2016). The value of connections in turbulent times: Evidence from the United States. *J. Financ. Econ.* **121** 368–391.

AHN, S. C., LEE, Y. H. and SCHMIDT, P. (2013). Panel data models with multiple time-varying individual effects. *J. Econometrics* **174** 1–14. MR3036957

AMJAD, M., SHAH, D. and SHEN, D. (2018). Robust synthetic control. *J. Mach. Learn. Res.* **19** Paper No. 22, 51. MR3862429

ANDREWS, D. W. K. (2003). End-of-sample instability tests. *Econometrica* **71** 1661–1694. MR2015416

ANGRIST, J. D. and PISCHKE, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton Univ. Press, Princeton, NJ.

ANTONAKIS, J., BENDAHAN, S., JACQUART, P. and LALIVE, R. (2010). On making causal claims: A review and recommendations. *Leadersh. Q.* **21** 1086–1120.

ASHENFELTER, O. (1978). Estimating the effect of training programs on earnings. *Rev. Econ. Stat.* **60** 47–57.

Pantelis Samartsidis is Investigator Statistician, MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, CB2 0SR, United Kingdom (e-mail: pantelis.samartsidis@mrc-bsu.cam.ac.uk). Shaun R. Seaman is Senior Research Associate, MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, CB2 0SR, United Kingdom (e-mail: shaun.seaman@mrc-bsu.cam.ac.uk). Anne M. Presanis is Senior Investigator Statistician, MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, CB2 0SR, United Kingdom (e-mail: anne.presanis@mrc-bsu.cam.ac.uk). Matthew Hickman is Professor, MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, CB2 0SR, United Kingdom (e-mail: matthew.hickman@bristol.ac.uk). Daniela De Angelis is Programme Leader, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2PS, United Kingdom (e-mail: daniela.deangelis@mrc-bsu.cam.ac.uk).

ASHENFELTER, O. and CARD, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *Rev. Econ. Stat.* **67** 648–660.

ATANASOV, V. and BLACK, B. (2016). Shock-based causal inference in corporate finance and accounting research. *Critical Finance Review* **5** 207–304.

ATHEY, S. and IMBENS, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica* **74** 431–497. MR2207397

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2017). Matrix completion methods for causal panel data models. Preprint. Available at arXiv:1710.10251.

AUSTIN, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. Behav. Res.* **46** 399–424.

AYTUĞ, H., KÜTÜK, M. M., ODUNCU, A. and TOGAN, S. (2017). Twenty years of the EU–Turkey customs union: A synthetic control method analysis. *J. Common Mark. Stud.* **55** 419–431.

BAI, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77** 1229–1279. MR2547073

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2018). The augmented synthetic control method. Preprint. Available at arXiv:1811.04170.

BERTRAND, M., DUFLO, E. and MULLAINATHAN, S. (2004). How much should we trust differences-in-differences estimates? *Q. J. Econ.* **119** 249–275.

BILLMEIER, A. and NANNICINI, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Rev. Econ. Stat.* **95** 983–1001.

BLUNDELL, R., DIAS, M. C., MEGHIR, C. and VAN REENEN, J. (2004). Evaluating the employment impact of a mandatory job search program. *J. Eur. Econ. Assoc.* **2** 569–606.

BRANAS, C. C., CHENEY, R. A., MACDONALD, J. M., TAM, V. W., JACKSON, T. D. and TEN HAVE, T. R. (2011). A difference-in-differences analysis of health, safety, and greening vacant urban space. *Am. J. Epidemiol.* **174** 1296–1306.

BRODERSEN, K. H., GALLUSSER, F., KOEHLER, J., REMY, N. and SCOTT, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat.* **9** 247–274. MR3341115

BRUHN, C. A., HETTERICH, S., SCHUCK-PAIM, C., KÜRÜM, E., TAYLOR, R. J., LUSTIG, R., SHAPIRO, E. D., WARREN, J. L., SIMONSEN, L. et al. (2017). Estimating the population-level impact of vaccines using synthetic controls. *Proc. Natl. Acad. Sci. USA* **114** 1524–1529.

CARD, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Ind. Labor Relat. Rev.* **43** 245–257.

CARD, D. and KRUEGER, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *Am. Econ. Rev.* **84** 772–793.

CARVALHO, C., MASINI, R. and MEDEIROS, M. C. (2018). ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *J. Econometrics* **207** 352–380. MR3868187

CAVALLO, E., GALIANI, S., NOY, I. and PANTANO, J. (2013). Catastrophic natural disasters and economic growth. *Rev. Econ. Stat.* **95** 1549–1561.

CHAN, M. K. and KWOK, S. (2016). Policy evaluation with interactive fixed effects. Preprint. Available at https://ideas.repec.org/p/syd/wpaper/2016-11.html.

CHERNOZHUKOV, V., WÜTHRICH, K. and ZHU, Y. (2017). An exact and robust conformal inference method for counterfactual and synthetic controls. Preprint. Available at arXiv:1712.09089.

DE VOCHT, F. (2016). Inferring the 1985–2014 impact of mobile phone use on selected brain cancer subtypes using Bayesian structural time series and synthetic controls. *Environ. Int.* **97** 100–107.

DE VOCHT, F., TILLING, K., PLIAKAS, T., ANGUS, C., EGAN, M., BRENNAN, A., CAMPBELL, R. and HICKMAN, M. (2017). The intervention effect of local alcohol licensing policies on hospital admission and crime: A natural experiment using a novel Bayesian synthetic time-series method. *J. Epidemiol. Community Health* **71** 912–918.

DONALD, S. G. and LANG, K. (2007). Inference with difference-in-differences and other panel data. *Rev. Econ. Stat.* **89** 221–233.

DOUDCHENKO, N. and IMBENS, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Preprint. Available at arXiv:1610.07748.

DUBE, A. and ZIPPERER, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies.

FERMAN, B. and PINTO, C. (2016). Revisiting the synthetic control estimator.

FERMAN, B., PINTO, C. and POSSEBOM, V. (2016). Cherry picking with synthetic controls.

FILZMOSER, P., MARONNA, R. and WERNER, M. (2008). Outlier identification in high dimensions. *Comput. Statist. Data Anal.* **52** 1694–1711. MR2422764

FIRPO, S. and POSSEBOM, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *J. Causal Inference* **6**.

FUJIKI, H. and HSIAO, C. (2015). Disentangling the effects of multiple treatments—measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake. *J. Econometrics* **186** 66–73. MR3321526

GALIANI, S., GERTLER, P. and SCHARGRODSKY, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *J. Polit. Econ.* **113** 83–120.

GARDEAZABAL, J. and VEGA-BAYO, A. (2017). An empirical comparison between the synthetic control method and Hsiao *et al.*'s panel data approach to program evaluation. *J. Appl. Econometrics* **32** 983–1002. MR3689015

GLASS, T. A., GOODMAN, S. N., HERNÁN, M. A. and SAMET, J. M. (2013). Causal inference in public health. *Annu. Rev. Public Health* **34** 61–75.

GOBILLON, L. and MAGNAC, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Rev. Econ. Stat.* **98** 535–551.

GONZÁLEZ, R. and HOSODA, E. B. (2016). Environmental impact of aircraft emissions and aviation fuel tax in Japan. *J. Air Transp. Manag.* **57** 234–240.

HAHN, J. and SHI, R. (2017). Synthetic control and inference. *Econometrics* **5** 52.

HANSEN, C. and LIAO, Y. (2019). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory* **35** 465–509. MR3947253

HAZLETT, C. and XU, Y. (2018). Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data.

HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. MR0867618

HSIAO, C., CHING, H. S. and WAN, S. K. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. *J. Appl. Econometrics* **27** 705–740. MR2949893

IMAI, K., KIM, I. S. and WANG, E. (2018). Matching methods for causal inference with time-series cross-section data.

IMBENS, G. W. and WOOLDRIDGE, J. M. (2009). Recent developments in the econometrics of program evaluation. *J. Econ. Lit.* **47** 5–86.

JONES, A. M. and RICE, N. (2011). Econometric evaluation of health policies. In *The Oxford Handbook of Health Economics* (S. Glied and P. C. Smith, eds.). *Oxford Handbooks* 890–923. Oxford Univ. Press, Oxford.

KEELE, L. (2015). The statistics of causal inference: A view from political methodology. *Polit. Anal.* **23** 313–335.

KEELE, L. and MINOZZI, W. (2013). How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data. *Polit. Anal.* **21** 193–216.

KIM, D. and OKA, T. (2014). Divorce law reforms and divorce rates in the USA: An interactive fixed-effects approach. *J. Appl. Econometrics* **29** 231–245. MR3233244

KING, M., ESSICK, C., BEARMAN, P. and ROSS, J. S. (2013). Medical school gift restriction policies and physician prescribing of newly marketed psychotropic medications: Difference-in-differences analysis. *BMJ* **346** f264.

KINN, D. (2018). Synthetic control methods and big data. Preprint. Available at arXiv:1803.00096.

KREIF, N., GRIEVE, R., HANGARTNER, D., TURNER, A. J., NIKOLOVA, S. and SUTTON, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ.* **25** 1514–1528.

LI, K. (2018). Inference for factor model based average treatment effects.

LI, K. T. and BELL, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *J. Econometrics* **197** 65–75. MR3598646

LOPES, H. F., SALAZAR, E. and GAMERMAN, D. (2008). Spatial dynamic factor analysis. *Bayesian Anal.* **3** 759–792. MR2469799

LOPEZ BERNAL, J., CUMMINS, S. and GASPARRINI, A. (2016). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *Int. J. Epidemiol.* **46** 348–355.

MORGAN, S. L. and WINSHIP, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research. Analytical Methods for Social Research*. Cambridge Univ. Press, Cambridge.

O'NEILL, S., KREIF, N., GRIEVE, R., SUTTON, M. and SEKHON, J. S. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Serv. Outcomes Res. Methodol.* **16** 1–21.

R CORE TEAM (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

ROBBINS, M. W., SAUNDERS, J. and KILMER, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *J. Amer. Statist. Assoc.* **112** 109–126. MR3646556

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

ROTHMAN, K. J. and GREENLAND, S. (2005). Causation and causal inference in epidemiology. *Am. J. Publ. Health* **95** S144–S150.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

RUBIN, D. B. (1990). Formal mode of statistical inference for causal effects. *J. Statist. Plann. Inference* **25** 279–292.

RUBIN, D. B. and WATERMAN, R. P. (2006). Estimating the causal effects of marketing interventions using propensity score methodology. *Statist. Sci.* **21** 206–222. MR2324079

RYAN, A. M., KRINSKY, S., KONTOPANTELIS, E. and DORAN, T. (2016). Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: A population study. *Lancet* **388** 268–274.

RYAN, A. M., KONTOPANTELIS, E., LINDEN, A. and BURGESS, J. F. JR. (2018). Now trending: Coping with non-parallel trends in difference-in-differences analysis. *Stat. Methods Med. Res.* DOI:10.1177/0962280218814570.

SANSO-NAVARRO, M., SANZ-GRACIA, F. and VERA-CABELLO, M. (2018). The demographic impact of terrorism: Evidence from municipalities in the Basque Country and Navarre. *Regional Studies* **0** 1–11.

SAUNDERS, J., LUNDBERG, R., BRAGA, A. A., RIDGEWAY, G. and MILES, J. (2015). A synthetic control approach to evaluating place-based crime interventions. *J. Quant. Criminol.* **31** 413–434.

SCHMITT, E., TULL, C. and ATWATER, P. (2018). Extending Bayesian structural time-series estimates of causal impact to many-household conservation initiatives. *Ann. Appl. Stat.* **12** 2517–2539. MR3875710

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812

VARIAN, H. R. (2016). Causal inference in economics and marketing. *Proc. Natl. Acad. Sci. USA* **113** 7310–7315.

VEGA-BAYO, A. (2015). An R package for the panel approach method for program evaluation. *R J.* **7**.

VIBOUD, C., BOËLLE, P.-Y., CARRAT, F., VALLERON, A.-J. and FLAHAULT, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *Am. J. Epidemiol.* **158** 996–1006.

VIZZOTTI, C., JUAREZ, M. V., BERGEL, E., ROMANIN, V., CALIFANO, G., SAGRADINI, S., RANCAÑO, C., AQUINO, A., LIBSTER, R. et al. (2016). Impact of a maternal immunization program against pertussis in a developing country. *Vaccine* **34** 6223–6228.

WING, C., SIMON, K. and BELLO-GOMEZ, R. A. (2018). Designing difference in difference studies: Best practices for public health policy research. *Annu. Rev. Public Health* **39** 453–469.

WOOLDRIDGE, J. M. (2013). *Introductory Econometrics*: *A Modern Approach*. Cengage Learning, Mason.

XU, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Polit*. *Anal*. **25** 57–76.

Statistical Science
2019, Vol. 34, No. 3, 504–521
https://doi.org/10.1214/19-STS703

A Conversation with Peter Diggle

Peter M. Atkinson and Jorge Mateu

Abstract. Peter John Diggle was born on February 24, 1950, in Lancashire, England. Peter went to school in Scotland, and it was at the end of his school years that he found that he was good at maths and actually enjoyed it. Peter went to Edinburgh to do a maths degree, but transferred halfway through to Liverpool where he completed his degree. Peter studied for a year at Oxford and was then appointed in 1974 as a lecturer in statistics at the University of Newcastle-upon-Tyne where he gained his PhD, and was promoted to Reader in 1983. A sabbatical at the Swedish Royal College of Forestry gave him his first exposure to real scientific data and problems, prompting a move to CSIRO, Australia. After five years with CSIRO where he was Senior, then Principal, then Chief Research Scientist and Chief of the Division of Mathematics and Statistics, he returned to the UK in 1988, to a Chair at Lancaster University. Since 2011 Peter has held appointments at Lancaster and Liverpool, together with honorary appointments at Johns Hopkins, Columbia and Yale. At Lancaster, Peter was the founder and Director of the Medical Statistics Unit (1995–2001), University Dean for Research (1998–2001), EPSRC Senior Fellow (2004–2008), Associate Dean for Research at the School of Health and Medicine (2007–2011), Distinguished University Professor, and leader of the CHICAS Research Group (2007–2017). A Fellow of the Royal Statistical Society since 1974, he was a Member of Council (1983–1985), Joint Editor of *JRSSB* (1984–1987), Honorary Secretary (1990–1996), awarded the Guy Medal in Silver (1997) and the Barnett Award (2018), Associate Editor of *Applied Statistics* (1998–2000), Chair of the Research Section Committee (1998–2000), and President (2014–2016). Away from work, Peter enjoys music, playing folk-blues guitar and tenor recorder, and listening to jazz. His running days are behind him, but he can just about hold his own in mixed-doubles badminton with his family. His boyhoood hero was Stirling Moss, and he retains an enthusiasm for classic cars, not least his 1988 Porsche 924S. His favorite authors are George Orwell, Primo Levi and Nigel Slater. This interview was done prior to the fourth Spatial Statistics conference held in Lancaster, July 2017 where a session was dedicated to Peter celebrating his contributions to statistics.

Key words and phrases: CSIRO, geostatistics, Lancaster, longitudinal data analysis, point pattern analysis, spatial statistics, CHICAS.

Peter M. Atkinson is Professor of Spatial Data Science, Dean of the Faculty of Science and Technology and Dean of the Faculty of Health and Medicine at Lancaster University, Bailrigg, Lancaster LA1 4YW, UK. Peter is also currently Visiting Research Professor at Queen's University at Belfast, UK, Visiting Professor at University of Southampton, Southampton, UK and Visiting Professor at the Chinese Academy of Sciences, Beijing, China (e-mail: pma@lancaster.ac.uk). Jorge Mateu is Professor of Statistics, Department of Mathematics, University Jaume I of Castellon, E-12071 Castellon, Spain; Director of the Research Unit Eurocop: Statistical Modelling of Crime Data, based at the Department of Mathematics, University Jaume I of Castellon (e-mail: mateu@uji.es).

REFERENCES

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **47** 299–350. MR1626544

GIORGI, E. and DIGGLE, P. J. (2017). PrevMap: An R package for prevalence mapping. *J. Stat. Softw.* **78**.

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

MATÉRN, B. (1960). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. *Meddelanden Fran Statens Skogsforskningsinstitut* **49** Nr. 5. Stockholm. MR0169346

RIBEIRO, P. J. JR and DIGGLE, P. J. (2001). geoR: A package for geostatistical analysis. *R News* **1**.

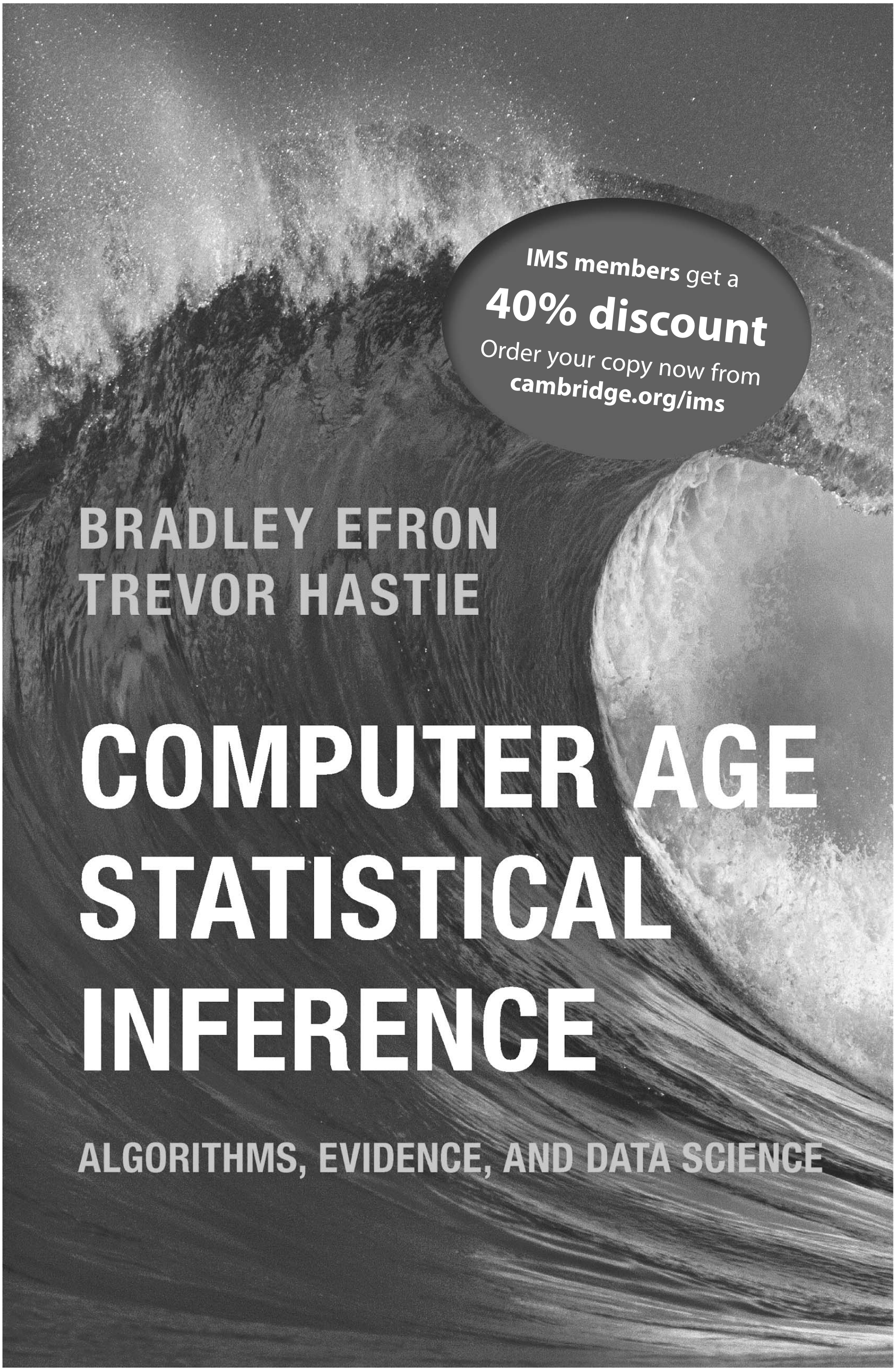
ROWLINGSON, B. S. and DIGGLE, P. J. (1993). Splancs: Spatial point pattern analysis code in S-plus. *Comput. Geosci.* **19** 627–655.

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE