STATISTICAL SCIENCE

Volume 35, Number 4 November 2020

***Statistical Science* [ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 35, Number 4, November 2020. 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 35, Number 4 (555–671) November 2020**

Volume 35 Number 4 November 2020

Statistical Science
2020, Vol. 35, No. 4, 555–578
https://doi.org/10.1214/19-STS701

Sparse Regression: Scalable Algorithms and Empirical Performance

Dimitris Bertsimas, Jean Pauphilet and Bart Van Parys

Abstract. In this paper, we review state-of-the-art methods for feature selection in statistics with an application-oriented eye. Indeed, sparsity is a valuable property and the profusion of research on the topic might have provided little guidance to practitioners. We demonstrate empirically how noise and correlation impact both the accuracy—the number of correct features selected—and the false detection—the number of incorrect features selected—for five methods: the cardinality-constrained formulation, its Boolean relaxation, ℓ_1 regularization and two methods with non-convex penalties. A cogent feature selection method is expected to exhibit a two-fold convergence, namely the accuracy and false detection rate should converge to 1 and 0 respectively, as the sample size increases. As a result, proper method should recover all and nothing but true features. Empirically, the integer optimization formulation and its Boolean relaxation are the closest to exhibit this two properties consistently in various regimes of noise and correlation. In addition, apart from the discrete optimization approach which requires a substantial, yet often affordable, computational time, all methods terminate in times comparable with the `glmnet` package for Lasso. We released code for methods that were not publicly implemented. Jointly considered, accuracy, false detection and computational time provide a comprehensive assessment of each feature selection method and shed light on alternatives to the Lasso-regularization which are not as popular in practice yet.

Key words and phrases: Feature selection.

REFERENCES

[1] BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci*. **2** 183–202. MR2486527 https://doi.org/10.1137/080716542

[2] BERTSEKAS, D. P. (2016). *Nonlinear Programming*, 3rd ed. *Athena Scientific Optimization and Computation Series*. Athena Scientific, Belmont, MA. MR3587371

[3] BERTSIMAS, D. and FERTIS, A. (2009). On the equivalence of robust optimization and regularization in statistics. Technical Report. Massachusetts Institute of Technology.

[4] BERTSIMAS, D. and KING, A. (2017). Logistic regression: From art to science. *Statist. Sci*. **32** 367–384. MR3696001 https://doi.org/10.1214/16-STS602

[5] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist*. **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

[6] BERTSIMAS, D., PAUPHILET, J. and VAN PARYS, B. (2020). Supplement to "Sparse regression: Scalable algorithms and empirical performance." https://doi.org/10.1214/19-STS701SUPP.

[7] BERTSIMAS, D., PAUPHILET, J. and VAN PARYS, B. (2017). Sparse classification: A scalable discrete optimization perspective. Preprint. arXiv:1710.01352.

[8] BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist*. **48** 300–323. https://doi.org/10.1214/18-AOS1804

[9] BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat*. **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388

[10] BREIMAN, L. (1996). Heuristics of instability and stabilization in model selection. *Ann. Statist*. **24** 2350–2383. MR1425957

Dimitris Bertsimas is the Boeing Professor of Operations Research, Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E62-560, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: dbertsim@mit.edu). Jean Pauphilet is a Doctoral Student, Operations Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E40, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: jpauphilet@london.edu). Bart Van Parys is an Assistant Professor, Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E62-569, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: vanparys@mit.edu).

https://doi.org/10.1214/aos/1032181158
[11] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124
[12] CHU, B. Y., HO, C. H., TSAI, C. H., LIN, C. Y. and LIN, C. J. (2015). Warm start for parameter selection of linear classifiers. In *Proceedings of the* 21*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 149–158. ACM, New York.
[13] DONOHO, D. L. and HUO, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory* **47** 2845–2862. MR1872845 https://doi.org/10.1109/18.959265
[14] DUNNING, I., HUCHETTE, J. and LUBIN, M. (2017). JuMP: A modeling language for mathematical optimization. *SIAM Rev.* **59** 295–320. MR3646493 https://doi.org/10.1137/15M1020575
[15] DURAN, M. A. and GROSSMANN, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Program.* **36** 307–339. MR0866413 https://doi.org/10.1007/BF02592064
[16] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067
[17] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273
[18] FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38** 3567–3604. MR2766861 https://doi.org/10.1214/10-AOS798
[19] FAN, R. E., CHANG, K.-W., HSIEH, C.-J. and LIN, C.-J. (2008). LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.* **9** 1871–1874.
[20] FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. MR2415737 https://doi.org/10.1214/07-AOAS131
[21] 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. and TIBSHIRANI, R. (2013). GLMNet: Lasso and elastic-net regularized generalized linear models. R package version 1.9–5.
[23] FURNIVAL, G. M. and WILSON, R. W. (1974). Regressions by leaps and bounds. *Technometrics* **16** 499–511.
[24] GAMARNIK, D. and ZADIK, I. (2017). High-dimensional regression with binary coefficients. Estimating squared error and a phase transition. *Proc. Mach. Learn. Res.* **65** 948–953.
[25] GAO, H.-Y. and BRUCE, A. G. (1997). WaveShrink with firm shrinkage. *Statist. Sinica* **7** 855–874. MR1488646
[26] GUROBI OPTIMIZATION, I (2016). Gurobi optimizer reference manual; 2016. Available at http://www.gurobi.com.
[27] GUYON, I., WESTON, J., BARNHILL, S. and VAPNIK, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46** 389–422.
[28] HASTIE, T., TIBSHIRANI, R. and TIBSHIRANI, R. J. (2020). Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. *Statist. Sci.* **35** 579–592.
[29] 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
[30] HOERL, A. E. and KENNARD, E. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.
[31] IBM ILOG CPLEX (2009). V12.1: User's manual for CPLEX. International Business Machines Corporation **46** 157.
[32] IGUCHI, T., MIXON, D. G., PETERSON, J. and VILLAR, S. (2015). On the tightness of an SDP relaxation of k-means. Preprint. arXiv:1505.04778.
[33] KENNEY, A., CHIAROMONTE, F. and FELICI, G. (2018). Efficient and effective L_0 feature selection. Preprint. arXiv:1808.02526.
[34] LOH, P.-L. and WAINWRIGHT, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *Ann. Statist.* **45** 2455–2482. MR3737898 https://doi.org/10.1214/16-AOS1530
[35] LUBIN, M. and DUNNING, I. (2015). Computing in operations research using Julia. *INFORMS J. Comput.* **27** 238–248. MR3347876 https://doi.org/10.1287/ijoc.2014.0623
[36] MALLAT, S. G. and ZHANG, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41** 3397–3415.
[37] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281
[38] NATARAJAN, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM J. Comput.* **24** 227–234. MR1320206 https://doi.org/10.1137/S0097539792240406
[39] NEDIĆ, A. and OZDAGLAR, A. (2008). Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optim.* **19** 1757–1780. MR2486049 https://doi.org/10.1137/070708111
[40] PILANCI, M., WAINWRIGHT, M. J. and EL GHAOUI, L. (2015). Sparse learning via Boolean relaxations. *Math. Program.* **151** 63–87. MR3347550 https://doi.org/10.1007/s10107-015-0894-1
[41] SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026
[42] SU, W., BOGDAN, M. and CANDÈS, E. (2017). False discoveries occur early on the Lasso path. *Ann. Statist.* **45** 2133–2150. MR3718164 https://doi.org/10.1214/16-AOS1521
[43] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
[44] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018
[45] WAINWRIGHT, M. J. (2009). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory* **55** 5728–5741. MR2597190 https://doi.org/10.1109/TIT.2009.2032816
[46] WANG, W., WAINWRIGHT, M. J. and RAMCHANDRAN, K. (2010). Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices. *IEEE Trans. Inf. Theory* **56** 2967–2979. MR2683451 https://doi.org/10.1109/TIT.2010.2046199
[47] WU, T. T. and LANGE, K. (2008). Coordinate descent algorithms for lasso penalized regression. *Ann. Appl. Stat.* **2** 224–244. MR2415601 https://doi.org/10.1214/07-AOAS147
[48] XU, H., CARAMANIS, C. and MANNOR, S. (2009). Robustness and regularization of support vector machines. *J. Mach. Learn. Res.* **10** 1485–1510. MR2534869
[49] YE, Y. (2016). Data randomness makes optimization problems easier to solve?
[50] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729
[51] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

[52] 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 https://doi.org/10.1111/j.1467-9868.2005.00503.x

[53] Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

Statistical Science
2020, Vol. 35, No. 4, 579–592
https://doi.org/10.1214/19-STS733

Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons

Trevor Hastie, Robert Tibshirani and Ryan Tibshirani

Abstract. In exciting recent work, Bertsimas, King and Mazumder (*Ann. Statist.* **44** (2016) 813–852) showed that the classical best subset selection problem in regression modeling can be formulated as a mixed integer optimization (MIO) problem. Using recent advances in MIO algorithms, they demonstrated that best subset selection can now be solved at much larger problem sizes than what was thought possible in the statistics community. They presented empirical comparisons of best subset with other popular variable selection procedures, in particular, the lasso and forward stepwise selection. Surprisingly (to us), their simulations suggested that best subset consistently outperformed both methods in terms of prediction accuracy. Here, we present an expanded set of simulations to shed more light on these comparisons. The summary is roughly as follows:

- neither best subset nor the lasso uniformly dominate the other, with best subset generally performing better in very high signal-to-noise (SNR) ratio regimes, and the lasso better in low SNR regimes;
- for a large proportion of the settings considered, best subset and forward stepwise perform similarly, but in certain cases in the high SNR regime, best subset performs better;
- forward stepwise and best subsets tend to yield sparser models (when tuned on a validation set), especially in the high SNR regime;
- the relaxed lasso (actually, a simplified version of the original relaxed estimator defined in Meinshausen (*Comput. Statist. Data Anal.* **52** (2007) 374–393)) is the overall winner, performing just about as well as the lasso in low SNR scenarios, and nearly as well as best subset in high SNR scenarios.

Key words and phrases: Regression, selection, penalization.

REFERENCES

BEALE, E. M. L., KENDALL, M. G. and MANN, D. W. (1967). The discarding of variables in multivariate analysis. *Biometrika* **54** 357–366. MR0221656 https://doi.org/10.1093/biomet/54.3-4.357

BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324 https://doi.org/10.1093/biomet/asr043

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523

CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094 https://doi.org/10.1137/S1064827596304010

DAVIS, G., MALLAT, S. and ZHANG, Z. (1994). Adaptive time-

Trevor Hastie is Professor, Departments of Statistics and Biomedical Data Science, Stanford University, Stanford, California 94305, USA (e-mail: hastie@stanford.edu). Robert Tibshirani is Professor, Departments of Statistics and Biomedical Data Science, Stanford University, Stanford, California 94305, USA (e-mail: tibs@stanford.edu). Ryan Tibshirani is Associate Professor, Departments of Statistics and Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania 15217, USA (e-mail: ryantibs@cmu.edu).

frequency decompositions with matching pursuit. *Wavelet Appl*. **402** 402–413.

DRAPER, N. R. and SMITH, H. (1966). *Applied Regression Analysis*. Wiley, New York. MR0211537

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann*. *Statist*. **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

EFROYMSON, M. (1966). Stepwise regression—A backward and forward look. Paper presented at the *Eastern Regional Meetings of the Institute of Mathematical Statistics*.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist*. *Assoc*. **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

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

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann*. *Appl*. *Stat*. **1** 302–332. MR2415737 https://doi.org/10.1214/07-AOAS131

FURNIVAL, G. and WILSON, R. (1974). Regression by leaps and bounds. *Technometrics* **16** 499–511.

GOLUB, G. H. and VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd ed. *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1417720

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 https://doi.org/10.1007/978-0-387-84858-7

HASTIE, T., TIBSHIRANI, R. and TIBSHIRANI, R. (2020). Supplement to "Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons." https://doi.org/10.1214/19-STS733SUPP

HAZIMEH, H. and MAZUMDER, R. (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. Preprints. Available at arXiv:1803.01454.

HOCKING, R. R. and LESLIE, R. N. (1967). Selection of the best subset in regression analysis. *Technometrics* **9** 531–540. MR0219192 https://doi.org/10.2307/1266192

JANSON, L., FITHIAN, W. and HASTIE, T. J. (2015). Effective degrees of freedom: A flawed metaphor. *Biometrika* **102** 479–485. MR3371017 https://doi.org/10.1093/biomet/asv019

KAUFMAN, S. and ROSSET, S. (2014). When does more regularization imply fewer degrees of freedom? Sufficient conditions and counterexamples. *Biometrika* **101** 771–784. MR3286916 https://doi.org/10.1093/biomet/asu034

MALLAT, S. and ZHANG, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans*. *Signal Process*. **41** 3397–3415.

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). *SparseNet*: Coordinate descent with nonconvex penalties. *J. Amer. Statist*. *Assoc*. **106** 1125–1138. MR2894769 https://doi.org/10.1198/jasa.2011.tm09738

MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput*. *Statist*. *Data Anal*. **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

NATARAJAN, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM J. Comput*. **24** 227–234. MR1320206 https://doi.org/10.1137/S0097539792240406

RADCHENKO, P. and JAMES, G. M. (2011). Improved variable selection with forward-Lasso adaptive shrinkage. *Ann*. *Appl*. *Stat*. **5** 427–448. MR2810404 https://doi.org/10.1214/10-AOAS375

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

TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron*. *J. Stat*. **7** 1456–1490. MR3066375 https://doi.org/10.1214/13-EJS815

TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann*. *Statist*. **40** 1198–1232. MR2985948 https://doi.org/10.1214/12-AOS1003

TIBSHIRANI, R., BIEN, J., FRIEDMAN, J., HASTIE, T., SIMON, N., TAYLOR, J. and TIBSHIRANI, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *J. R. Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **74** 245–266. MR2899862 https://doi.org/10.1111/j.1467-9868.2011.01004.x

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann*. *Statist*. **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the lasso. *Ann*. *Statist*. **35** 2173–2192. MR2363967 https://doi.org/10.1214/009053607000000127

Statistical Science
2020, Vol. 35, No. 4, 593–601
https://doi.org/10.1214/20-STS806

A Discussion on Practical Considerations with Sparse Regression Methodologies

Owais Sarwar, Benjamin Sauk and Nikolaos V. Sahinidis

Abstract. Sparse linear regression is a vast field and there are many different algorithms available to build models. Two new papers published in *Statistical Science* study the comparative performance of several sparse regression methodologies, including the lasso and subset selection. Comprehensive empirical analyses allow the researchers to demonstrate the relative merits of each estimator and provide guidance to practitioners. In this discussion, we summarize and compare the two studies and we examine points of agreement and divergence, aiming to provide clarity and value to users. The authors have started a highly constructive dialogue, our goal is to continue it.

Key words and phrases: Sparse regression, subset selection, lasso.

REFERENCES

[1] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388
[2] BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804
[3] BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388
[4] BREIMAN, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37** 373–384. MR1365720 https://doi.org/10.2307/1269730
[5] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523
[6] CELENTANO, M. and MONTANARI, A. (2019). Fundamental barriers to high-dimensional regression with convex penalties. Available at arXiv:1903.10603.
[7] COZAD, A., SAHINIDIS, N. V. and MILLER, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE J.* **60** 2211–2227. https://doi.org/10.1002/aic.14418
[8] COZAD, A., SAHINIDIS, N. V. and MILLER, D. C. (2015). A combined first-principles and data-driven approach to model building. *Comput. Chem. Eng.* **73** 116–127. https://doi.org/10.1016/J.COMPCHEMENG.2014.11.010
[9] DAS, A. and KEMPE, D. (2018). Approximate submodularity and its applications: Subset selection, sparse approximation and dictionary selection. *J. Mach. Learn. Res.* **19** Paper No. 3, 34. MR3862410
[10] DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. With electronic supplementary materials available online. MR2546388 https://doi.org/10.1098/rsta.2009.0152
[11] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273
[12] FRANK, I. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–135.
[13] FURNIVAL, G. M. and WILSON, R. W. (1974). Regressions by leaps and bounds. *Technometrics* **16** 499–511.
[14] GAMARNIK, D. and ZADIK, I. (2017). High-dimensional regression with binary coefficients. Estimating squared error and a phase transition. Available at arXiv:1701.04455.
[15] 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
[16] HAZIMEH, H. and MAZUMDER, R. (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. Available at arXiv:1803.01454.
[17] HAZIMEH, H., MAZUMDER, R. and SAAB, A. (2020). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. Available at arXiv:2004:06152.
[18] HOCKING, R. R. and LESLIE, R. N. (1967). Selection of the best subset in regression analysis. *Technometrics* **9** 531–540. MR0219192 https://doi.org/10.2307/1266192

Owais Sarwar is a PhD student, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: osarwar@andrew.cmu.edu). Benjamin Sauk is a PhD student, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: bsauk@andrew.cmu.edu). Nikolaos V. Sahinidis is the Gary C. Butler Family Chair and Professor, H. Milton Stewart School of Industrial & Systems Engineering and School of Chemical & Biomolecular Engineering, Georgia Institute of Technology Atlanta, Georgia 30332, USA (e-mail: nikos@gatech.edu).

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

[20] LOH, P.-L. and WAINWRIGHT, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *Ann. Statist.* **45** 2455–2482. MR3737898 https://doi.org/10.1214/16-AOS1530

[21] MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769 https://doi.org/10.1198/jasa.2011.tm09738

[22] MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput. Statist. Data Anal.* **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

[23] MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523 https://doi.org/10.1111/j.1467-9868.2010.00740.x

[24] PILANCI, M., WAINWRIGHT, M. J. and EL GHAOUI, L. (2015). Sparse learning via Boolean relaxations. *Math. Program.* **151** 63–87. MR3347550 https://doi.org/10.1007/s10107-015-0894-1

[25] STIGLER, S. M. (1988). Gauss and the invention of least squares. *Ann. Statist.* **2** 347–370.

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

[27] WAINWRIGHT, M. J. (2009). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory* **55** 5728–5741. MR2597190 https://doi.org/10.1109/TIT.2009.2032816

[28] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

[29] WANG, F., MUKHERJEE, S., RICHARDSON, S. and HILL, S. M. (2020). High-dimensional regression in practice: An empirical study of finite-sample prediction, variable selection and ranking. *Stat. Comput.* **30** 697–719. MR4065227 https://doi.org/10.1007/s11222-019-09914-9

[30] WANG, S., NAN, B., ROSSET, S. and ZHU, J. (2011). Random Lasso. *Ann. Appl. Stat.* **5** 468–485. MR2810406 https://doi.org/10.1214/10-AOAS377

[31] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

[32] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

[33] 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 https://doi.org/10.1111/j.1467-9868.2005.00503.x

Statistical Science
2020, Vol. 35, No. 4, 602–608
https://doi.org/10.1214/20-STS807

Discussion of "Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons"

Rahul Mazumder

REFERENCES

[1] ASUR, S. and HUBERMAN, B. A. (2010). Predicting the future with social media. In 2010 *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* **1** 492–499. IEEE, New York.

[2] ATAMTURK, A. and GOMEZ, A. (2019). Rank-one convexification for sparse regression. Preprint. Available at arXiv:1901.10334.

[3] BAARDMAN, L., COHEN, M. C., PANCHAMGAM, K., PERAKIS, G. and SEGEV, D. (2019). Scheduling promotion vehicles to boost profits. *Manage. Sci.* **65** 50–70.

[4] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

[5] BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

[6] BOLLEN, J., MAO, H. and ZENG, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science* **2** 1–8.

[7] COHEN, M. C., LEUNG, N.-H. Z., PANCHAMGAM, K., PERAKIS, G. and SMITH, A. (2017). The impact of linear optimization on promotion planning. *Oper. Res.* **65** 446–468. MR3647851 https://doi.org/10.1287/opre.2016.1573

[8] DONG, H., CHEN, K. and LINDEROTH, J. (2015). Regularization vs. Relaxation: A conic optimization perspective of statistical variable selection. ArXiv e-prints.

[9] FREUND, R. M., GRIGAS, P. and MAZUMDER, R. (2017). A new perspective on boosting in linear regression via subgradient optimization and relatives. *Ann. Statist.* **45** 2328–2364. MR3737894 https://doi.org/10.1214/16-AOS1505

[10] GHOSE, A. and IPEIROTIS, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng.* **23** 1498–1512.

[11] 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 https://doi.org/10.1007/978-0-387-84858-7

[12] 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

[13] HAZIMEH, H. and MAZUMDER, R. (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Oper. Res.* **68** 1285–1624.

[14] HAZIMEH, H., MAZUMDER, R. and SAAB, A. (2020). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. Preprint. Available at arXiv:2004.06152.

[15] HAZIMEH, H., PONOMAREVA, N., MOL, P., TAN, Z. and MAZUMDER, R. (2020). The tree ensemble layer: Differentiability meets conditional computation. In *Proceedings of the 37th Annual International Conference on Machine Learning, ICML. Vol. 20.*

[16] MAZUMDER, R., RADCHENKO, P. and DEDIEU, A. (2017). Subset selection with shrinkage: Sparse linear modeling when the snr is low. Preprint. Available at arXiv:1708.03288.

[17] MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput. Statist. Data Anal.* **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

[18] SMITH, S. A. and ACHABAL, D. D. (1998). Clearance pricing and inventory policies for retail chains. *Manage. Sci.* **44** 285–300.

[19] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

[20] XIE, W. and DENG, X. (2020). Scalable algorithms for the sparse ridge regression.

[21] ZHOU, D., ZHENG, L., ZHU, Y., LI, J. and HE, J. (2020). Domain adaptive multi-modality neural attention network for financial forecasting. In *Proceedings of the Web Conference* 2020 2230–2240.

Rahul Mazumder is the Robert G. James Career Development Professor and Associate Professor of Operations Research and Statistics at the MIT Sloan School of Management. He is also afilliated with the Operations Research Center, MIT Center for Statistics and MIT IBM Watson AI Lab. Address: 100 Main Street, Cambridge-02142, USA (e-mail: rahulmaz@mit.edu).

Statistical Science
2020, Vol. 35, No. 4, 609–613
https://doi.org/10.1214/20-STS808

Modern Variable Selection in Action: Comment on the Papers by HTT and BPV

Edward I. George

REFERENCES

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BERTSIMAS, D., PAUPHILET, J. and VAN PARYS, B. (2017). Sparse Classification: a scalable discrete optimization perspective. Under revision.

BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

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

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

ROČKOVÁ, V. and GEORGE, E. I. (2018). The spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. MR3803476 https://doi.org/10.1080/01621459.2016.1260469

STEIN, C. M. (1960). *Multiple Regression. Essays in Honor of Harold Hotelling*. Stanford Univ. Press, Palo Alto, CA.

ZHOU, H. H. and HWANG, J. T. G. (2005). Minimax estimation with thresholding and its application to wavelet analysis. *Ann. Statist.* **33** 101–125. MR2157797 https://doi.org/10.1214/009053604000000977

Edward I. George is the Universal Furniture Professor of Statistics and Economics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA (e-mail: edgeorge@wharton.upenn.edu).

Statistical Science
2020, Vol. 35, No. 4, 614–622
https://doi.org/10.1214/20-STS809

A Look at Robustness and Stability of ℓ_1- versus ℓ_0-Regularization: Discussion of Papers by Bertsimas et al. and Hastie et al.

Yuansi Chen, Armeen Taeb and Peter Bühlmann

Abstract. We congratulate the authors Bertsimas, Pauphilet and van Parys (hereafter BPvP) and Hastie, Tibshirani and Tibshirani (hereafter HTT) for providing fresh and insightful views on the problem of variable selection and prediction in linear models. Their contributions at the fundamental level provide guidance for more complex models and procedures.

Key words and phrases: Distributional robustness, high-dimensional estimation, latent variables, low-rank estimation, variable selection.

REFERENCES

ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J. et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. MR0483125

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210

BELLEC, P. (2018). The noise barrier and the large signal bias of the Lasso and other convex estimators. Preprint. Available at arXiv:1804.01230.

BERTSIMAS, D. and COPENHAVER, M. S. (2018). Characterization of the equivalence of robustification and regularization in linear and matrix regression. *European J. Oper. Res.* **270** 931–942. MR3814540 https://doi.org/10.1016/j.ejor.2017.03.051

BERTSIMAS, D., COPENHAVER, M. S. and MAZUMDER, R. (2017). Certifiably optimal low rank factor analysis. *J. Mach. Learn. Res.* **18** Paper No. 29, 53. MR3634896

BERTSIMAS, D. and KING, A. (2017). Logistic regression: From art to science. *Statist. Sci.* **32** 367–384. MR3696001 https://doi.org/10.1214/16-STS602

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BERTSIMAS, D. and SHIODA, R. (2009). Algorithm for cardinality-constrained quadratic optimizaiton. *Comput. Optim. Appl.* **43** 1–22. MR2501042 https://doi.org/10.1007/s10589-007-9126-9

BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

BREIMAN, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37** 373–384. MR1365720 https://doi.org/10.2307/1269730

BREIMAN, L. (1996a). Bagging predictors. *Mach. Learn.* **24** 123–140.

BREIMAN, L. (1996b). Heuristics of instability and stabilization in model selection. *Ann. Statist.* **24** 2350–2383. MR1425957 https://doi.org/10.1214/aos/1032181158

BÜHLMANN, P. (2006). Boosting for high-dimensional linear models. *Ann. Statist.* **34** 559–583. MR2281878 https://doi.org/10.1214/009053606000000092

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

BÜHLMANN, P. and VAN DE GEER, S. (2018). Statistics for big data: A perspective. *Statist. Probab. Lett.* **136** 37–41. MR3806834 https://doi.org/10.1016/j.spl.2018.02.016

BURER, S. and MONTEIRO, R. D. C. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.* **95** 329–357. MR1976484 https://doi.org/10.1007/s10107-002-0352-8

CAI, T. T. and WANG, L. (2011). Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* **57** 4680–4688. MR2840484 https://doi.org/10.1109/TIT.2011.2146090

CEVID, D., BÜHLMANN, P. and MEINSHAUSEN, N. (2018). Spectral deconfounding and perturbed sparse linear models. Preprint. Available at arXiv:1811.05352.

Yuansi Chen is Postdoctoral Fellow, Seminar for Statistics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland (e-mail: chen@stat.math.ethz.ch). Armeen Taeb is Postdoctoral Fellow, Seminar for Statistics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland (e-mail: armeen.taeb@stat.math.ethz.ch). Peter Bühlmann is Professor, Seminar for Statistics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland (e-mail: buhlmann@stat.math.ethz.ch).

CHEN, Y. and CHEN, Y. (2018). Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization. *IEEE Signal Process. Mag.* **35** 14–31.

CHEN, S. and DONOHO, D. (1994). Basis pursuit. In *Proceedings of* 1994 28*th Asilomar Conference on Signals, Systems and Computers* **1** 41–44. IEEE, New York.

CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev.* **43** 129–159. MR1854649 https://doi.org/10.1137/S003614450037906X

CHEN, Y., TAEB, A. and BÜHLMANN, P. (2020). Supplement to "A look at robustness and stability of ℓ_1- versus ℓ_0-regularization: discussion of papers by Bertsimas et al. and Hastie et al." https://doi.org/10.1214/20-STS809SUPP

CHICKERING, D. M. (2003). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554. MR1991085 https://doi.org/10.1162/153244303321897717

DONOHO, D. L. (2006). For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.* **59** 797–829. MR2217606 https://doi.org/10.1002/cpa.20132

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

FAZEL, M. (2002). Matrix rank minimization with applications. Ph.D. thesis, Stanford, CA.

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

FRIEDRICH, J., ZHOU, P. and PANINSKI, L. (2017). Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13** 1–26.

FUCHS, J.-J. (2004). On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory* **50** 1341–1344. MR2094894 https://doi.org/10.1109/TIT.2004.828141

GATU, C. and KONTOGHIORGHES, E. J. (2006). Branch-and-bound algorithms for computing the best-subset regression models. *J. Comput. Graph. Statist.* **15** 139–156. MR2269366 https://doi.org/10.1198/106186006X100290

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 https://doi.org/10.1007/978-0-387-84858-7

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

HOFMANN, M., GATU, C. and KONTOGHIORGHES, E. J. (2007). Efficient algorithms for computing the best subset regression models for large-scale problems. *Comput. Statist. Data Anal.* **52** 16–29. MR2409961 https://doi.org/10.1016/j.csda.2007.03.017

JAIN, P., TEWARI, A. and KAR, P. (2014). On iterative hard thresholding methods for high-dimensional M-estimation. In *Advances in Neural Information Processing Systems* 685–693.

JEWELL, S. and WITTEN, D. (2018). Exact spike train inference via ℓ_0 optimization. *Ann. Appl. Stat.* **12** 2457–2482. MR3875708 https://doi.org/10.1214/18-AOAS1162

KOTZ, S. and JOHNSON, N. L., eds. (1992). *Breakthroughs in Statistics. Vol. I: Foundations and Basic Theory. Springer Series in Statistics. Perspectives in Statistics*. Springer, New York. MR1182794

MALLAT, S. and ZHANG, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41** 3397–3415.

MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput. Statist. Data Anal.* **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

MILLER, A. J. (1990). *Subset Selection in Regression. Monographs on Statistics and Applied Probability* **40**. CRC Press, London. MR1072361 https://doi.org/10.1007/978-1-4899-2939-6

PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N. et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* 8026–8037.

ROTHENHÄUSLER, D., MEINSHAUSEN, N., BÜHLMANN, P. and PETERS, J. (2018). Anchor regression: Heterogeneous data meets causality. Preprint. Available at arXiv:1801.06229. To appear in J. Roy. Statist. Soc. Ser. B.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

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

TROPP, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* **50** 2231–2242. MR2097044 https://doi.org/10.1109/TIT.2004.834793

VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316 https://doi.org/10.1214/09-EJS506

VAN DE GEER, S. and BÜHLMANN, P. (2013). ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. *Ann. Statist.* **41** 536–567. MR3099113 https://doi.org/10.1214/13-AOS1085

VAN DE GEER, S., BÜHLMANN, P. and ZHOU, S. (2011). The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso). *Electron. J. Stat.* **5** 688–749. MR2820636 https://doi.org/10.1214/11-EJS624

WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

XU, H., CARAMANIS, C. and MANNOR, S. (2009). Robust regression and Lasso. In *Advances in Neural Information Processing Systems* 1801–1808.

ZHANG, T. (2011). Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inf. Theory* **57** 6215–6221. MR2857968 https://doi.org/10.1109/TIT.2011.2162263

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 https://doi.org/10.1198/016214506000000735

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 https://doi.org/10.1111/j.1467-9868.2005.00503.x

Statistical Science
2020, Vol. 35, No. 4, 623–624
https://doi.org/10.1214/20-STS701REJ

Rejoinder: Sparse Regression: Scalable Algorithms and Empirical Performance

Dimitris Bertsimas, Jean Pauphilet and Bart Van Parys

REFERENCES

[1] BERTSIMAS, D. and COPENHAVER, M. S. (2018). Characterization of the equivalence of robustification and regularization in linear and matrix regression. *European J. Oper. Res.* **270** 931–942. MR3814540 https://doi.org/10.1016/j.ejor.2017.03.051

[2] BERTSIMAS, D. and DUNN, J. (2019). *Machine Learning Under a Modern Optimization Lens*. Dynamic Ideas Press, Belmont, MA.

[3] BERTSIMAS, D. and KING, A. (2016). OR forum—an algorithmic approach to linear regression. *Oper. Res.* **64** 2–16. MR3463258 https://doi.org/10.1287/opre.2015.1436

[4] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

[5] BERTSIMAS, D. and LI, M. L. (2020). Scalable holistic linear regression. *Oper. Res. Lett.* **48** 203–208. MR4077406 https://doi.org/10.1016/j.orl.2020.02.008

[6] BERTSIMAS, D., PAUPHILLET, J. and VAN PARYS, B. (2020). Sparse regression: Scalable algorithms and empirical performance. *Statist. Sci.* **35** 555–578.

[7] BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

[8] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124

[9] CHEN, Y., TAEB, A. and BÜHLMANN, P. (2020). A Look at Robustness and Stability of ℓ_1-versus ℓ_0-Regularization: Discussion of Papers by Bertsimas et al. and Hastie et al. *Statist. Sci.* **35** 614–622.

[10] GEORGE, E. (2020). Modern variable selection in action: Comment on the papers by HTT and BPV. *Statist. Sci.* **35** 609–613.

[11] HASTIE, T., TIBSHIRANI, R. and TIBSHIRANI, R. (2020). Best subset, forward stepwise, or lasso? Analysis and recommendations based on extensive comparisons. *Statist. Sci.* **35** 579–592.

[12] 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

[13] HAZIMEH, H. and MAZUMDER, R. (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. Available at arXiv:1803.01454.

[14] HAZIMEH, H., MAZUMDER, R. and SAAB, A. (2020). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. Available at arXiv:2004:06152.

[15] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

[16] SARWAR, O., BENJAMIN SAUK, B. and SAHINIDIS, N. (2020). A discussion on practical considerations with sparse regression methodologies. *Statist. Sci.* **35** 593–601.

[17] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

[18] XU, H., CARAMANIS, C. and MANNOR, S. (2009). Robustness and regularization of support vector machines. *J. Mach. Learn. Res.* **10** 1485–1510. MR2534869

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

Dimitris Bertsimas is the Boeing Professor of Operations Research, Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E62-560, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: dbertsim@mit.edu). Jean Pauphilet is a Doctoral Student, Operations Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E40, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: jpauph@mit.edu). Bart Van Parys is an Assistant Professor, Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue Bldg. E62-569, Cambridge, Belmont, Massachusetts 02139, USA (e-mail: vanparys@mit.edu).

Statistical Science
2020, Vol. 35, No. 4, 625–626
https://doi.org/10.1214/20-STS733REJ

Rejoinder: Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons

Trevor Hastie, Robert Tibshirani and Ryan J. Tibshirani

REFERENCES

HAZIMEH, H. and MAZUMDER, R. (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Oper. Res.* **68** 1517–1537.

JOHNDROW, J. E., ORENSTEIN, P. and BHATTACHARYA, A. (2020). Bayes shrinkage at GWAS scale: Convergence and approximation theory of a scalable MCMC algorithm for the horseshoe prior. *J. Mach. Learn. Res.*. To appear.

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769 https://doi.org/10.1198/jasa.2011.tm09738

Trevor Hastie is Professor, Departments of Statistics and Biomedical Data Science, Stanford University, Stanford, California 94305, USA (e-mail: hastie@stanford.edu). Robert Tibshirani is Professor, Departments of Statistics and Biomedical Data Science, Stanford University, Stanford, California 94305, USA (e-mail: tibs@stanford.edu). Ryan J. Tibshirani is Associate Professor, Departments of Statistics and Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania 15217, USA (e-mail: ryantibs@cmu.edu).

Statistical Science
2020, Vol. 35, No. 4, 627–662
https://doi.org/10.1214/19-STS743

Exponential-Family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios

Michael Schweinberger, Pavel N. Krivitsky, Carter T. Butts and Jonathan R. Stewart

Abstract. Exponential-family Random Graph Models (ERGMs) constitute a large statistical framework for modeling dense and sparse random graphs with short- or long-tailed degree distributions, covariate effects and a wide range of complex dependencies. Special cases of ERGMs include network equivalents of generalized linear models (GLMs), Bernoulli random graphs, β-models, p_1-models and models related to Markov random fields in spatial statistics and image processing. While ERGMs are widely used in practice, questions have been raised about their theoretical properties. These include concerns that some ERGMs are near-degenerate and that many ERGMs are non-projective. To address such questions, careful attention must be paid to model specifications and their underlying assumptions, and to the inferential settings in which models are employed. As we discuss, near-degeneracy can affect simplistic ERGMs lacking structure, but well-posed ERGMs with additional structure can be well-behaved. Likewise, lack of projectivity can affect non-likelihood-based inference, but likelihood-based inference does not require projectivity. Here, we review well-posed ERGMs along with likelihood-based inference. We first clarify the core statistical notions of "sample" and "population" in the ERGM framework, separating the process that generates the population graph from the observation process. We then review likelihood-based inference in finite, super and infinite population scenarios. We conclude with consistency results, and an application to human brain networks.

Key words and phrases: Social network, exponential-family random graph model, ERGM, model degeneracy, projectivity.

REFERENCES

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

ALMQUIST, Z. W. and BAGOZZI, B. E. (2019). Using radical environmentalist texts to uncover network structure and network features. *Sociol. Methods Res.* **48** 905–960. MR4003633 https://doi.org/10.1177/0049124117729696

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 https://doi.org/10.1214/13-AOS1138

ARISTOFF, D. and RADIN, C. (2013). Emergent structures in large networks. *J. Appl. Probab.* **50** 883–888. MR3102521 https://doi.org/10.1239/jap/1378401243

ASUNCION, A., LIU, Q., IHLER, A. T. and SMYTH, P. (2010). Learning with blocks: Composite likelihood and contrastive divergence. In *Thirteenth International Conference on AI and Statistics* 33–40.

ATCHADÉ, Y. F., LARTILLOT, N. and ROBERT, C. (2013). Bayesian computation for statistical models with intractable normalizing constants. *Braz. J. Probab. Stat.* **27** 416–436. MR3105037 https://doi.org/10.1214/11-BJPS174

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

Michael Schweinberger is Assistant Professor, Department of Statistics, Rice University, Houston, Texas 77005, USA (e-mail: m.s@rice.edu). Pavel N. Krivitsky is Lecturer, Department of Statistics, University of New South Wales, Sydney, New South Wales 2052, Australia (e-mail: p.krivitsky@unsw.edu.au). Carter T. Butts is Professor, Departments of Sociology, Statistics, Computer Science and Electrical Engineering and Computer Science, University of California, Irvine, California 92697, USA (e-mail: buttsc@uci.edu). Jonathan R. Stewart is graduate student, Department of Statistics, Rice University, Houston, Texas 77005, USA (e-mail: jrs6@rice.edu).

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

BEARMAN, P. S., MOODY, J. and STOVEL, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *Amer. J. Sociol.* **110** 44–91.

BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Stat.* **43** 193–204. MR0298810

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

BHAMIDI, S., BRESLER, G. and SLY, A. (2008). Mixing time of exponential random graphs. In 2008 *IEEE* 49*th Annual IEEE Symposium on Foundations of Computer Science* 803–812.

BHAMIDI, S., BRESLER, G. and SLY, A. (2011). Mixing time of exponential random graphs. *Ann. Appl. Probab.* **21** 2146–2170. MR2895412 https://doi.org/10.1214/10-AAP740

BHAMIDI, S., CHAKRABORTY, S., CRANMER, S. and DESMARAIS, B. (2018). Weighted exponential random graph models: Scope and large network limits. *J. Stat. Phys.* **173** 704–735. MR3876904 https://doi.org/10.1007/s10955-018-2103-0

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. In *Proceedings of the National Academy of Sciences* **106** 21068–21073.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

BINKIEWICZ, N., VOGELSTEIN, J. T. and ROHE, K. (2017). Covariate-assisted spectral clustering. *Biometrika* **104** 361–377. MR3698259 https://doi.org/10.1093/biomet/asx008

BOLLOBÁS, B. (1985). *Random Graphs*. Academic Press [Harcourt Brace Jovanovich, Publishers], London. MR0809996

BOLLOBÁS, B., RIORDAN, O., SPENCER, J. and TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277 https://doi.org/10.1002/rsa.1009

BORGS, C., CHAYES, J. T., COHN, H. and VEITCH, V. (2019). Sampling perspectives on sparse exchangeable graphs. *Ann. Probab.* **47** 2754–2800. MR4021236 https://doi.org/10.1214/18-AOP1320

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

BRAILLY, J., FAVRE, G., CHATELLET, J. and LAZEGA, E. (2016). Embeddedness as a multilevel problem: A case study in economic sociology. *Soc. Netw.* **44** 319–333.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

BUTTS, C. T. (2008). A relational event framework for social action. *Sociol. Method.* **38** 155–200.

BUTTS, C. T. (2011). Bernoulli graph bounds for general random graph models. *Sociol. Method.* **41** 299–345.

BUTTS, C. T. (2015). A novel simulation method for binary discrete exponential families, with application to social networks. *J. Math. Sociol.* **39** 174–202. MR3367717 https://doi.org/10.1080/0022250X.2015.1022279

BUTTS, C. T. (2018). A perfect sampling method for exponential family random graph models. *J. Math. Sociol.* **42** 17–36. MR3764794 https://doi.org/10.1080/0022250X.2017.1396985

BUTTS, C. T. (2019). A dynamic process interpretation of the sparse ERGM reference model. *J. Math. Sociol.* **43** 40–57. MR3891856 https://doi.org/10.1080/0022250X.2018.1490737

BUTTS, C. T. and ACTON, R. M. (2011). Spatial modeling of social networks. In *The SAGE Handbook of GIS and Society Research* (T. Nyerges, H. Couclelis and R. McMaster, eds.) 222–250 12. SAGE, Thousand Oaks.

BUTTS, C. T. and ALMQUIST, Z. W. (2015). A flexible parameterization for baseline mean degree in multiple-network ERGMs. *J. Math. Sociol.* **39** 163–167. MR3367715 https://doi.org/10.1080/0022250X.2014.967851

BYSHKIN, M., STIVALA, A., MIRA, A., ROBINS, G. and LOMI, A. (2018). Fast maximum likelihood estimation via equilibrium expectation for large network data. *Sci. Rep.* **8** 11509. https://doi.org/10.1038/s41598-018-29725-8

CAI, D., CAMPBELL, T. and BRODERICK, T. (2016). Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 4249–4257.

CAIMO, A. and FRIEL, N. (2011). Bayesian inference for exponential random graph models. *Soc. Netw.* **33** 41–55.

CAIMO, A. and FRIEL, N. (2013). Bayesian model selection for exponential random graph models. *Soc. Netw.* **35** 11–24.

CAIMO, A. and GOLLINI, I. (2020). A multilayer exponential random graph modelling approach for weighted networks. *Comput. Statist. Data Anal.* **142** 106825, 18. MR3995271 https://doi.org/10.1016/j.csda.2019.106825

CAIMO, A. and MIRA, A. (2015). Efficient computational strategies for doubly intractable problems with applications to Bayesian social networks. *Stat. Comput.* **25** 113–125. MR3304912 https://doi.org/10.1007/s11222-014-9516-7

CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. MR3731666 https://doi.org/10.1111/rssb.12233

CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053

CORANDER, J., DAHMSTRÖM, K. and DAHMSTRÖM, P. (1998). Maximum likelihood estimation for Markov graphs Technical Report Department of Statistics, Univ. Stockholm.

CORANDER, J., DAHMSTRÖM, K. and DAHMSTRÖM, P. (2002). Maximum likelihood estimation for exponential random graph models. In *Contributions to Social Network Analysis*, *Information Theory*, *and Other Topics in Statistics*; *A Festschrift in Honour of Ove Frank* (J. Hagberg, ed.) 1–17. Dept. Statistics, Univ. Stockholm.

CRANE, H. (2018). *Probabilistic Foundations of Statistical Network Analysis*. *Monographs on Statistics and Applied Probability* **157**. CRC Press, Boca Raton, FL. MR3791467

CRANE, H. and DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113** 1311–1326. MR3862359 https://doi.org/10.1080/01621459.2017.1341413

CRANE, H. and DEMPSEY, W. (2020). A statistical framework for modern network science. *Statist. Sci.* To appear.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

DAHMSTRÖM, K. and DAHMSTRÖM, P. (1993). ML-estimation of the clustering parameter in a Markov graph model Technical Report Univ. Stockholm, Department of Statistics.

DAHMSTRÖM, K. and DAHMSTRÖM, P. (1999). Properties of different estimators of the parameters in Markov graphs. In

Bulletin of the International Statistical Institute 1–2. International Statistical Institute. Available at https://tilastokeskus.fi/isi99/proceedings/arkisto/varasto/dahm0777.pdf.

DAWID, A. P. and DICKEY, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.* **72** 845–850. MR0471124

DESMARAIS, B. A. and CRANMER, S. J. (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PLoS ONE* **7** 1–12.

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

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3** 1189–1242. MR0428531

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376. MR0471152

ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

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

EVERITT, R. G. (2012). Bayesian parameter estimation for latent Markov random fields and social networks. *J. Comput. Graph. Statist.* **21** 940–960. MR3005805 https://doi.org/10.1080/10618600.2012.687493

FELLOWS, I. and HANDCOCK, M. S. (2012). Exponential-family random network models. Available at arXiv:1208.0121.

FELLOWS, I. and HANDCOCK, M. S. (2017). Removing phase transitions from Gibbs measures. In *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.) **54** 289–297. Proceedings of Machine Learning Research.

FIENBERG, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *J. Comput. Graph. Statist.* **21** 825–839. MR3005799 https://doi.org/10.1080/10618600.2012.738106

FIENBERG, S. E. and SLAVKOVIC, A. (2010). Data privacy and confidentiality. In *International Encyclopedia of Statistical Science* 342–345. Springer, Berlin.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Ser. A* **222** 309–368.

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society A* **144** 285–307.

FOSDICK, B. K. and HOFF, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *J. Amer. Statist. Assoc.* **110** 1047–1056. MR3420683 https://doi.org/10.1080/01621459.2015.1008697

FOSDICK, B. K., MCCORMICK, T. H., MURPHY, T. B., NG, T. L. J. and WESTLING, T. (2019). Multiresolution network models. *J. Comput. Graph. Statist.* **28** 185–196. MR3939381 https://doi.org/10.1080/10618600.2018.1505633

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

FRIEZE, A. and KAROŃSKI, M. (2016). *Introduction to Random Graphs*. Cambridge Univ. Press, Cambridge. MR3675279 https://doi.org/10.1017/CBO9781316339831

GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

GEYER, C. J. and THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. Roy. Statist. Soc. Ser. B* **54** 657–699. MR1185217

GILBERT, E. N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098

GILE, K. J. (2011). Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *J. Amer. Statist. Assoc.* **106** 135–146. MR2816708 https://doi.org/10.1198/jasa.2011.ap09475

GILE, K. and HANDCOCK, M. S. (2006). Model-based assessment of the impact of missing data on inference for networks Technical Report Center for Statistics and the Social Sciences, Univ. Washington, Seattle. Available at https://www.csss.washington.edu/Papers/wp66.pdf.

GILE, K. and HANDCOCK, M. H. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociol. Method.* **40** 285–327.

GILE, K. J. and HANDCOCK, M. S. (2017). Analysis of networks with missing data with application to the National Longitudinal Study of Adolescent Health. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 501–519. MR3632339 https://doi.org/10.1111/rssc.12184

GILL, P. S. and SWARTZ, T. B. (2004). Bayesian analysis of directed graphs data with applications to social networks. *J. Roy. Statist. Soc. Ser. C* **53** 249–260. MR2055619 https://doi.org/10.1046/j.1467-9876.2003.05215.x

GJOKA, M., SMITH, E. J. and BUTTS, C. T. (2014). Estimating clique composition and size distributions from sampled network data. *Proceedings of the Sixth IEEE Workshop on Network Science for Communication Networks* (*NetSciCom* 2014).

GJOKA, M., SMITH, E. and BUTTS, C. T. (2015). Estimating subgraph frequencies with or without attributes from egocentrically sampled data. Available at arxiv.org/abs/1510.08119.

GOLDBERGER, A. L., AMARAL, L. A., GLASS, L., HAUSDORFF, J. M., IVANOV, P. C., MARK, R. G., MIETUS, J. E., MOODY, G. B., PENG, C.-K. et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101** e215–e220.

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2009). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

GONDAL, N. (2018). Duality of departmental specializations and PhD exchange: A Weberian analysis of status in interaction using multilevel exponential random graph models (mERGM). *Soc. Netw.* **55** 202–212.

GOODMAN, L. A. (1961). Snowball sampling. *Ann. Math. Stat.* **32** 148–170. MR0124140 https://doi.org/10.1214/aoms/1177705148

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.

GOODREAU, S. M., HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T. and MORRIS, M. (2008). A statnet tutorial. *J. Stat. Softw.* **24** 1–27.

GRAZIOLI, G., MARTIN, R. W. and BUTTS, C. T. (2019). Comparative exploratory analysis of intrinsically disordered protein dynamics using machine learning and network analytic methods. *Frontiers in Molecular Biosciences*, *Biological Modeling and Simulation* **6**.

GRAZIOLI, G., YU, Y., UNHELKAR, M. H., MARTIN, R. W. and BUTTS, C. T. (2019). Network-based classification and modeling of amyloid fibrils. *J Phys Chem B* **123** 5452–5462. https://doi.org/10.1021/acs.jpcb.9b03494

GROENDYKE, C., WELCH, D. and HUNTER, D. R. (2012). A network-based analysis of the 1861 Hagelloch measles data. *Biometrics* **68** 755–765. MR3055180 https://doi.org/10.1111/j.1541-0420.2012.01748.x

HÄGGSTRÖM, O. and JONASSON, J. (1999). Phase transition in the random triangle model. *J. Appl. Probab.* **36** 1101–1115. MR1742153 https://doi.org/10.1017/s0021900200017897

HANDCOCK, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In *Dynamic Social Network Modeling and Analysis*: *Workshop Summary and Papers* (R. Breiger, K. Carley and P. Pattison, eds.) 1–12. National Academies Press, Washington, DC.

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling social networks from sampled data. *Ann*. *Appl*. *Stat*. **4** 5–25. MR2758082 https://doi.org/10.1214/08-AOAS221

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 https://doi.org/10.1111/j.1467-985X.2007.00471.x

HANNEKE, S., FU, W. and XING, E. P. (2010). Discrete temporal models of social networks. *Electron*. *J*. *Stat*. **4** 585–605. MR2660534 https://doi.org/10.1214/09-EJS548

HARRIS, J. K. (2013). *An Introduction to Exponential Random Graph Modeling*. Sage, Thousand Oaks.

HARTLEY, H. O. and SIELKEN, R. L. JR. (1975). A "super-population viewpoint" for finite population sampling. *Biometrics* **31** 411–422. MR0386084 https://doi.org/10.2307/2529429

HE, R. and ZHENG, T. (2015). GLMLE: Graph-limit enabled fast computation for fitting exponential random graph models to large social networks. *Soc*. *Netw*. *Anal*. *Min*. **5** 1–19.

HECKATHORN, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Soc*. *Probl*. **44** 174–199.

HOFF, P. D. (2003). Random effects models for network data. In *Dynamic Social Network Modeling and Analysis*: *Workshop Summary and Papers* (R. Breiger, K. Carley and P. Pattison, eds.) 303–312. National Academies Press, Washington, DC.

HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J*. *Amer*. *Statist*. *Assoc*. **100** 286–295. MR2156838 https://doi.org/10.1198/016214504000001015

HOFF, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* 20 (J. C. Platt, D. Koller, Y. Singer and S. Roweis, eds.) 657–664. MIT Press, Cambridge, MA.

HOFF, P. D. (2009). A hierarchical eigenmodel for pooled covariance estimation. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 971–992. MR2750253 https://doi.org/10.1111/j.1467-9868.2009.00716.x

HOFF, P. D. (2020). Additive and multiplicative effects network models. *Statist*. *Sci*. To appear.

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 https://doi.org/10.1198/016214502388618906

HOLLAND, P. W. and LEINHARDT, S. (1970). A method for detecting structure in sociometric data. *Amer*. *J*. *Sociol*. **76** 492–513.

HOLLAND, P. W. and LEINHARDT, S. (1972). Some evidence on the transitivity of positive interpersonal sentiment. *Amer*. *J*. *Sociol*. **77** 1205–1209.

HOLLAND, P. W. and LEINHARDT, S. (1976). Local structure in social networks. *Sociol*. *Method*. 1–45.

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

HOLLWAY, J. and KOSKINEN, J. (2016). Multilevel embeddedness: The case of the global fisheries governance complex. *Soc*. *Netw*. **44** 281–294.

HOLLWAY, J., LOMI, A., PALLOTTI, F. and STADTFELD, C. (2017). Multilevel social spaces: The network dynamics of organizational fields. *Netw*. *Sci*. **5** 187–212.

HOMANS, G. C. (1950). *The Human Group*. Harcourt, Brace, New York.

HUITSING, G., VAN DUIJN, M. A. J., SNIJDERS, T. A. B., WANG, P., SAINIO, M., SALMIVALLI, C. and VEENSTRA, R. (2012). Univariate and multivariate models of positive and negative networks: Liking, disliking, and bully–victim relationships. *Soc*. *Netw*. **34** 645–657.

HUMMEL, R. M., HUNTER, D. R. and HANDCOCK, M. S. (2012). Improving simulation-based algorithms for fitting ERGMs. *J*. *Comput*. *Graph*. *Statist*. **21** 920–939. MR3005804 https://doi.org/10.1080/10618600.2012.679224

HUNTER, D. R. (2007). Curved exponential family models for social networks. *Soc*. *Netw*. **29** 216–230. https://doi.org/10.1016/j.socnet.2006.08.005

HUNTER, D. R., GOODREAU, S. M. and HANDCOCK, M. S. (2008). Goodness of fit of social network models. *J*. *Amer*. *Statist*. *Assoc*. **103** 248–258. MR2394635 https://doi.org/10.1198/016214507000000446

HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J*. *Comput*. *Graph*. *Statist*. **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069

HUNTER, D. R., KRIVITSKY, P. N. and SCHWEINBERGER, M. (2012). Computational statistical methods for social network models. *J*. *Comput*. *Graph*. *Statist*. **21** 856–882. MR3005801 https://doi.org/10.1080/10618600.2012.732921

HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J*. *Stat*. *Softw*. **24** 1–29.

ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Z*. *Phys*. **31** 253–258.

JANSON, S. (2018). On edge exchangeable random graphs. *J*. *Stat*. *Phys*. **173** 448–484. MR3876897 https://doi.org/10.1007/s10955-017-1832-9

JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

JIN, J. (2015). Fast community detection by SCORE. *Ann*. *Statist*. **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

JIN, I. H. and LIANG, F. (2013). Fitting social network models using varying truncation stochastic approximation MCMC algorithm. *J*. *Comput*. *Graph*. *Statist*. **22** 927–952. MR3173750 https://doi.org/10.1080/10618600.2012.680851

JIN, I. H., YUAN, Y. and LIANG, F. (2013). Bayesian analysis for exponential random graph models using the adaptive exchange sampler. *Stat*. *Interface* **6** 559–576. MR3164659 https://doi.org/10.4310/SII.2013.v6.n4.a13

JONASSON, J. (1999). The random triangle model. *J*. *Appl*. *Probab*. **36** 852–867. MR1737058 https://doi.org/10.1239/jap/1032374639

KARWA, V., KRIVITSKY, P. N. and SLAVKOVIĆ, A. B. (2017). Sharing social network data: Differentially private estimation of exponential family random-graph models. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **66** 481–500. MR3632338 https://doi.org/10.1111/rssc.12185

KARWA, V., PETROVIĆ, S. and BAJIĆ, D. (2016). DERGMs: Degeneracy-restricted exponential random graph models. Preprint. Available at arXiv:1612.03054.

KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann*. *Statist*. **44** 87–112. MR3449763 https://doi.org/10.1214/15-AOS1358

KENYON, R. and YIN, M. (2017). On the asymptotics of constrained exponential random graphs. *J*. *Appl*. *Probab*. **54** 165–180. MR3632612 https://doi.org/10.1017/jpr.2016.93

KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. *Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1

KOSKINEN, J. (2004). Essays on Bayesian inference for social networks. Ph.D. thesis Stockholm Univ., Dept. of Statistics, Sweden.

KOSKINEN, J. H. (2009). Using latent variables to account for heterogeneity in exponential family random graph models. In *Proceedings of the* 6*th St*. *Petersburg Workshop on Simulation* (S. M. Ermakov, V. B. Melas and A. N. Pepelyshev, eds.) **2** 845–849. St. Petersburg State Univ., St. Petersburg, Russia.

KOSKINEN, J. H., ROBINS, G. L. and PATTISON, P. E. (2010). Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Stat*. *Methodol*. **7** 366–384. MR2643608 https://doi.org/10.1016/j.stamet.2009.09.007

KRACKHARDT, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Soc*. *Netw*. **10** 359–381. MR0984597 https://doi.org/10.1016/0378-8733(88)90004-4

KRIVITSKY, P. N. (2012). Exponential-family random graph models for valued networks. *Electron*. *J*. *Stat*. **6** 1100–1128. MR2988440 https://doi.org/10.1214/12-EJS696

KRIVITSKY, P. N. (2017). Using contrastive divergence to seed Monte Carlo MLE for exponential-family random graph models. *Comput*. *Statist*. *Data Anal*. **107** 149–161. MR3575065 https://doi.org/10.1016/j.csda.2016.10.015

KRIVITSKY, P. N. and BUTTS, C. T. (2017). Exponential-family random graph models for rank-order relational data. *Sociol*. *Method*. **47** 68–112.

KRIVITSKY, P. N. and HANDCOCK, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *J*. *Stat*. *Softw*. **24**. https://doi.org/10.18637/jss.v024.i05

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **76** 29–46. MR3153932 https://doi.org/10.1111/rssb.12014

KRIVITSKY, P. N., HANDCOCK, M. S. and MORRIS, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Stat*. *Methodol*. **8** 319–339. MR2800354 https://doi.org/10.1016/j.stamet.2011.01.005

KRIVITSKY, P. N. and KOLACZYK, E. D. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statist*. *Sci*. **30** 184–198. MR3353102 https://doi.org/10.1214/14-STS502

KRIVITSKY, P. N., MARCUM, C. S. and KOEHLY, L. (2019). Exponential-family random graph models for multi-layer networks.

KRIVITSKY, P. N. and MORRIS, M. (2017). Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann*. *Appl*. *Stat*. **11** 427–455. MR3634330 https://doi.org/10.1214/16-AOAS1010

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. https://doi.org/10.1016/j.socnet.2009.04.001

KURANT, M., MARKOPOULOU, A. and THIRAN, P. (2011). Towards unbiased BFS sampling. *IEEE J*. *Sel*. *Areas Commun*. **29** 1799–1809.

KURANT, M., GJOKA, M., WANG, Y., ALMQUIST, Z. W., BUTTS, C. T. and MARKOPOULOU, A. (2012). Coarse-grained topology estimation via graph sampling. In *Proceedings of ACM SIGCOMM Workshop on Online Social Networks* (*WOSN*) '12.

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford University Press, New York. MR1419991

LAURITZEN, S. L. (2008). Exchangeable Rasch matrices. *Rend*. *Mat*. *Appl*. (7) **28** 83–95. MR2463441

LAURITZEN, S., RINALDO, A. and SADEGHI, K. (2018). Random networks, graphical models and exchangeability. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **80** 481–508. MR3798875 https://doi.org/10.1111/rssb.12266

LAZEGA, E. and PATTISON, P. E. (1999). Multiplexity, generalized exchange and cooperation in organizations: A case study. *Soc*. *Netw*. **21** 67–90.

LAZEGA, E. and SNIJDERS, T. A. B., eds. (2016). *Multilevel Network Analysis for the Social Sciences*. Springer, Cham.

LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. *Springer Texts in Statistics*. Springer, New York. MR1663158 https://doi.org/10.1007/b98855

LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann*. *Statist*. **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

LEIFELD, P., CRANMER, S. J. and DESMARAIS, B. A. (2018). Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. *J*. *Stat*. *Softw*. **83** 1–36.

LIANG, F. and JIN, I.-H. (2013). A Monte Carlo Metropolis–Hastings algorithm for sampling from distributions with intractable normalizing constants. *Neural Comput*. **25** 2199–2234. MR3100001 https://doi.org/10.1162/NECO_a_00466

LIANG, F., JIN, I. H., SONG, Q. and LIU, J. S. (2016). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *J*. *Amer*. *Statist*. *Assoc*. **111** 377–393. MR3494666 https://doi.org/10.1080/01621459.2015.1009072

LOMI, A., ROBINS, G. and TRANMER, M. (2016). Introduction to multilevel social networks. *Soc*. *Netw*. **44** 266–268. https://doi.org/10.1016/j.socnet.2015.10.006

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060

LUBBERS, M. J. (2003). Group composition and network structure in school classes: A multilevel application of the p* model. *Soc*. *Netw*. **25** 309–332.

LUBBERS, M. J. and SNIJDERS, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Soc*. *Netw*. **29** 489–507.

LUNAGOMEZ, S. and AIROLDI, E. (2014). Bayesian inference from non-ignorable network sampling designs. Available at arXiv:1401.4718.

LUSHER, D., KOSKINEN, J. and ROBINS, G. (2013). *Exponential Random Graph Models for Social Networks*. Cambridge Univ. Press, Cambridge, UK.

LYNE, A.-M., GIROLAMI, M., ATCHADÉ, Y., STRATHMANN, H. and SIMPSON, D. (2015). On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Statist*. *Sci*. **30** 443–467. MR3432836 https://doi.org/10.1214/15-STS523

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0727836 https://doi.org/10.1007/978-1-4899-3244-0

MCPHERSON, J. M. (1983). An ecology of affiliation. *Am*. *Sociol*. *Rev*. **48** 519–532.

MELE, A. (2017). A structural model of dense network formation. *Econometrica* **85** 825–850. MR3664180 https://doi.org/10.3982/ECTA10400

MEREDITH, C., VAN DEN NOORTGATE, W., STRUYVE, C., GIELEN, S. and KYNDT, E. (2017). Information seeking in secondary schools: A multilevel network approach. *Soc*. *Netw*. **50** 35–45.

MØLLER, J., PETTITT, A. N., REEVES, R. and BERTHELSEN, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93** 451–458. MR2278096 https://doi.org/10.1093/biomet/93.2.451

MORRIS, M., HANDCOCK, M. S. and HUNTER, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *J. Stat. Softw.* **24** 1–24.

MUKHERJEE, S. (2013). Phase transition in the two star exponential random graph model. Available at arXiv:1310.4164.

MUKHERJEE, S. (2020). Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics. *Bernoulli*. To appear.

MUKHERJEE, R., MUKHERJEE, S. and SEN, S. (2018). Detection thresholds for the β-model on sparse graphs. *Ann. Statist.* **46** 1288–1317. MR3798004 https://doi.org/10.1214/17-AOS1585

MURRAY, I., GHAHRAMANI, Z. and MACKAY, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence* 359–366. AUAI Press, Corvallis, OR.

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

OBANDO, C. and DE VICO FALLANI, F. (2017). A statistical model for brain networks inferred from large-scale electrophysiological signals. *J. R. Soc. Interface* 1–10.

OKABAYASHI, S. and GEYER, C. J. (2012). Long range search for maximum likelihood in exponential families. *Electron. J. Stat.* **6** 123–147. MR2879674 https://doi.org/10.1214/11-EJS664

ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

OUZIENKO, V., GUO, Y. and OBRADOVIC, Z. (2011). A decoupled exponential random graph model for prediction of structure and attributes in temporal social networks. *Stat. Anal. Data Min.* **4** 470–486. MR2842404 https://doi.org/10.1002/sam.10130

PARK, J. and HARAN, M. (2018). Bayesian inference in the presence of intractable normalizing functions. *J. Amer. Statist. Assoc.* **113** 1372–1390. MR3862364 https://doi.org/10.1080/01621459.2018.1448824

PARK, J. and NEWMAN, M. E. J. (2004). Solution of the two-star model of a network. *Phys. Rev. E* (3) **70** 066146, 5. MR2133810 https://doi.org/10.1103/PhysRevE.70.066146

PARK, J. and NEWMAN, M. E. J. (2005). Solution for the properties of a clustered network. *Phys. Rev. E* **72** 026136.

PATTISON, P. and ROBINS, G. (2002). Neighborhood-based models for social networks. In *Sociological Methodology* (R. M. Stolzenberg, ed.) **32** 301–337. Blackwell Publishing, Boston, MA.

PATTISON, P. and WASSERMAN, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *Br. J. Math. Stat. Psychol.* **52** 169–193.

PATTISON, P. E., ROBINS, G. L., SNIJDERS, T. A. B. and WANG, P. (2013). Conditional estimation of exponential random graph models from snowball sampling designs. *J. Math. Psych.* **57** 284–296. MR3137882 https://doi.org/10.1016/j.jmp.2013.05.004

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

RADIN, C. and YIN, M. (2013). Phase transitions in exponential random graphs. *Ann. Appl. Probab.* **23** 2458–2471. MR3127941 https://doi.org/10.1214/12-AAP907

RAFTERY, A. E., NIU, X., HOFF, P. D. and YEUNG, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *J. Comput. Graph. Statist.* **21** 901–919. MR3005803 https://doi.org/10.1080/10618600.2012.679240

RAPOPORT, A. (1979/80). A probabilistic approach to networks. *Soc. Netw.* **2** 1–18. MR0551137 https://doi.org/10.1016/0378-8733(79)90008-X

RASTELLI, R., FRIEL, N. and RAFTERY, A. E. (2016). Properties of latent variable network models. *Netw. Sci.* **4** 407–432.

RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691

RICHARDSON, M. and DOMINGOS, P. (2006). Markov logic networks. *Mach. Learn.* **62** 107–136.

RINALDO, A., FIENBERG, S. E. and ZHOU, Y. (2009). On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Stat.* **3** 446–484. MR2507456 https://doi.org/10.1214/08-EJS350

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078

ROBINS, G. and PATTISON, P. (2001). Random graph models for temporal processes in social networks. *J. Math. Sociol.* **25** 5–41.

ROBINS, G. L., PATTISON, P. E. and WANG, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. *Soc. Netw.* **31** 105–117.

ROBINS, G., PATTISON, P. and WASSERMAN, S. (1999). Logit models and logistic regressions for social networks. III. Valued relations. *Psychometrika* **64** 371–394. MR1720089 https://doi.org/10.1007/BF02294302

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

ROLLS, D. A., WANG, P., JENKINSON, R., PATTISON, P. E., ROBINS, G. L., SACKS-DAVIS, R., DARAGANOVA, G., HELLARD, M. and MCBRYDE, E. (2013). Modelling a disease-relevant contact network of people who inject drugs. *Soc. Netw.* **35** 699–710.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

SALGANIK, M. J. and HECKATHORN, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol. Method.* **34** 193–239.

SALTER-TOWNSHEND, M. and MURPHY, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Comput. Statist. Data Anal.* **57** 661–671. MR2981116 https://doi.org/10.1016/j.csda.2012.08.004

SALTER-TOWNSHEND, M. and MURPHY, T. B. (2015). Role analysis in networks using mixtures of exponential random graph models. *J. Comput. Graph. Statist.* **24** 520–538. MR3357393 https://doi.org/10.1080/10618600.2014.923777

SALTER-TOWNSHEND, M., WHITE, A., GOLLINI, I. and MURPHY, T. B. (2012). Review of statistical network analysis: Models, algorithms, and software. *Stat. Anal. Data Min.* **5** 260–264. MR2958152 https://doi.org/10.1002/sam.11146

SCHALK, G., MCFARLAND, D. J., HINTERBERGER, T., BIRBAUMER, N. and WOLPAW, J. R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **51** 1034–1043.

SCHWEINBERGER, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841 https://doi.org/10.1198/jasa.2011.tm10747

SCHWEINBERGER, M. (2020). Consistent structure estimation of exponential-family random graph models with block structure. *Bernoulli*. To appear.

SCHWEINBERGER, M. and HANDCOCK, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 https://doi.org/10.1111/rssb.12081

SCHWEINBERGER, M. and LUNA, P. (2018). HERGM: Hierarchical exponential-family random graph models. *J. Stat. Softw.* **85** 1–39.

SCHWEINBERGER, M. and SNIJDERS, T. A. B. (2003). Settings in social networks: A measurement model. In *Sociological Methodology* (R. M. Stolzenberg, ed.) **33** 307–341 10. Basil Blackwell, Boston & Oxford.

SCHWEINBERGER, M. and STEWART, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* To appear.

SENGUPTA, S. and CHEN, Y. (2018). A block model for node popularity in networks with community structure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 365–386. MR3763696 https://doi.org/10.1111/rssb.12245

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061 https://doi.org/10.1080/01621459.2014.988214

SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112 https://doi.org/10.1214/12-AOS1044

SIMPSON, S. L., BOWMAN, F. D. and LAURIENTI, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Stat. Surv.* **7** 1–36. MR3161730 https://doi.org/10.1214/13-SS103

SIMPSON, S. L., HAYASAKA, S. and LAURIENTI, P. J. (2011). Exponential random graph modeling for complex brain networks. *PLoS ONE* **6** e20039. https://doi.org/10.1371/journal.pone.0020039

SIMPSON, S. L., MOUSSA, M. N. and LAURIENTI, P. J. (2012). An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *NeuroImage* **60** 1117–1126.

SINKE, M. R. T., DIJKHUIZEN, R. M., CAIMO, A., STAM, C. J. and OTTE, W. M. (2016). Bayesian exponential random graph modeling of whole-brain structural networks across lifespan. *NeuroImage* **135** 79–91.

SLAUGHTER, A. J. and KOEHLY, L. M. (2016). Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Soc. Netw.* **44** 334–345.

SMITH, T. W., MARSDEN, P., HOUT, M. and KIM, J. (1972–2016). General Social Surveys Technical Report NORC at the Univ. Chicago.

SNIJDERS, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *J. Soc. Struct.* **3** 1–40.

SNIJDERS, T. A. B. (2010). Conditional marginalization for exponential random graph models. *J. Math. Sociol.* **34** 239–252.

SNIJDERS, T. A. B. and BOSKER, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd ed. Sage Publications, Los Angeles, CA. MR3137621

SNIJDERS, T. A. B. and VAN DUIJN, M. A. J. (2002). Conditional maximum likelihood estimation under various specifications of exponential random graph models. In *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in Honour of Ove Frank* (J. Hagberg, ed.) 117–134. Dept. Statistics, Univ. Stockholm.

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

STEIN, M. L. (1999). *Interpolation of Spatial Data. Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

STEWART, J., SCHWEINBERGER, M., BOJANOWSKI, M. and MORRIS, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Soc. Netw.* **59** 98–119.

STRAUSS, D. (1986). On a general class of models for interaction. *SIAM Rev.* **28** 513–527. MR0867682 https://doi.org/10.1137/1028156

STRAUSS, D. and IKEDA, M. (1990). Pseudolikelihood estimation for social networks. *J. Amer. Statist. Assoc.* **85** 204–212. MR1137368

SUESSE, T. (2012). Marginalized exponential random graph models. *J. Comput. Graph. Statist.* **21** 883–900. MR3005802 https://doi.org/10.1080/10618600.2012.694750

TALAGRAND, M. (1996). A new look at independence. *Ann. Probab.* **24** 1–34. MR1387624 https://doi.org/10.1214/aop/1042644705

TANG, M., SUSSMAN, D. L. and PRIEBE, C. E. (2013). Universally consistent vertex classification for latent positions graphs. *Ann. Statist.* **41** 1406–1430. MR3113816 https://doi.org/10.1214/13-AOS1112

THIEMICHEN, S. and KAUERMANN, G. (2017). Stable exponential random graph models with non-parametric components for large dense networks. *Soc. Netw.* **49** 67–80.

THIEMICHEN, S., FRIEL, N., CAIMO, A. and KAUERMANN, G. (2016). Bayesian exponential random graph models with nodal random effects. *Soc. Netw.* **46** 11–28.

THOMPSON, S. K. (2012). *Sampling*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2894042 https://doi.org/10.1002/9781118162934

THOMPSON, S. and FRANK, O. (2000). Model-based estimation with link-tracing sampling designs. *Surv. Methodol.* **26** 87–98.

VAN DUIJN, M. A. J. (1995). Estimation of a random effects model for directed graphs. In *Toeval Zit Overal: Programmatuur voor Random-Coëffciënt Modellen* (T. A. B. Snijders, B. Engel, J. C. Van Houwelingen, A. Keen, G. J. Stemerdink and M. Verbeek, eds.) 113–131. IEC ProGAMMA, Groningen.

VAN DUIJN, M. A. J., GILE, K. and HANDCOCK, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Soc. Netw.* **31** 52–62.

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 https://doi.org/10.1046/j.0039-0402.2003.00258.x

VEITCH, V. and ROY, D. M. (2015). The class of random graphs arising from exchangeable random measures. Preprint. Available at arXiv:1512.03099.

VEITCH, V. and ROY, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *Ann. Statist.* **47** 3274–3299. MR4025742 https://doi.org/10.1214/18-AOS1778

WANG, J. and ATCHADÉ, Y. F. (2014). Approximate Bayesian computation for exponential random graph models for large social networks. *Comm. Statist. Simulation Comput.* **43** 359–377. MR3200975 https://doi.org/10.1080/03610918.2012.703359

WANG, P., ROBINS, G. and PATTISON, P. (2006). PNet. program for the simulation and estimation of exponential random graph (p*) models. Melbourne School of Psychological Sciences, University of Melbourne.

WANG, P., ROBINS, G., PATTISON, P. and LAZEGA, E. (2013). Exponential random graph models for multilevel networks. *Soc. Netw.* **35** 96–115.

WANG, P., ROBINS, G., PATTISON, P. and LAZEGA, E. (2016a). Social selection models for multilevel networks. *Soc. Netw.* **44** 346–362.

WANG, C., BUTTS, C. T., HIPP, J. R., JOSE, R. and LAKON, C. M. (2016b). Multiple imputation for missing edge data: A predictive evaluation method with application to add health. *Soc. Netw.* **45** 89–98. https://doi.org/10.1016/j.socnet.2015.12.003

WANG, Y., FANG, H., YANG, D., ZHAO, H. and DENG, M. (2018). Network clustering analysis using mixture exponential-family random graph models and its application in genetic interaction data.

IEEE/ACM Trans. Comput. Biol. Bioinform. https://doi.org/10.1109/TCBB.2017.2743711.

WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, Cambridge.

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 https://doi.org/10.1007/BF02294547

WILLINGER, W., ALDERSON, D. and DOYLE, J. C. (2009). Mathematics and the Internet: A source of enormous confusion and great potential. *Notices Amer. Math. Soc.* **56** 586–599. MR2509062

WYATT, D., CHOUDHURY, T. and BILMES, J. (2008). Learning hidden curved exponential random graph models to infer face-to-face interaction networks from situated speech data. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence* 732–738.

XIANG, R. and NEVILLE, J. (2011). Relational learning with one network: An asymptotic analysis. In *Proceedings of the* 14*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 1–10.

YAN, T., LENG, C. and ZHU, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.* **44** 31–57. MR3449761 https://doi.org/10.1214/15-AOS1343

YAN, T., QIN, H. and WANG, H. (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statist. Sinica* **26** 273–293. MR3468353

YAN, T., ZHAO, Y. and QIN, H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *J. Multivariate Anal.* **133** 61–76. MR3282018 https://doi.org/10.1016/j.jmva.2014.08.013

YAN, T., JIANG, B., FIENBERG, S. E. and LENG, C. (2019). Statistical inference in a directed network model with covariates. *J. Amer. Statist. Assoc.* **114** 857–868. MR3963186 https://doi.org/10.1080/01621459.2018.1448829

YANG, X., RINALDO, A. and FIENBERG, S. E. (2014). Estimation for dyadic-dependent exponential random graph models. *J. Algebr. Stat.* **5** 39–63. MR3279953 https://doi.org/10.18409/jas.v5i1.24

YANG, E., RAVIKUMAR, P., ALLEN, G. I. and LIU, Z. (2015). Graphical models via univariate exponential family distributions. *J. Mach. Learn. Res.* **16** 3813–3847. MR3450553

YIN, M., RINALDO, A. and FADNAVIS, S. (2016). Asymptotic quantization of exponential random graphs. *Ann. Appl. Probab.* **26** 3251–3285. MR3582803 https://doi.org/10.1214/16-AAP1175

ZAPPA, P. and LOMI, A. (2015). The analysis of multilevel networks in organizations: Models and empirical tests. *Organ. Res. Methods* **18** 542–569.

ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. MR3546450 https://doi.org/10.1214/15-AOS1428

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083 https://doi.org/10.1214/12-AOS1036

Statistical Science
2020, Vol. 35, No. 4, 663–671
https://doi.org/10.1214/19-STS766

A Conversation with J. Stuart (Stu) Hunter

Richard D. De Veaux

Abstract. J. Stuart (Stu) Hunter has been an inspiration and mentor to a generation of statisticians, especially to those working in industry. Born on June 3, 1923 in Holyoke, Massachusetts, Stu moved to Linden, New Jersey at the age of 2 where he spent the rest of his childhood, graduating from high school at age 16. After receiving a bachelor's degree in Electrical Engineering in 1947, he went on to receive a master's degree in applied mathematics in 1949 and a PhD in statistics in 1954, all from North Carolina State. His research centered on experimental design, in particular the study of fractional factorial designs and response surface methods. He was the founding editor of *Technometrics*. Stu joined the faculty at Princeton University as an assistant professor in the Engineering School in 1961. He was a first-rate teacher, and his courses at Princeton were often rated among the top courses at the University. The interviewer had the good fortune to take his Engineering Statistics course in 1970 which began a life-long friendship. Stu was a consultant for many companies and the co-author of the influential book *Statistics for Experimenters* with George Box and William Hunter. His short courses in industry were legendary. He served as the 1993 president of the American Statistical Association (ASA) and has received many honors and awards from the ASA, the ASQ and other organizations. In 2005, he was named as a fellow to the National Academy of Engineering. The Stu Hunter Research Conference was established in 2012 to "honor one of the pioneers in applied statistics." Stu retired from Princeton in 1984, but remains active consulting, mentoring and traveling to this day.

Key words and phrases: Experimental design, fractional factorial design, applied statistics, engineering statistics.

REFERENCES

[1] BOX, G. E. P. and HUNTER, J. S. (1954). A confidence region for the solution of a set of simultaneous equations with an application to experimental design. *Biometrika* **41** 190–199. MR0062388 https://doi.org/10.1093/biomet/41.1-2.190

[2] BOX, G. E. P. and HUNTER, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Stat*. **28** 195–241. MR0085679 https://doi.org/10.1214/aoms/1177707047

[3] BOX, G. E. P. and HUNTER, J. S. (1961). The 2^{k-p} fractional factorial designs. Part I. *Technometrics* **3** 311–351. MR0131937 https://doi.org/10.2307/1266725

[4] BOX, G. E. P. and HUNTER, J. S. (1961). The 2^{k-p} fractional factorial designs. Part II. *Technometrics* **3** 449–458. MR0131938 https://doi.org/10.1080/00401706.1961.10489967

[5] BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters*: *Design*, *Innovation*, *and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2140250

[6] BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters*: *An Introduction to Design*, *Data Analysis*, *and Model Building*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0483116

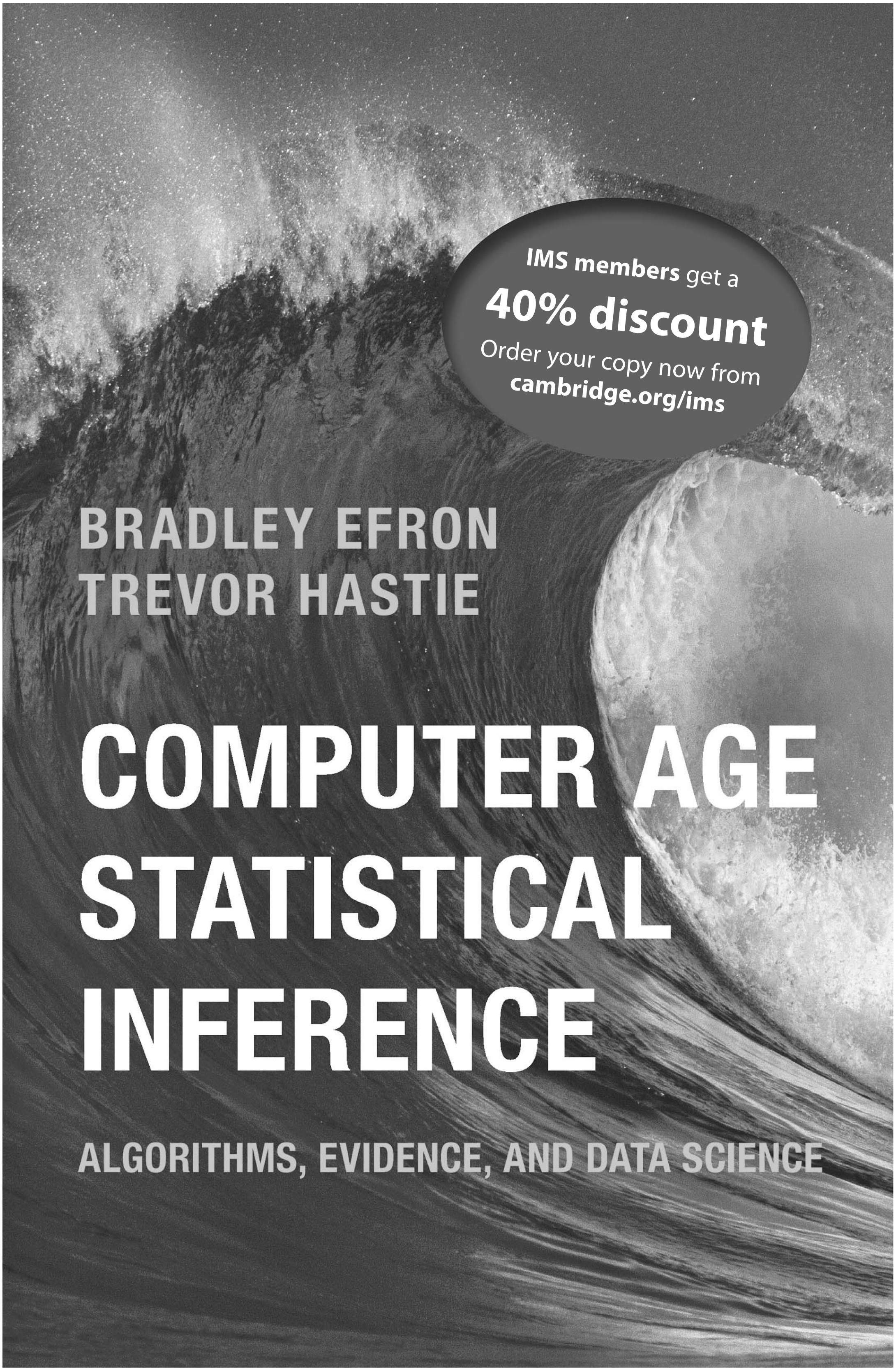
[7] YOUDEN, W. J. and HUNTER, J. S. (1955). Partially replicated Latin squares. *Biometrics* **11** 399–405.

Richard D. De Veaux is the chair of the Department of Mathematics and Statistics and C. Carlilse and M. Tippit Professor of Statistics, Williams College, Williamstown, Massachusetts 01267, USA (e-mail: rdeveaux@williams.edu).

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