

# THE ANNALS *of* STATISTICS

*AN OFFICIAL JOURNAL OF THE*  
INSTITUTE OF MATHEMATICAL STATISTICS

|                                                                                                                                                                            |                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| A likelihood ratio framework for high-dimensional semiparametric regression<br>YANG NING, TIANQI ZHAO AND HAN LIU                                                          | 2299                |
| A new perspective on boosting in linear regression via subgradient optimization and<br>relatives ..... ROBERT M. FREUND, PAUL GRIGAS AND RAHUL MAZUMDER                    | 2328                |
| On the validity of resampling methods under long memory<br>SHUYANG BAI AND MURAD S. TAQQU                                                                                  | 2365                |
| CoCoLasso for high-dimensional error-in-variables regression<br>ABHIRUP DATTA AND HUI ZOU                                                                                  | 2400                |
| Consistent parameter estimation for LASSO and approximate message passing<br>ALI MOUSAVI, ARIAN MALEKI AND RICHARD G. BARANIUK                                             | 2427                |
| Support recovery without incoherence: A case for nonconvex regularization<br>PO-LING LOH AND MARTIN J. WAINWRIGHT                                                          | 2455                |
| Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays<br>YUAN-LUNG LIN, FREDERICK KIN HING PHOA AND MING-HUNG KAO                                  | 2483                |
| Adaptive Bernstein–von Mises theorems in Gaussian white noise ..... KOLYAN RAY                                                                                             | 2511                |
| Targeted sequential design for targeted learning inference of the optimal treatment rule<br>and its mean reward<br>ANTOINE CHAMBAZ, WENJING ZHENG AND MARK J. VAN DER LAAN | 2537                |
| Nonparametric goodness-of-fit tests for uniform stochastic ordering<br>CHUAN-FA TANG, DEWEI WANG AND JOSHUA M. TEBBS                                                       | 2565                |
| Selecting the number of principal components: Estimation of the true rank of a noisy<br>matrix ..... YUNJIN CHOI, JONATHAN TAYLOR AND ROBERT TIBSHIRANI                    | 2590                |
| Extended conditional independence and applications in causal inference<br>PANAYIOTA CONSTANTINOOU AND A. PHILIP DAWID                                                      | 2618                |
| A weight-relaxed model averaging approach for high-dimensional generalized linear<br>models ..... TOMOHIRO ANDO AND KER-CHAU LI                                            | 2654                |
| Structural similarity and difference testing on multiple sparse Gaussian graphical models<br>WEIDONG LIU                                                                   | 2680                |
| Estimating a probability mass function with unknown labels<br>DRAGI ANEVSKI, RICHARD D. GILL AND STEFAN ZOHREN                                                             | 2708                |
| Optimal sequential detection in multi-stream data .....                                                                                                                    | HOCK PENG CHAN 2736 |

THE ANNALS OF STATISTICS

Vol. 45, No. 6, pp. 2299–2763 December 2017

# 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 OFFICERS

**President:** Jon Wellner, Department of Statistics, University of Washington, Seattle, Washington 98195-4322, USA

**President-Elect:** Alison Etheridge, Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom

**Past President:** Richard Davis, Department of Statistics, Columbia University, New York, New York 10027, USA

**Executive Secretary:** Aurore Delaigle, Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia

**Treasurer:** Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706, USA

**Program Secretary:** Judith Rousseau, Université Paris Dauphine, Place du Maréchal DeLattre de Tassigny, 75016 Paris, France

## IMS EDITORS

**The Annals of Statistics.** *Editors:* Edward I. George, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104, USA; Tailen Hsing, Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1107 USA

**The Annals of Applied Statistics.** *Editor-in-Chief:* Tilmann Gneiting, Heidelberg Institute for Theoretical Studies, HITS gGmbH, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

**The Annals of Probability.** *Editor:* Maria Eulália Vares, Instituto de Matemática, Universidade Federal do Rio de Janeiro, 21941-909 Rio de Janeiro, RJ, Brazil

**The Annals of Applied Probability.** *Editor:* Bálint Tóth, School of Mathematics, University of Bristol, University Walk, BS8 1TW, Bristol, UK and Alfréd Rényi, Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

**Statistical Science.** *Editor:* Cun-Hui Zhang, Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA

**The IMS Bulletin.** *Editor:* Anirban DasGupta, Department of Statistics, Purdue University, West Lafayette, Indiana 47907-2068, USA

*The Annals of Statistics* [ISSN 0090-5364 (print); ISSN 2168-8966 (online)], Volume 45, Number 6, December 2017. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

**POSTMASTER:** Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

## A LIKELIHOOD RATIO FRAMEWORK FOR HIGH-DIMENSIONAL SEMIPARAMETRIC REGRESSION

BY YANG NING\*, TIANQI ZHAO<sup>†</sup> AND HAN LIU<sup>†</sup>

*Cornell University\* and Princeton University<sup>†</sup>*

We propose a new inferential framework for high-dimensional semiparametric generalized linear models. This framework addresses a variety of challenging problems in high-dimensional data analysis, including incomplete data, selection bias and heterogeneity. Our work has three main contributions: (i) We develop a regularized statistical chromatography approach to infer the parameter of interest under the proposed semiparametric generalized linear model without the need of estimating the unknown base measure function. (ii) We propose a new likelihood ratio based framework to construct post-regularization confidence regions and tests for the low dimensional components of high-dimensional parameters. Unlike existing post-regularization inferential methods, our approach is based on a novel directional likelihood. (iii) We develop new concentration inequalities and normal approximation results for U-statistics with unbounded kernels, which are of independent interest. We further extend the theoretical results to the problems of missing data and multiple datasets inference. Extensive simulation studies and real data analysis are provided to illustrate the proposed approach.

## REFERENCES

- [1] ARGYRIOU, A., EVGENIOU, T. and PONTIL, M. (2008). Convex multi-task feature learning. *Mach. Learn.* **73** 243–272.
- [2] BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429. [MR3001131](#)
- [3] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. [MR3335097](#)
- [4] BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2016). Post-selection inference for generalized linear models with many controls. *J. Bus. Econom. Statist.* **34** 606–619. [MR3547999](#)
- [5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- [6] BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007). Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.* **1** 169–194. [MR2312149](#)
- [7] CHAN, K. C. G. (2013). Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation. *Biometrika* **100** 269–276. [MR3034342](#)

---

*MSC2010 subject classifications.* Primary 62E20, 62G20; secondary 62G10.

*Key words and phrases.* Post-regularization inference, likelihood ratio test, semiparametric sparsity, confidence interval, nonconvex penalty, U-statistics.

- [8] CHEN, Y., NING, Y., HONG, C. and WANG, S. (2014). Semiparametric tests for identifying differentially methylated loci with case-control designs using Illumina arrays. *Genet. Epidemiol.* **38** 42–50.
- [9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#)
- [10] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling: From Dependence to Independence*. Springer, New York. [MR1666908](#)
- [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](#)
- [12] FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. [MR3210988](#)
- [13] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#)
- [14] HOFF, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.* **1** 265–283. [MR2393851](#)
- [15] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. [MR3277152](#)
- [16] KALBFLEISCH, J. D. (1978). Likelihood methods and nonparametric tests. *J. Amer. Statist. Assoc.* **73** 167–170. [MR0518600](#)
- [17] LAWLESS, J. F., KALBFLEISCH, J. D. and WILD, C. J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 413–438. [MR1680310](#)
- [18] LIANG, K.-Y. and QIN, J. (2000). Regression analysis under non-standard situations: A pair-wise pseudolikelihood approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 773–786. [MR1796291](#)
- [19] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- [20] LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. [MR3015038](#)
- [21] LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized  $M$ -estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616. [MR3335800](#)
- [22] LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. [MR2893865](#)
- [23] LUO, X. and TSAI, W. Y. (2012). A proportional likelihood ratio model. *Biometrika* **99** 211–222. [MR2899674](#)
- [24] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- [25] MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. [MR2758523](#)
- [26] MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009).  $p$ -values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. [MR2750584](#)
- [27] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. [MR2488351](#)
- [28] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. [MR3161450](#)
- [29] NING, Y., ZHAO, T. and LIU, H. (2017). Supplement to “A likelihood ratio framework for high dimensional semiparametric regression.” DOI:[10.1214/16-AOS1483SUPP](#).

- [30] PETTITT, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **44** 234–243. [MR0676214](#)
- [31] SEVERINI, T. A. and WONG, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20** 1768–1802. [MR1193312](#)
- [32] SHAH, R. D. and SAMWORTH, R. J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. [MR3008271](#)
- [33] SRIVASTAVA, V. K. and GILES, D. E. A. (1987). *Seemingly Unrelated Regression Equations Models: Estimation and Inference. Statistics: Textbooks and Monographs* **80**. Dekker, New York. [MR0930104](#)
- [34] STÄDLER, N. and BÜHLMANN, P. (2012). Missing values: Sparse inverse covariance estimation and an extension to sparse regression. *Stat. Comput.* **22** 219–235. [MR2865066](#)
- [35] TAYLOR, J., LOCKHART, R., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). Post-selection adaptive inference for least angle regression and the lasso. arXiv preprint [arXiv:1401.3889](#).
- [36] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. [MR1379242](#)
- [37] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. [MR3224285](#)
- [38] WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. [MR3127873](#)
- [39] WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. [MR2543689](#)
- [40] ZAHN, J. M., POOSALA, S., OWEN, A. B., INGRAM, D. K., LUSTIG, A., CARTER, A., WEERARATNA, A. T., TAUB, D. D., GOROSPE, M., MAZAN-MAMCZARZ, K. et al. (2007). AGEMAP: A gene expression database for aging in mice. *PLoS Genet.* **3** 2326–2337.
- [41] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. [MR3153940](#)
- [42] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)

## A NEW PERSPECTIVE ON BOOSTING IN LINEAR REGRESSION VIA SUBGRADIENT OPTIMIZATION AND RELATIVES

BY ROBERT M. FREUND<sup>1,\*</sup>, PAUL GRIGAS<sup>2,†</sup> AND RAHUL MAZUMDER<sup>3,\*</sup>

*Massachusetts Institute of Technology\** and *University of California, Berkeley*<sup>†</sup>

We analyze boosting algorithms [*Ann. Statist.* **29** (2001) 1189–1232; *Ann. Statist.* **28** (2000) 337–407; *Ann. Statist.* **32** (2004) 407–499] in linear regression from a new perspective: that of modern first-order methods in convex optimization. We show that classic boosting algorithms in linear regression, namely the incremental forward stagewise algorithm ( $\text{FS}_\varepsilon$ ) and least squares boosting [ $\text{LS-BOOST}(\varepsilon)$ ], can be viewed as subgradient descent to minimize the loss function defined as the maximum absolute correlation between the features and residuals. We also propose a minor modification of  $\text{FS}_\varepsilon$  that yields an algorithm for the LASSO, and that may be easily extended to an algorithm that computes the LASSO path for different values of the regularization parameter. Furthermore, we show that these new algorithms for the LASSO may also be interpreted as the same master algorithm (subgradient descent), applied to a regularized version of the maximum absolute correlation loss function. We derive novel, comprehensive computational guarantees for several boosting algorithms in linear regression (including  $\text{LS-BOOST}(\varepsilon)$  and  $\text{FS}_\varepsilon$ ) by using techniques of first-order methods in convex optimization. Our computational guarantees inform us about the statistical properties of boosting algorithms. In particular, they provide, for the first time, a precise theoretical description of the amount of data-fidelity and regularization imparted by running a boosting algorithm with a prespecified learning rate for a fixed but arbitrary number of iterations, for *any* dataset.

## REFERENCES

- [1] BACH, F. (2015). Duality between subgradient and conditional gradient methods. *SIAM J. Optim.* **25** 115–129. [MR3296634](#)
- [2] BERTSEKAS, D. (1999). *Nonlinear Programming*. Athena Scientific, Belmont, MA.
- [3] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575](#)
- [4] BREIMAN, L. (1998). Arcing classifiers. *Ann. Statist.* **26** 801–849. [MR1635406](#)
- [5] BREIMAN, L. (1999). Prediction games and arcing algorithms. *Neural Comput.* **11** 1493–1517.
- [6] BÜHLMANN, P. (2006). Boosting for high-dimensional linear models. *Ann. Statist.* 559–583.
- [7] BÜHLMANN, P. and HOTHORN, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statist. Sci.* **22** 477–505. [MR2420454](#)
- [8] BÜHLMANN, P. and YU, B. (2003). Boosting with the L2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339.
- [9] BÜHLMANN, P. and YU, B. (2006). Sparse boosting. *J. Mach. Learn. Res.* **7** 1001–1024.

---

*MSC2010 subject classifications.* Primary 62J05, 62J07k; secondary 90C25.

*Key words and phrases.* Boosting, forward stagewise regression, linear regression, bias-variance tradeoff, shrinkage, convex optimization, computational guarantees.

- [10] DETTLING, M. and BÜHLMANN, P. (2003). Boosting for tumor classification with gene expression data. *Bioinformatics* **19** 1061–1069.
- [11] DUCHI, J. and SINGER, Y. (2009). Boosting with structural sparsity. In *Proceedings of the 26th Annual International Conference on Machine Learning* 297–304. ACM, New York.
- [12] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression (with discussion). *Ann. Statist.* **32** 407–499. [MR2060166](#)
- [13] FRANK, M. and WOLFE, P. (1956). An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **3** 95–110. [MR0089102](#)
- [14] FREUND, R., GRIGAS, P. and MAZUMDER, R. (2017). Supplement to “A new perspective on boosting in linear regression via subgradient optimization and relatives.” DOI:[10.1214/16-AOS1505SUPP](#).
- [15] FREUND, R. M. and GRIGAS, P. (2014). New analysis and results for the Frank–Wolfe method. *Math. Program.* To appear.
- [16] FREUND, R. M., GRIGAS, P. and MAZUMDER, R. (2013). AdaBoost and forward stage-wise regression are first-order convex optimization methods. Preprint. Available at [arXiv:1307.1192](#).
- [17] FREUND, Y. (1995). Boosting a weak learning algorithm by majority. *Inform. and Comput.* **121** 256–285. [MR1348530](#)
- [18] FREUND, Y. and SCHAPIRE, R. (1996). Experiments with a new boosting algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference* 148–156. Morgan Kaufman, San Francisco.
- [19] FRIEDMAN, J. (2008). Fast sparse regression and classification. Technical Report, Dept. Statistics, Stanford Univ.
- [20] FRIEDMAN, J., HASTIE, T., HOEFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **2** 302–332.
- [21] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting. *Ann. Statist.* **28** 337–407. [MR1790002](#)
- [22] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. [MR1873328](#)
- [23] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. [MR1873328](#)
- [24] FRIEDMAN, J. H. (2012). Fast sparse regression and classification. *Int. J. Forecast.* **28** 722–738.
- [25] FRIEDMAN, J. H. and POPESCU, B. E. (2003). Importance sampled learning ensembles. *J. Mach. Learn. Res.* **9** 305.
- [26] GÄRTNER, B., JAGGI, M. and MARIA, C. (2012). An exponential lower bound on the complexity of regularization paths. *J. Comput. Geom.* **3** 168–195. [MR3030324](#)
- [27] GIESEN, J., JAGGI, M. and LAUE, S. (2012). Optimizing over the Growing Spectrahedron. In *Algorithms—ESA 2012: 20th Annual European Symposium, Ljubljana, Slovenia, September 10–12, 2012. Proceedings* 503–514. Springer, Berlin.
- [28] GIESEN, J., JAGGI, M. and LAUE, S. (2012). Approximating parameterized convex optimization problems. *ACM Trans. Algorithms* **9** Art. 10, 17. [MR3008305](#)
- [29] GIESEN, J., MUELLER, J., LAUE, S. and SWIERCY, S. (2012). Approximating concavely parameterized optimization problems. In *Advances in Neural Information Processing Systems* **25** (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 2105–2113. Curran Associates, Red Hook, NY.
- [30] HASTIE, T., TAYLOR, J., TIBSHIRANI, R. and WALTHER, G. (2007). Forward stagewise regression and the monotone lasso. *Electron. J. Stat.* **1** 1–29.
- [31] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.



- [32] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. Chapman & Hall, London. [MR1082147](#)
- [33] JAGGI, M. (2013). Revisiting Frank–Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* 427–435.
- [34] MALLAT, S. G. and ZHANG, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41** 3397–3415.
- [35] MASON, L., BAXTER, J., BARTLETT, P. and FREAN, M. (2000). Boosting algorithms as gradient descent **12** 512–518.
- [36] MILLER, A. (2002). *Subset Selection in Regression*. CRC Press, Boca Raton, FL.
- [37] NESTEROV, Y. E. (2003). *Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization* **87**. Kluwer Academic, Boston, MA.
- [38] PATI, Y. C., REZAIEFAR, R. and KRISHNAPRASAD, P. S. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993 40–44. IEEE, New York.
- [39] POLYAK, B. (1987). *Introduction to Optimization*. Optimization Software, New York.
- [40] RÄTSCH, G., ONODA, T. and MÜLLER, K.-R. (2001). Soft margins for AdaBoost. *Mach. Learn.* **42** 287–320.
- [41] ROSSET, S., SWIRSZCZ, G., SREBRO, N. and ZHU, J. (2007).  $\ell_1$  regularization in infinite dimensional feature spaces. In *Conference on Learning Theory* 544–558. Springer, Berlin.
- [42] ROSSET, S., ZHU, J. and HASTIE, T. (2003/2004). Boosting as a regularized path to a maximum margin classifier. *J. Mach. Learn. Res.* **5** 941–973. [MR2248005](#)
- [43] SCHAPIRE, R. (1990). The strength of weak learnability. *Mach. Learn.* **5** 197–227.
- [44] SCHAPIRE, R. E. and FREUND, Y. (2012). *Boosting: Foundations and Algorithms. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2920188](#)
- [45] SHOR, N. Z. (1985). *Minimization Methods for Nondifferentiable Functions. Springer Series in Computational Mathematics* **3**. Springer, Berlin. [MR0775136](#)
- [46] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [47] TIBSHIRANI, R. J. (2015). A general framework for fast stagewise algorithms. *J. Mach. Learn. Res.* **16** 2543–2588.
- [48] TUKEY, J. (1977). *Exploratory Data Analysis*. Addison-Wesley, Boston, MA.
- [49] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [50] WEISBERG, S. (1980). *Applied Linear Regression*. Wiley, New York. [MR0591462](#)
- [51] ZHAO, P. and YU, B. (2007). Stagewise lasso. *J. Mach. Learn. Res.* **8** 2701–2726.

# ON THE VALIDITY OF RESAMPLING METHODS UNDER LONG MEMORY<sup>1</sup>

BY SHUYANG BAI AND MURAD S. TAQQU

*University of Georgia and Boston University*

For long-memory time series, inference based on resampling is of crucial importance, since the asymptotic distribution can often be non-Gaussian and is difficult to determine statistically. However, due to the strong dependence, establishing the asymptotic validity of resampling methods is nontrivial. In this paper, we derive an efficient bound for the canonical correlation between two finite blocks of a long-memory time series. We show how this bound can be applied to establish the asymptotic consistency of subsampling procedures for general statistics under long memory. It allows the subsample size  $b$  to be  $o(n)$ , where  $n$  is the sample size, irrespective of the strength of the memory. We are then able to improve many results found in the literature. We also consider applications of subsampling procedures under long memory to the sample covariance, M-estimation and empirical processes.

## REFERENCES

- ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107. [MR0368354](#)
- BAI, S., TAQQU, M. S. and ZHANG, T. (2016). A unified approach to self-normalized block sampling. *Stochastic Process. Appl.* **126** 2465–2493. [MR3505234](#)
- BERAN, J. (1991). M estimators of location for Gaussian and related processes with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86** 704–708. [MR1147095](#)
- BERAN, J., FENG, Y., GHOSH, S. and KULIK, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer, Heidelberg. [MR3075595](#)
- BETKEN, A. and WENDLER, M. (2015). Subsampling for general statistics under long range dependence. Preprint. Available at [arXiv:1509.05720](#).
- BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions*. Kendrick Press, Utah.
- BREUER, P. and MAJOR, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441. [MR0716933](#)
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR1093459](#)
- CHAMBERS, D. and SLUD, E. (1989). Central limit theorems for nonlinear functionals of stationary Gaussian processes. *Probab. Theory Related Fields* **80** 323–346. [MR0976529](#)
- CONTI, P. L., DE GIOVANNI, L., STOEV, S. A. and TAQQU, M. S. (2008). Confidence intervals for the long memory parameter based on wavelets and resampling. *Statist. Sinica* **18** 559–579. [MR2411615](#)
- DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* 3–113. Birkhäuser, Boston, MA. [MR1958777](#)

---

*MSC2010 subject classifications.* 62M10, 62G09.

*Key words and phrases.* Long memory, long-range dependence, resampling, subsampling, sampling window, block sampling, noncentral limit theorems, canonical correlation.

- DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to  $U$ -statistics. *Ann. Statist.* **17** 1767–1783. [MR1026312](#)
- DEHLING, H. and TAQQU, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference* **28** 153–165.
- DEO, R., HSIEH, M. and HURVICH, C. M. (2010). Long memory in intertrade durations, counts and realized volatility of NYSE stocks. *J. Statist. Plann. Inference* **140** 3715–3733. [MR2674160](#)
- DOBRUSHIN, R. L. (1979). Gaussian and their subordinated self-similar random generalized fields. *Ann. Probab.* **7** 1–28.
- DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52. [MR0550122](#)
- DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S., eds. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston, MA. [MR1956041](#)
- GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. World Scientific, Singapore.
- HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* **8** 1189–1204. [MR1666245](#)
- HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. [MR0614953](#)
- HOSKING, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econometrics* **73** 261–284.
- HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
- IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes. Applications of Mathematics* **9**. Springer, New York. Translated from the Russian by A. B. Aries. [MR0543837](#)
- JACH, A., MCELROY, T. and POLITIS, D. N. (2012). Subsampling inference for the mean of heavy-tailed long-memory time series. *J. Time Series Anal.* **33** 96–111. [MR2877610](#)
- KIM, Y. M. and NORDMAN, D. J. (2011). Properties of a block bootstrap under long-range dependence. *Sankhya A* **73** 79–109. [MR2887088](#)
- KOLMOGOROV, A. N. and ROZANOV, Y. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory Probab. Appl.* **5** 204–208.
- LAHIRI, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18** 405–413.
- LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Science & Business Media.
- LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96** 1066–1076. [MR1947254](#)
- MCELROY, T. and JACH, A. (2012). Subsampling inference for the autocovariances and autocorrelations of long-memory heavy-tailed linear time series. *J. Time Series Anal.* **33** 935–953. [MR2991910](#)
- NORDMAN, D. J. and LAHIRI, S. N. (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* **21** 1087–1111.
- PIPIRAS, V. and TAQQU, M. S. (2017). *Long-Range Dependence and Self-Similarity*. Cambridge Univ. Press, Cambridge.
- POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. [MR1329181](#)
- POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York.
- PSARADAKIS, Z. (2010). On inference based on the one-sample sign statistic for long-range dependent data. *Comput. Statist.* **25** 329–340. [MR2639814](#)
- ROMANO, J. P. and WOLF, M. (1999). Subsampling inference for the mean in the heavy-tailed case. *Metrika* **50** 55–69.
- RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, New York.
- SEBER, G. A. F. (2008). *A Matrix Handbook for Statisticians*, Vol. 15. John Wiley & Sons, Hoboken, NJ.

- SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. [MR2758116](#)
- SHERMAN, M. and CARLSTEIN, E. (1996). Replicate histograms. *J. Amer. Statist. Assoc.* **91** 566–576.
- SINAI, Y. G. (1976). Self-similar probability distributions. *Theory Probab. Appl.* **21** 64–80.
- SLY, A. and HEYDE, C. (2008). Nonstandard limit theorem for infinite variance functionals. *Ann. Probab.* **36** 796–805. [MR2393998](#)
- SOULIER, P. (2009). Some applications of regular variation in probability and statistics. Unpublished Notes.
- TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Probab. Theory Related Fields* **50** 53–83.
- TAQQU, M. S. (2003). Fractional Brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence* 5–38. Birkhäuser, Boston, MA.
- TERRIN, N. and TAQQU, M. S. (1990). A noncentral limit theorem for quadratic forms of Gaussian stationary sequences. *J. Theoret. Probab.* **3** 449–475. [MR1057525](#)
- TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable. International Series of Monographs in Pure and Applied Mathematics* **34**. Macmillan, New York. [MR0192238](#)
- ZHANG, T., HO, H.-C., WENDLER, M. and WU, W. B. (2013). Block sampling under strong dependence. *Stochastic Process. Appl.* **123** 2323–2339.
- ZYGMUND, A. (2002). *Trigonometric Series, Vols. I, II*, 3rd ed. Cambridge Univ. Press, Cambridge. [MR1963498](#)

# COCOLASSO FOR HIGH-DIMENSIONAL ERROR-IN-VARIABLES REGRESSION

BY ABHIRUP DATTA AND HUI ZOU<sup>1</sup>

*Johns Hopkins University and University of Minnesota*

Much theoretical and applied work has been devoted to high-dimensional regression with clean data. However, we often face corrupted data in many applications where missing data and measurement errors cannot be ignored. Loh and Wainwright [*Ann. Statist.* **40** (2012) 1637–1664] proposed a non-convex modification of the Lasso for doing high-dimensional regression with noisy and missing data. It is generally agreed that the virtues of convexity contribute fundamentally the success and popularity of the Lasso. In light of this, we propose a new method named CoCoLasso that is convex and can handle a general class of corrupted datasets. We establish the estimation error bounds of CoCoLasso and its asymptotic sign-consistent selection property. We further elucidate how the standard cross validation techniques can be misleading in presence of measurement error and develop a novel calibrated cross-validation technique by using the basic idea in CoCoLasso. The calibrated cross-validation has its own importance. We demonstrate the superior performance of our method over the nonconvex approach by simulation studies.

## REFERENCES

- [1] BELLONI, A., ROSENBAUM, M. and TSYBAKOV, A. B. (2014). Linear and conic programming estimators in high-dimensional errors-in-variables models. Preprint. Available at [arXiv:1408.0241](https://arxiv.org/abs/1408.0241).
- [2] BELLONI, A., ROSENBAUM, M. and TSYBAKOV, A. B. (2016). An  $\{\ell_1, \ell_2, \ell_\infty\}$ -regularization approach to high-dimensional errors-in-variables models. *Electron. J. Stat.* **10** 1729–1750. [MR3522659](https://arxiv.org/abs/1603.04469)
- [3] BENJAMINI, Y. and SPEED, T. P. (2012). Estimation and correction for GC-content bias in high throughput sequencing. *Nucleic Acids Res.* **40** 72.
- [4] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](https://arxiv.org/abs/0807.3214)
- [5] BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Faund. Trends Mach. Learn.* **3** 1–122.
- [6] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](https://arxiv.org/abs/1008.2927)
- [7] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144.

---

*MSC2010 subject classifications.* Primary 62J07; secondary 62F12.

*Key words and phrases.* Convex optimization, error in variables, high-dimensional regression, LASSO, missing data.

- [8] CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35** 2313–2351.
- [9] DUCHI, J., SHALEV-SHWARTZ, S., SINGER, Y. and CHANDRA, T. (2008). Efficient projections onto the  $L_1$ -ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning (ICML '08)* 272–279. ACM, New York.
- [10] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499.
- [11] EFRON, B., HASTIE, T. and TIBSHIRANI, R. (2007). Discussion: “The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ” [Ann. Statist. **35** (2007), no. 6, 2313–2351; MR2382644] by E. Candes and T. Tao. *Ann. Statist.* **35** 2358–2364. [MR2382646](#)
- [12] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- [13] FAN, J. and LI, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *International Congress of Mathematicians. Vol. III* 595–622. Eur. Math. Soc., Zürich. [MR2275698](#)
- [14] FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20** 101–148. [MR2640659](#)
- [15] FRIEDMAN, J. H., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.
- [16] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York.
- [17] LOH, P. L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Ann. Statist.* **40** 1637–1664.
- [18] MAI, Q., ZOU, H. and YUAN, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42.
- [19] PURDOM, E. and HOLMES, S. P. (2005). Error distribution for gene expression data. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 16, 35. [MR2170432](#)
- [20] ROSENBAUM, M. and TSYBAKOV, A. B. (2010). Sparse recovery under matrix uncertainty. *Ann. Statist.* **38** 2620–2651.
- [21] ROSENBAUM, M. and TSYBAKOV, A. B. (2013). Improved matrix uncertainty selector. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat. (IMS) Collect.* **9** 276–290. IMS, Beachwood, OH. [MR3202640](#)
- [22] SLIJEPCJEVIC, S., MEGERIAN, S. and POTKONJAK, M. (2002). Location errors in wireless embedded sensor networks: Sources, models, and effects on applications. *Mob. Comput. Commun. Rev.* **6** 67–78.
- [23] SØRENSEN, Ø., FRIGESSI, A. and THORESEN, M. (2013). Measurement error in LASSO: Impact and likelihood bias correction. *Statist. Sinica* **23**. To appear.
- [24] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc., B* **58** 267–288. [MR1379242](#)
- [25] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. [MR2136641](#)
- [26] 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](#)
- [27] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [28] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. [MR2729873](#)
- [29] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)

- [30] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)
- [31] 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](#)

## CONSISTENT PARAMETER ESTIMATION FOR LASSO AND APPROXIMATE MESSAGE PASSING

BY ALI MOUSAVI\*, ARIAN MALEKI<sup>†</sup> AND RICHARD G. BARANIUK\*

*Rice University\** and *Columbia University<sup>†</sup>*

This paper studies the optimal tuning of the regularization parameter in LASSO or the threshold parameters in approximate message passing (AMP). Considering a model in which the design matrix and noise are zero-mean i.i.d. Gaussian, we propose a data-driven approach for estimating the regularization parameter of LASSO and the threshold parameters in AMP. Our estimates are consistent, that is, they converge to their asymptotically optimal values in probability as  $n$ , the number of observations, and  $p$ , the ambient dimension of the sparse vector, grow to infinity, while  $n/p$  converges to a fixed number  $\delta$ . As a byproduct of our analysis, we will shed light on the asymptotic properties of the solution paths of LASSO and AMP.

### REFERENCES

- [1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: A geometric theory of phase transitions in convex optimization. Preprint. Available at [arXiv:1303.6672](https://arxiv.org/abs/1303.6672).
- [2] BAYATI, M., LELARGE, M. and MONTANARI, A. (2012). Universality in polytope phase transitions and message passing algorithms. Preprint. Available at [arXiv:1207.7321](https://arxiv.org/abs/1207.7321).
- [3] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inform. Theory* **58** 1997–2017. [MR2951312](#)
- [4] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory* **57** 764–785.
- [5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of LASSO and Dantzig selector. *Ann. Statist.* 1705–1732. [MR2533469](#)
- [6] BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007). Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.* **1** 169–194. [MR2312149](#)
- [7] CAI, T. T., WANG, L. and XU, G. (2010). Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.* **58** 1300–1308.
- [8] CAI, T. T., WANG, L. and XU, G. (2010). Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inform. Theory* **56** 3516–3522. [MR2799010](#)
- [9] CAI, T. T., XU, G. and ZHANG, J. (2009). On recovery of sparse signals via  $\ell_1$ -minimization. *IEEE Trans. Inform. Theory* **55** 3388–3397.
- [10] 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](#)
- [11] CHATTERJEE, S. and JAFAROV, J. (2015). Prediction error of cross-validated lasso. Preprint. Available at [arXiv:1502.06291](https://arxiv.org/abs/1502.06291).
- [12] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. [MR1639094](#)

---

*MSC2010 subject classifications.* 62G05, 62J05.

*Key words and phrases.* LASSO, estimation, sparsity, approximate message passing.



- [13] 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. [MR2546388](#)
- [14] DONOHO, D. L. (2006). High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.* **35** 617–652. [MR2225676](#)
- [15] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [16] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). Noise sensitivity phase transition. *IEEE Trans. Inform. Theory* **57** 6920–6941. [MR2882271](#)
- [17] DONOHO, D. L. and TANNER, J. (2005). Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA* **102** 9452–9457. [MR2168716](#)
- [18] DONOHO, D. L. and TANNER, J. (2009). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22** 1–53. [MR2449053](#)
- [19] DONOHO, D. L. and TANNER, J. (2010). Precise undersampling theorems. *Proc. IEEE* **98** 913–924.
- [20] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#)
- [21] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. and FRANKLIN, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *Math. Intelligencer* **27** 83–85.
- [22] HOMRIGHAUSEN, D. and McDONALD, D. J. (2014). Leave-one-out cross-validation is risk consistent for lasso. *Mach. Learn.* **97** 65–78.
- [23] KNIGHT, K. and FU, W. (2000). Asymptotics for LASSO-type estimators. *Ann. Statist.* **28** 1356–1378.
- [24] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2012). A significance test for the LASSO. Preprint. Available at [arXiv:1301.7161](#).
- [25] MALEKI, A. (2010). Approximate message passing algorithm for compressed sensing. Ph.D. Thesis, Stanford Univ., Stanford, CA.
- [26] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- [27] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. [MR2488351](#)
- [28] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2013). Parameterless optimal approximate message passing. Preprint. Available at [arXiv:1311.0035](#).
- [29] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2017). Supplement to “Consistent parameter estimation for LASSO and approximate message passing.” DOI:[10.1214/16-AOS1529SUPP](#).
- [30] OYMAK, S. and HASSIBI, B. (2012). On a relation between the minimax risk and the phase transitions of compressed recovery. In *Proc. Allerton Conf. Communication, Control, and Computing* 1018–1025, Urbana-Champaign, IL.
- [31] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over  $\ell_q$ -Balls. *IEEE Trans. Inform. Theory* **57** 6976–6994.
- [32] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* 1135–1151.
- [33] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [34] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. [MR3066375](#)
- [35] 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](#)

- [36] WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. [MR2543689](#)
- [37] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- [38] ZHANG, T. (2009). Some sharp performance bounds for least squares regression with  $L_1$  regularization. *Ann. Statist.* **37** 2109–2144. [MR2543687](#)
- [39] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)
- [40] ZHENG, L., MALEKI, A., WANG, X. and LONG, T. (2015). Does  $\ell_p$ -minimization outperform  $\ell_1$ -minimization? Preprint. Available at [arXiv:1501.03704](#).
- [41] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the degrees of freedom of the LASSO. *Ann. Statist.* **35** 2173–2192. [MR2363967](#)

## SUPPORT RECOVERY WITHOUT INCOHERENCE: A CASE FOR NONCONVEX REGULARIZATION

BY PO-LING LOH<sup>1,2</sup> AND MARTIN J. WAINWRIGHT<sup>2</sup>

*University of Wisconsin—Madison and University of California, Berkeley*

We develop a new primal-dual witness proof framework that may be used to establish variable selection consistency and  $\ell_\infty$ -bounds for sparse regression problems, even when the loss function and regularizer are nonconvex. We use this method to prove two theorems concerning support recovery and  $\ell_\infty$ -guarantees for a regression estimator in a general setting. Notably, our theory applies to all potential stationary points of the objective and certifies that the stationary point is unique under mild conditions. Our results provide a strong theoretical justification for the use of nonconvex regularization: For certain nonconvex regularizers with vanishing derivative away from the origin, any stationary point can be used to recover the support without requiring the typical incoherence conditions present in  $\ell_1$ -based methods. We also derive corollaries illustrating the implications of our theorems for composite objective functions involving losses such as least squares, nonconvex modified least squares for errors-in-variables linear regression, the negative log likelihood for generalized linear models and the graphical Lasso. We conclude with empirical studies that corroborate our theoretical predictions.

## REFERENCES

- [1] BERTSEKAS, D. P. (1999). *Nonlinear Programming*. Athena Scientific, Belmont, MA.
- [2] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- [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](#)
- [4] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- [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](#)
- [6] CANDÈS, E. J., WAKIN, M. B. and BOYD, S. P. (2008). Enhancing sparsity by reweighted  $l_1$  minimization. *J. Fourier Anal. Appl.* **14** 877–905. [MR2461611](#)
- [7] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. [MR1639094](#)
- [8] CLARKE, F. H. (1975). Generalized gradients and applications. *Trans. Amer. Math. Soc.* **205** 247–262. [MR0367131](#)
- [9] DONOHO, D. L. and STARK, P. B. (1989). Uncertainty principles and signal recovery. *SIAM J. Appl. Math.* **49** 906–931.

---

*MSC2010 subject classifications.* 62F12.

*Key words and phrases.*  $M$ -estimator, Lasso, sparsity, nonconvex regularizer, high-dimensional statistics, variable selection.

- [10] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.
- [11] FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory* **57** 5467–5484. [MR2849368](#)
- [12] FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961.
- [13] FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. [MR3210988](#)
- [14] FLETCHER, R. and WATSON, G. A. (1980). First- and second-order conditions for a class of nondifferentiable optimization problems. *Math. Program.* **18** 291–307. [MR0571992](#)
- [15] HUNTER, D. R. and LI, R. (2005). Variable selection using MM algorithms. *Ann. Statist.* **33** 1617–1642.
- [16] LEE, J. D., SUN, Y. and TAYLOR, J. E. (2015). On model selection consistency of regularized M-estimators. *Electron. J. Stat.* **9** 608–642. [MR3331852](#)
- [17] LOH, P. and WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616.
- [18] LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. [MR3015038](#)
- [19] LOH, P.-L. and WAINWRIGHT, M. J. (2017). Supplement to “Support recovery without incoherence: A case for nonconvex regularization.” DOI:10.1214/16-AOS1530SUPP.
- [20] LOUNICI, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electron. J. Stat.* **2** 90–102. [MR2386087](#)
- [21] MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). *SparseNet*: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. [MR2894769](#)
- [22] NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. [MR3025133](#)
- [23] NESTEROV, Y. and NEMIROVSKII, A. (1987). *Interior Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, PA.
- [24] PAN, Z. and ZHANG, C. (2015). Relaxed sparse eigenvalue conditions for sparse estimation via non-convex regularized regression. *Pattern Recognit.* **48** 231–243.
- [25] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259.
- [26] RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using  $\ell_1$ -regularized logistic regression. *Ann. Statist.* **38** 1287–1319. [MR2662343](#)
- [27] RUDELSON, M. and ZHOU, S. (2013). Reconstruction from anisotropic random measurements. *IEEE Trans. Inform. Theory* **59** 3434–3447. [MR3061256](#)
- [28] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [29] 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](#)
- [30] VAVASIS, S. A. (1995). Complexity issues in global optimization: A survey. In *Handbook of Global Optimization. Nonconvex Optim. Appl.* **2** 27–41. Kluwer Academic, Dordrecht. [MR1377083](#)
- [31] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. [MR2729873](#)
- [32] WAINWRIGHT, M. J. (2009). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inform. Theory* **55** 5728–5741.

- [33] WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201.
- [34] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- [35] ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. [MR3025135](#)
- [36] ZHANG, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.* **11** 1081–1107. [MR2629825](#)
- [37] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators. *Electron. J. Stat.* **11** 752–799. [MR3622646](#)
- [38] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2567.
- [39] ZHENG, Z., FAN, Y. and LV, J. (2014). High dimensional thresholded regression and shrinkage effect. *J. Roy. Statist. Soc. Ser. B* **76** 627–649.
- [40] ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. [MR2435443](#)

# OPTIMAL DESIGN OF FMRI EXPERIMENTS USING CIRCULANT (ALMOST-)ORTHOGONAL ARRAYS<sup>1</sup>

BY YUAN-LUNG LIN<sup>\*</sup>, FREDERICK KIN HING PHOA<sup>\*</sup> AND  
 MING-HUNG KAO<sup>†</sup>

*Academia Sinica<sup>\*</sup> and Arizona State University<sup>†</sup>*

Functional magnetic resonance imaging (fMRI) is a pioneering technology for studying brain activity in response to mental stimuli. Although efficient designs on these fMRI experiments are important for rendering precise statistical inference on brain functions, they are not systematically constructed. Design with circulant property is crucial for estimating a hemodynamic response function (HRF) and discussing fMRI experimental optimality. In this paper, we develop a theory that not only successfully explains the structure of a circulant design, but also provides a method of constructing efficient fMRI designs systematically. We further provide a class of two-level circulant designs with good performance (statistically optimal), and they can be used to estimate the HRF of a stimulus type and study the comparison of two HRFs. Some efficient three- and four-levels circulant designs are also provided, and we proved the existence of a class of circulant orthogonal arrays.

## REFERENCES

- [1] BURACAS, G. T. and BOYNTON, G. M. (2002). Efficient design of event-related fmri experiments using m-sequences. *Neuroimage* **16** 801–813.
- [2] CHAKRAVARTI, I. M. (1956). Fractional replication in asymmetrical factorial designs and partially balanced arrays. *Sankhyā* **17** 143–164. [MR0085677](#)
- [3] CHAKRAVARTI, I. M. (1961). On some methods of construction of partially balanced arrays. *Ann. Math. Stat.* **32** 1181–1185. [MR0130770](#)
- [4] CHÊNG, C. S. (1978). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6** 1239–1261. [MR0523760](#)
- [5] CHÊNG, C. S. (1980). Orthogonal arrays with variable numbers of symbols. *Ann. Statist.* **8** 447–453. [MR0560740](#)
- [6] CHENG, C.-S. (1995). Some projection properties of orthogonal arrays. *Ann. Statist.* **23** 1223–1233. [MR1353503](#)
- [7] CHENG, C.-S. (2014). Optimal biased weighing designs and two-level main-effect plans. *J. Stat. Theory Pract.* **8** 83–99. [MR3196641](#)
- [8] CHENG, C.-S. and KAO, M.-H. (2015). Optimal experimental designs for fMRI via circulant biased weighing designs. *Ann. Statist.* **43** 2565–2587. [MR3405604](#)
- [9] CHENG, C.-S., KAO, M.-H. and PHOA, F. K. H. (2017). Optimal and efficient designs for functional brain imaging experiments. *J. Statist. Plann. Inference* **181** 71–80. [MR3567999](#)

---

*MSC2010 subject classifications.* Primary 05B10, 05B15; secondary 62K15.

*Key words and phrases.* Circulant almost orthogonal arrays, complete difference system, design efficiency, hemodynamic response function.

- [10] CRAIGEN, R., FAUCHER, G., LOW, R. and WARES, T. (2013). Circulant partial Hadamard matrices. *Linear Algebra Appl.* **439** 3307–3317. [MR3119854](#)
- [11] DALE, A. M. (1999). Optimal experimental design for event-related fmri. *Human Brain Mapping* **8** 109–114.
- [12] GOLOMB, S. W. and GONG, G. (2005). *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge Univ. Press, Cambridge. [MR2156522](#)
- [13] HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York. [MR1693498](#)
- [14] JANSMA, J. M., DE ZWART, J. A., VAN GELDEREN, P., DUYN, J. H., DREVETS, W. C. and FUREY, M. L. (2013). In vivo evaluation of the effect of stimulus distribution on fir statistical efficiency in event-related fmri. *Journal of Neuroscience Methods* **215** 190–195.
- [15] JONES, B. and NACHTSHEIM, C. J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology* **43** 1–15.
- [16] KAO, M.-H. (2013). On the optimality of extended maximal length linear feedback shift register sequences. *Statist. Probab. Lett.* **83** 1479–1483. [MR3048312](#)
- [17] KAO, M.-H. (2014). A new type of experimental designs for event-related fMRI via Hadamard matrices. *Statist. Probab. Lett.* **84** 108–112. [MR3131263](#)
- [18] KAO, M.-H. (2015). Universally optimal fMRI designs for comparing hemodynamic response functions. *Statist. Sinica* **25** 499–506. [MR3379084](#)
- [19] KAO, M.-H., MANDAL, A., LAZAR, N. and STUFKEN, J. (2009). Multi-objective optimal experimental designs for event-related fmri studies. *NeuroImage* **44** 849–856.
- [20] LAZAR, N. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, New York.
- [21] LIN, Y.-L., PHOA, F. K. H. and KAO, M.-H. (2016+). Circulant partial Hadamard designs: Construction via general difference sets and its application to fmri experiments. Submitted.
- [22] LIN, Y.-L., PHOA, F. K. H. and KAO, M.-H. (2017). Supplement to “Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays.” DOI:[10.1214/16-AOS1531SUPP](#).
- [23] LIN, Y.-L. and PHOA, F. K. H. (2016). Constructing near-Hadamard designs with (almost)  $D$ -optimality by general supplementary difference sets. *Statist. Sinica* **26** 413–427. [MR3468358](#)
- [24] LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. [MR2530545](#)
- [25] LIU, T. T. (2004). Efficiency, power, and entropy in event-related fmri with multiple trial types: Part II: Design of experiments. *Neuroimage* **21** 401–413.
- [26] LIU, T. T. and FRANK, L. R. (2004). Efficiency, power, and entropy in event-related FMRI with multiple trial types. Part I: Theory. *Neuroimage* **21** 387–400.
- [27] LOW, R. M., STAMP, M., CRAIGEN, R. and FAUCHER, G. (2005). Unpredictable binary strings. *Congr. Numer.* **177** 65–75. [MR2198651](#)
- [28] MAUS, B., VAN BREUKELEN, G. J., GOEBEL, R. and BERGER, M. P. (2010). Robustness of optimal design of fmri experiments with application of a genetic algorithm. *NeuroImage* **49** 2433–2443.
- [29] PALEY, R. (1933). On orthogonal matrices. *J. Math. Phys.* 311–320.
- [30] PHOA, F. K. H. and LIN, D. K. J. (2015). A systematic approach for the construction of definitive screening designs. *Statist. Sinica* **25** 853–861. [MR3409727](#)
- [31] RADHAKRISHNA RAO, C. (1945). Finite geometries and certain derived results in theory of numbers. *Proc. Nat. Inst. Sci. India* **11** 136–149. [MR0023865](#)
- [32] RADHAKRISHNA RAO, C. (1946). Hypercubes of strength “ $d$ ” leading to confounded designs in factorial experiments. *Bull. Calcutta Math. Soc.* **38** 67–78. [MR0019291](#)

- [33] RAKTOE, B. L., HEDAYAT, A. and FEDERER, W. T. (1981). *Factorial Designs*. Wiley, New York. [MR0633756](#)
- [34] RYSER, H. J. (1963). *Combinatorial Mathematics. The Carus Mathematical Monographs* **14**. Wiley, New York. [MR0150048](#)
- [35] SINGER, J. (1938). A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.* **43** 377–385. [MR1501951](#)
- [36] STANTON, R. G. and SPROTT, D. A. (1958). A family of difference sets. *Canad. J. Math.* **10** 73–77. [MR0091249](#)



# ADAPTIVE BERNSTEIN–VON MISES THEOREMS IN GAUSSIAN WHITE NOISE<sup>1</sup>

BY KOLYAN RAY

*Leiden University*

We investigate Bernstein–von Mises theorems for adaptive nonparametric Bayesian procedures in the canonical Gaussian white noise model. We consider both a Hilbert space and multiscale setting with applications in  $L^2$  and  $L^\infty$ , respectively. This provides a theoretical justification for plug-in procedures, for example the use of certain credible sets for sufficiently smooth linear functionals. We use this general approach to construct optimal frequentist confidence sets based on the posterior distribution. We also provide simulations to numerically illustrate our approach and obtain a visual representation of the geometries involved.

## REFERENCES

- [1] ABRAMOVICH, F., SAPATINAS, T. and SILVERMAN, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 725–749. [MR1649547](#)
- [2] BELITSER, E. and GHOSAL, S. (2003). Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. *Ann. Statist.* **31** 536–559. [MR1983541](#)
- [3] BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *Ann. Statist.* **40** 206–237. [MR3013185](#)
- [4] BONTEMPS, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. [MR2906878](#)
- [5] BOUCHERON, S. and GASSIAT, E. (2009). A Bernstein–von Mises theorem for discrete probability distributions. *Electron. J. Stat.* **3** 114–148. [MR2471588](#)
- [6] BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. [MR2988456](#)
- [7] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. [MR2471287](#)
- [8] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99.
- [9] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. [MR3262477](#)
- [10] CASTILLO, I. (2015). Discussion of “Frequentist coverage of adaptive nonparametric Bayesian credible sets”. *Ann. Statist.* **43** 1437–1443. [MR3357863](#)
- [11] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028.
- [12] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. [MR3262473](#)

---

*MSC2010 subject classifications.* Primary 62G20; secondary 62G15, 62G08.

*Key words and phrases.* Bayesian inference, posterior asymptotics, adaptation, credible set, confidence set.

- [13] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383.
- [14] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. [MR3375874](#)
- [15] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. [MR3059077](#)
- [16] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. [MR1232525](#)
- [17] DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–67. [MR0829555](#)
- [18] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. [MR1740119](#)
- [19] GHOSAL, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331.
- [20] GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646.
- [21] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170.
- [22] GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **40**. Cambridge Univ. Press, Cambridge.
- [23] HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409.
- [24] HOFFMANN, M., ROUSSEAU, J. and SCHMIDT-HIEBER, J. (2015). On adaptive posterior concentration rates. *Ann. Statist.* **43** 2259–2295.
- [25] JOHNSTONE, I. M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. Collect.* **6** 87–98. IMS, Beachwood, OH.
- [26] KNAPIK, B. T., SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2016). Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Related Fields* **164** 771–813. [MR3477780](#)
- [27] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. [MR2906881](#)
- [28] KUEH, A. (2012). Locally adaptive density estimation on the unit sphere using needlets. *Constr. Approx.* **36** 433–458.
- [29] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. [MR2802048](#)
- [30] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory. Springer Series in Statistics*. Springer, New York. [MR0856411](#)
- [31] LOW, M. G. and MA, Z. (2015). Discussion of “Frequentist coverage of adaptive nonparametric Bayesian credible sets”. *Ann. Statist.* **43** 1448–1454. [MR3357865](#)
- [32] LOW, M. G. and ZHOU, H. H. (2007). A complement to Le Cam’s theorem. *Ann. Statist.* **35** 1146–1165. [MR2341701](#)
- [33] MEYER, Y. (1992). *Wavelets and Operators. Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge.
- [34] NICKL, R. (2015). Discussion of “Frequentist coverage of adaptive nonparametric Bayesian credible sets.” *Ann. Statist.* **43** 1429–1436.
- [35] NICKL, R. and SZABÓ, B. (2016). A sharp adaptive confidence ball for self-similar functions. *Stochastic Process. Appl.* **126** 3913–3934.
- [36] PETRONE, S., ROUSSEAU, J. and SCRICCILO, C. (2014). Bayes and empirical Bayes: Do they merge? *Biometrika* **101** 285–302.

- [37] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. [MR3117105](#)
- [38] RAY, K. (2014). Asymptotic theory for Bayesian nonparametric procedures in inverse problems Ph.D. thesis, University of Cambridge.
- [39] RAY, K. (2016). Supplement to “Adaptive Bernstein–von Mises theorems in Gaussian white noise”. DOI:[10.1214/16-AOS1533SUPP](#).
- [40] RAY, K. and SCHMIDT-HIEBER, J. (2016). The Le Cam distance between density estimation, Poisson processes and Gaussian white noise. ArXiv E-prints.
- [41] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523.
- [42] SZABÓ, B., VAN DER VAART, A. and VAN ZANTEN, H. (2015). Honest Bayesian confidence sets for the  $L^2$ -norm. *J. Statist. Plann. Inference* **166** 36–51.
- [43] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428.
- [44] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Rejoinder to discussions of “Frequentist coverage of adaptive nonparametric Bayesian credible sets”. *Ann. Statist.* **43** 1463–1470. [MR3357867](#)
- [45] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)

# TARGETED SEQUENTIAL DESIGN FOR TARGETED LEARNING INFERENCE OF THE OPTIMAL TREATMENT RULE AND ITS MEAN REWARD

BY ANTOINE CHAMBAZ<sup>\*,†,1,2</sup>, WENJING ZHENG<sup>†,1</sup> AND  
 MARK J. VAN DER LAAN<sup>†,3</sup>

*UPL, Université Paris Nanterre\* and University of California, Berkeley<sup>†</sup>*

This article studies the targeted sequential inference of an optimal treatment rule (TR) and its mean reward in the nonexceptional case, *that is*, assuming that there is no stratum of the baseline covariates where treatment is neither beneficial nor harmful, and under a companion margin assumption.

Our pivotal estimator, whose definition hinges on the targeted minimum loss estimation (TMLE) principle, actually infers the mean reward under the current estimate of the optimal TR. This data-adaptive statistical parameter is worthy of interest on its own. Our main result is a central limit theorem which enables the construction of confidence intervals on both mean rewards under the current estimate of the optimal TR and under the optimal TR itself. The asymptotic variance of the estimator takes the form of the variance of an efficient influence curve at a limiting distribution, allowing to discuss the efficiency of inference.

As a by product, we also derive confidence intervals on two cumulated pseudo-regrets, a key notion in the study of bandits problems.

A simulation study illustrates the procedure. One of the cornerstones of the theoretical study is a new maximal inequality for martingales with respect to the uniform entropy integral.

## REFERENCES

- [1] BALZER, L. B., PETERSEN, M. L., VAN DER LAAN, M. J. and THE SEARCH COLLABORATION (2016). Targeted estimation and inference for the sample average treatment effect in trials with and without pair-matching. *Stat. Med.* **35** 3717–3732. [MR3538043](#)
- [2] BUBECK, S. and CESA-BIANCHI, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Found. Trends Mach. Learn.* **5** 1–122.
- [3] CHAKRABORTY, B., LABER, E. B. and ZHAO, Y-Q. (2014). Inference about the expected performance of a data-driven dynamic treatment regime. *Clin. Trials* **11** 408–417.
- [4] CHAKRABORTY, B. and MOODIE, E. E. M. (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer, New York. [MR3112454](#)
- [5] CHAMBAZ, A. (2016). *tsml.cara.rct*: Targeted sequential minimum loss CARA RCT design and inference. R package version 0.1.0.
- [6] CHAMBAZ, A. and VAN DER LAAN, M. J. (2014). Inference in targeted group-sequential covariate-adjusted randomized clinical trials. *Scand. J. Stat.* **41** 104–140. [MR3181135](#)

---

*MSC2010 subject classifications.* Primary 62G05, 62G20; secondary 62K99.

*Key words and phrases.* Bandits, optimal treatment rule, precision medicine, pseudo-regret, targeted minimum loss estimation (TMLE).

- [7] CHAMBAZ, A., VAN DER LAAN, M. J. and ZHENG, W. (2015). Targeted covariate-adjusted response-adaptive lasso-based randomized controlled trials. In *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects* (O. Sverdlov, ed.) 345–368. CRC Press, Boca Raton, FL.
- [8] CHAMBAZ, A., ZHENG, W. and VAN DER LAAN, M. J. (2017). Supplement to “Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward.” DOI:[10.1214/16-AOS1534SUPP](https://doi.org/10.1214/16-AOS1534SUPP).
- [9] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling: From Dependence to Independence*. Springer, New York. [MR1666908](#)
- [10] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.
- [11] GARIVIER, A. and KAUFMANN, E. (2016). Optimal best arm identification with fixed confidence. *Proc. Mach. Learn. Res.* **49** 998–1027.
- [12] GOLDBERG, Y., SONG, R., ZENG, D. and KOSOROK, M. R. (2014). Comment on “Dynamic treatment regimes: Technical challenges and applications”. *Electron. J. Stat.* **8** 1290–1300.
- [13] IBRAGIMOV, I. A. and HAS’MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory. Applications of Mathematics 16*. Springer, New York–Berlin. Translated from the Russian by Samuel Kotz. [MR0620321](#)
- [14] JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, Boca Raton, FL. [MR1710781](#)
- [15] KAUFMANN, E. (2014). Analyse de stratégies bayésiennes et fréquentistes pour l’allocation séquentielle de ressources. Ph.D. thesis, TELECOM ParisTech. Available at <http://chercheurs.lille.inria.fr/ekaufman/TheseEmilie.pdf>.
- [16] LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **8** 1225–1272.
- [17] LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Rejoinder of “Dynamic treatment regimes: Technical challenges and applications” [[MR3263119](#); [MR3263120](#); [MR3263121](#); [MR3263122](#); [MR3263118](#)]. *Electron. J. Stat.* **8** 1312–1321. [MR3263123](#)
- [18] LUEDTKE, A. R. and VAN DER LAAN, M. J. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J. Causal Inference* **3** 61–95.
- [19] LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. [MR3476615](#)
- [20] LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.* **12** 305–332. [MR3505699](#)
- [21] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829.
- [22] PEARL, J. (2000). *Causality: Models, Reasoning and Inference, Vol. 29*. Cambridge Univ. Press, Cambridge.
- [23] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics 13*. Springer, New York–Berlin.
- [24] QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210.
- [25] ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proc. Second Seattle Symp. Biostat* (D. Y. Lin and P. Heagerty, eds.) 189–326.
- [26] ROBINS, J. M. and ROTNITZKY, A. (2014). Discussion of “Dynamic treatment regimes: Technical challenges and applications.” *Electron. J. Stat.* **8** 1273–1289.
- [27] RUBIN, D. B. and VAN DER LAAN, M. J. (2012). Statistical issues and limitations in personalized medicine research with clinical trials. *Int. J. Biostat.* **8** Article 1.

- [28] STANLEY, K. (2007). Design of randomized controlled trials. *Circulation* **115** 1164–1169.
- [29] R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.
- [30] VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York. [MR1958123](#)
- [31] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, New York.
- [32] VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11. [MR2306500](#)
- [33] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence*. Springer, Berlin.
- [34] ZHANG, B., TSIATIS, A., DAVIDIAN, M., ZHANG, M. and LABER, E. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018.
- [35] ZHANG, B., TSIATIS, A., DAVIDIAN, M., ZHANG, M. and LABER, E. (2012). Estimating optimal treatment regimes from a classification perspective. *Stat* **68** 103–114.
- [36] ZHAO, Y., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598.
- [37] ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118.
- [38] ZHENG, W., CHAMBAZ, A. and VAN DER LAAN, M. J. (2015). Drawing valid targeted inference when covariate-adjusted response-adaptive RCT meets data-adaptive loss-based estimation, with an application to the LASSO. Technical Report 339, Univ. California, Berkeley, Division of Biostatistics Working Paper Series.

# NONPARAMETRIC GOODNESS-OF-FIT TESTS FOR UNIFORM STOCHASTIC ORDERING<sup>1</sup>

BY CHUAN-FA TANG, DEWEI WANG AND JOSHUA M. TEBBS

*University of South Carolina*

We propose  $L^p$  distance-based goodness-of-fit (GOF) tests for uniform stochastic ordering with two continuous distributions  $F$  and  $G$ , both of which are unknown. Our tests are motivated by the fact that when  $F$  and  $G$  are uniformly stochastically ordered, the ordinal dominance curve  $R = FG^{-1}$  is star-shaped. We derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration of  $F$  and  $G$  for  $p \in [1, \infty]$ . We use simulation to assess finite-sample performance and demonstrate that a modified, one-sample version of our procedure (e.g., with  $G$  known) is more powerful than the one-sample GOF test suggested by Arcones and Samaniego [*Ann. Statist.* **28** (2000) 116–150]. We also discuss sample size determination. We illustrate our methods using data from a pharmacology study evaluating the effects of administering caffeine to prematurely born infants.

## REFERENCES

- ARANDA, J., BEHARRY, K., VALENCIA, G., NATARAJAN, G. and DAVIS, J. (2010). Caffeine impact on neonatal morbidities. *J. Matern.-Fetal Neonatal Med.* **23** 20–23.
- ARCONES, M. A. and SAMANIEGO, F. J. (2000). On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. *Ann. Statist.* **28** 116–150. [MR1762906](#)
- BAMBER, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psych.* **12** 387–415. [MR0384214](#)
- BEARE, B. and MOON, J. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* **31** 471–492. [MR3348455](#)
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. Wiley, New York.
- CAROLAN, C. A. and TEBBS, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika* **92** 159–171. [MR2158617](#)
- COX, C., HASHEM, N., TEBBS, J., BOOKSTAVAR, B. and ISKERSKY, V. (2015). Evaluation of caffeine and the development of necrotizing enterocolitis. *J. Neonatal-Perinatal Med.* **8** 339–347.
- DARDANONI, V. and FORCINA, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *J. Amer. Statist. Assoc.* **93** 1112–1123. [MR1649205](#)
- DAVIDOV, O. and HERMAN, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 825–847. [MR2988908](#)
- DOBSON, N. and HUNT, C. (2013). Caffeine use in neonates: Indications, pharmacokinetics, clinical effects, outcomes. *NeoReviews* **14** 540–550.

---

*MSC2010 subject classifications.* Primary 62G10; secondary 62F30.

*Key words and phrases.* Brownian bridge, hazard rate ordering, least favorable distribution, order-restricted inference, ordinal dominance curve, star-shaped ordering.

- DYKSTRA, R., KOCHAR, S. and ROBERTSON, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Ann. Statist.* **19** 870–888. [MR1105849](#)
- EL BARMÍ, H. and MCKEAGUE, I. (2016). Testing for uniform stochastic ordering via empirical likelihood. *Ann. Inst. Statist. Math.* **68** 955–976. [MR3538581](#)
- EL BARMÍ, H. and MUKERJEE, H. (2016). Consistent estimation of survival functions under uniform stochastic ordering: The  $k$ -sample case. *J. Multivariate Anal.* **144** 99–109. [MR3434943](#)
- HSIEH, F. and TURNBULL, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24** 25–40. [MR1389878](#)
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.
- LEHMANN, E. and ROJO, J. (1992). Invariant directional orderings. *Ann. Statist.* **20** 2100–2110. [MR1193328](#)
- MUKERJEE, H. (1996). Estimation of survival functions under uniform stochastic ordering. *J. Amer. Statist. Assoc.* **91** 1684–1689. [MR1439110](#)
- PARK, C., LEE, C. and ROBERTSON, T. (1998). Goodness-of-fit test for uniform stochastic ordering among several populations. *Canad. J. Statist.* **26** 69–81.
- ROJO, J. and SAMANIEGO, F. J. (1993). On estimating a survival curve subject to a uniform stochastic ordering constraint. *J. Amer. Statist. Assoc.* **88** 566–572. [MR1224382](#)
- SCHMIDT, B., ROBERTS, R., DAVIS, P., DOYLE, L., BARRINGTON, K., OHLSSON, A., SOLIMANO, A. and TIN, W. (2006). Caffeine therapy for apnea of prematurity. *N. Engl. J. Med.* **354** 2112–2121.
- SHAKED, M. and SHANTHIKUMAR, J. (2007). *Stochastic Orders*. Springer, New York.
- SHAPIRO, A. (1990). On concepts of directional differentiability. *J. Optim. Theory Appl.* **66** 477–487. [MR1080259](#)
- SHAPIRO, A. (1991). Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* **30** 169–186. [MR1118896](#)
- TANG, C., WANG, D. and TEBBS, J. (2017). Supplement to “Nonparametric goodness-of-fit tests for uniform stochastic ordering.” DOI:[10.1214/16-AOS1535SUPP](#).
- VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.



# SELECTING THE NUMBER OF PRINCIPAL COMPONENTS: ESTIMATION OF THE TRUE RANK OF A NOISY MATRIX

BY YUNJIN CHOI\*, JONATHAN TAYLOR<sup>†,1</sup> AND ROBERT TIBSHIRANI<sup>†,2</sup>

*National University of Singapore\* and Stanford University<sup>†</sup>*

Principal component analysis (PCA) is a well-known tool in multivariate statistics. One significant challenge in using PCA is the choice of the number of principal components. In order to address this challenge, we propose distribution-based methods with *exact* type 1 error controls for hypothesis testing and construction of confidence intervals for signals in a noisy matrix with finite samples. Assuming Gaussian noise, we derive exact type 1 error controls based on the conditional distribution of the singular values of a Gaussian matrix by utilizing a post-selection inference framework, and extending the approach of [Taylor, Loftus and Tibshirani (2013)] in a PCA setting. In simulation studies, we find that our proposed methods compare well to existing approaches.

## REFERENCES

- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- CAI, J.-F., CANDÈS, E. J. and SHEN, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20** 1956–1982. [MR2600248](#)
- CHOI, Y., TAYLOR, J. and TIBSHIRANI, R. (2017). Supplement to “Selecting the number of principal components: Estimation of the true rank of a noisy matrix.” DOI:[10.1214/16-AOS1536SUPP](#).
- EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#)
- FITHIAN, W., TAYLOR, J., TIBSHIRANI, R. and TIBSHIRANI, R. (2015). Selective Sequential Model Selection. Preprint. Available at [arXiv:1512.02565](#).
- G’SSELL, M. G., WAGER, S., CHOULDECHOVA, A. and TIBSHIRANI, R. (2013). Sequential Selection Procedures And False Discovery Rate Control. Preprint. Available at [arXiv:1309.5352](#).
- GAVISH, M. and DONOHO, D. L. (2014). The optimal hard threshold for singular values is  $4/\sqrt{3}$ . *IEEE Trans. Inform. Theory* **60** 5040–5053. [MR3245370](#)
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. [MR2722294](#)
- JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. [MR0181057](#)
- JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. Springer, New York. [MR2036084](#)
- JOSSE, J. and HUSSON, F. (2012). Selecting the number of components in principal component analysis using cross-validation approximations. *Comput. Statist. Data Anal.* **56** 1869–1879. [MR2892383](#)

---

*MSC2010 subject classifications.* 62J05, 62J07, 62F03.

*Key words and phrases.* Principal components, hypothesis test, exact  $p$ -value.

- KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.
- MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York-Toronto, Ont. [MR0560319](#)
- MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322.
- MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. [MR0652932](#)
- NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. [MR2485013](#)
- PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **77** 1617–1642.
- REID, S., TIBSHIRANI, R. and FRIEDMAN, J. (2016). A study of error variance estimation in Lasso regression. *Statist. Sinica* **26** 35–67. [MR3468344](#)
- SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. [MR897872](#)
- TAYLOR, J., LOFTUS, J. and TIBSHIRANI, R. (2013). Tests in adaptive regression via the Kac-Rice formula. Preprint. Available at [arXiv:1308.3020](#).
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- TIBSHIRANI, R., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2014). Exact Post-Selection Inference for Sequential Regression Procedures. Preprint. Available at [arXiv:1401.3889](#).

## EXTENDED CONDITIONAL INDEPENDENCE AND APPLICATIONS IN CAUSAL INFERENCE

BY PANAYIOTA CONSTANTINOU AND A. PHILIP DAWID

*University of Warwick and University of Cambridge*

The goal of this paper is to integrate the notions of stochastic conditional independence and variation conditional independence under a more general notion of extended conditional independence. We show that under appropriate assumptions the calculus that applies for the two cases separately (axioms of a separoid) still applies for the extended case. These results provide a rigorous basis for a wide range of statistical concepts, including ancillarity and sufficiency, and, in particular, the Decision Theoretic framework for statistical causality, which uses the language and calculus of conditional independence in order to express causal properties and make causal inferences.

### REFERENCES

- [1] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. [MR0489333](#)
- [2] BASU, D. (1964). Recovery of ancillary information. *Sankhya A* **26** 3–16. [MR0172366](#)
- [3] BERZUINI, C. R., DAWID, A. P. and DIDELEZ, V. (2012). Assessing dynamic treatment strategies. In *Causality: Statistical Perspectives and Applications* (C. R. Berzuini, A. P. Dawid and L. Bernardinelli, eds.) Chapter 8 85–100. Wiley, Chichester.
- [4] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. [MR1324786](#)
- [5] BORKAR, V. S. (1995). *Probability Theory: An Advanced Course*, 1st ed. Springer, New York. [MR1367959](#)
- [6] CASELLA, G. and BERGER, R. L. (2001). *Statistical Inference*, 2nd ed. Duxbury Press. [MR1051420](#)
- [7] CONSTANTINOU, P. (2013). Conditional independence and applications in statistical causality. PhD Dissertation, Univ. Cambridge.
- [8] CONSTANTINOU, P. and DAWID, A. P. (2017). Supplement to “Extended conditional independence and applications in causal inference.” DOI:[10.1214/16-AOS1537SUPP](#).
- [9] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (2007). *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*, 2nd ed. Springer, Berlin. [MR1697175](#)
- [10] DAWID, A. P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. R. Stat. Soc., B* **37** 248–258.
- [11] DAWID, A. P. (1976). Properties of diagnostic data distributions. *Biometrics* **32** 647–658.
- [12] DAWID, A. P. (1979). Conditional independence in statistical theory. *J. R. Stat. Soc., B* **41** 1–31.
- [13] DAWID, A. P. (1979). Some misleading arguments involving conditional independence. *J. R. Stat. Soc., B* **41** 249–252. [MR0547250](#)

---

*MSC2010 subject classifications.* Primary 62A99; secondary 60A05.

*Key words and phrases.* Conditional independence, extended conditional independence, sufficiency, ancillarity, causality, separoid.

- [14] DAWID, A. P. (1980). A Bayesian look at nuisance parameters. In *Bayesian Statistics (Valencia, 1979)* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 167–203. Univ. Press, Valencia. With discussion. [MR0638877](#)
- [15] DAWID, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8** 598–617. [MR0568723](#)
- [16] DAWID, A. P. (1985). Invariance and independence in multivariate distribution theory. *J. Multivariate Anal.* **17** 304–315. [MR0813238](#)
- [17] DAWID, A. P. (1998). Conditional independence. In *Encyclopedia of Statistical Sciences, Vol. 2* (S. Kotz, C. B. Read and D. L. Banks, eds.) 146–155. Wiley-Interscience.
- [18] DAWID, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.* **32** 335–372.
- [19] DAWID, A. P. (2001). Some variations on variation independence. In *Proceedings of Artificial Intelligence and Statistics 2001* (T. Jaakkola and T. Richardson, eds.) 187–191.
- [20] DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *Int. Stat. Rev.* **70** 161–189.
- [21] DAWID, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance. In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort and S. Richardson, eds.) *Oxford Statist. Sci. Ser.* **27** 45–81. Oxford Univ. Press, Oxford. With part A by Elja Arjas and part B by James M. Robins. [MR2082406](#)
- [22] DAWID, A. P. (2012). The decision-theoretic approach to causal inference. In *Causality: Statistical Perspectives and Applications* (C. R. Berzuini, A. P. Dawid and L. Bernardinelli, eds.) Chapter 4 25–42. Wiley, Chichester.
- [23] DAWID, A. P. (2015). Statistical causality from a decision-theoretic perspective. *Annual Review of Statistics and Its Application* **2** 273–303.
- [24] DAWID, A. P. and CONSTANTINOU, P. (2014). A formal treatment of sequential ignorability. *Stat. Biosci.* **6** 166–188.
- [25] DAWID, A. P. and DICKEY, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.* **72** 845–850.
- [26] DAWID, A. P. and DIDELEZ, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Stat. Surv.* **4** 184–231. [MR2740837](#)
- [27] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. [Correction: *Ann. Statist.* **23** (1995) 1864.]
- [28] DAWID, A. P. and MORTERA, J. (1996). Coherent analysis of forensic identification evidence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 425–443. [MR1377842](#)
- [29] DAWID, A. P. and MORTERA, J. (1998). Forensic identification with imperfect evidence. *Biometrika* **85** 835–849. [Correction: *Biometrika* **86** (1999) 974.] [MR1666687](#)
- [30] DAWID, A. P. and STUDENÝ, M. (1999). Conditional products: An alternative approach to conditional independence. In *Artificial Intelligence and Statistics 99* (D. Heckerman and J. Whittaker, eds.) 32–40. Morgan Kaufmann.
- [31] DIDELEZ, V., DAWID, A. P. and GENELETTI, S. (2006). Direct and indirect effects of sequential treatments. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. S. Richardson, eds.) 138–146. AUAI Press, Arlington, VA.
- [32] DURRETT, R. (2013). *Probability: Theory and Examples*, 4th ed. Cambridge Univ. Press. [MR2722836](#)
- [33] FISHER, R. A. (1925). Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* **22** 700–725.
- [34] GEIGER, D., VERMA, T. and PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534. [MR1064736](#)

- [35] GENELETTI, S. and DAWID, A. P. (2011). Defining and identifying the effect of treatment on the treated. In *Causality in the Sciences* (P. M. Illari, F. Russo and J. Williamson, eds.) 728–749.
- [36] GHOSH, M., REID, N. and FRASER, D. A. S. (2010). Ancillary statistics: A review. *Statist. Sinica* **20** 1309–1332. [MR2777327](#)
- [37] GUO, H. and DAWID, A. P. (2010). Sufficient covariates and linear propensity analysis. In *Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics (AISTATS) 2010* (Y. W. Teh and D. M. Titterington, eds.). *Journal of Machine Learning Research Workshop and Conference Proceedings* **9** 281–288.
- [38] GUO, H., DAWID, A. P. and BERZUINI, G. M. (2016). Sufficient covariate, propensity variable and doubly robust estimation. In *Statistical Causal Inferences and Their Applications in Public Health Research*. (H. He, P. Wu and D.-G. Chen, eds.). Springer. To appear.
- [39] HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **1** 225–241. [MR0030730](#)
- [40] KINGMAN, J. F. C. (1993). *Poisson Processes*, 2nd ed. *Oxford Studies in Probability* **3**. Oxford Univ. Press, New York. [MR1207584](#)
- [41] KOLMOGOROFF, A. (1942). Sur l’estimation statistique des paramètres de la loi de Gauss. *Bull. Acad. Sci. URSS. Sér. Math. [Izvestia Akad. Nauk SSSR]* **6** 3–32. [MR0007959](#)
- [42] LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505. [MR1064735](#)
- [43] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 1st ed. Morgan Kaufmann, San Mateo, CA. [MR0965765](#)
- [44] PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. [MR1744773](#)
- [45] PEARL, J. and PAZ, A. (1986). Graphoids: Graph-based logic for reasoning about relevance relations or when would  $x$  tell you more about  $y$  if you already know  $z$ ? In *Advances in Artificial Intelligence II, Seventh European Conference on Artificial Intelligence, ECAI 86* (J. B. H. du Boulay, D. Hogg and L. Steels, eds.) 357–363.
- [46] RESNICK, S. I. (2014). *A Probability Path*, 1st ed. Birkhäuser. [MR3135152](#)
- [47] ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Modelling* **7** 1393–1512. [MR0877758](#)
- [48] ROBINS, J. (1987). Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect”. *Comput. Math. Appl.* **14** 923–945. [MR0922792](#)
- [49] ROBINS, J. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS* (L. Sechrest, H. Freeman and A. Mulley, eds.) 113–159.
- [50] SPOHN, W. (1994). On the properties of conditional independence. In *Patrick Suppes: Scientific Philosopher* (P. Humphreys, ed.). *Synthese Lib.* **233** 173–196. Kluwer Acad. Publ., Dordrecht. [MR1349350](#)

## A WEIGHT-RELAXED MODEL AVERAGING APPROACH FOR HIGH-DIMENSIONAL GENERALIZED LINEAR MODELS

BY TOMOHIRO ANDO<sup>\*,1</sup> AND KER-CHAU LI<sup>†,‡,2</sup>

*University of Melbourne<sup>\*</sup>, University of California, Los Angeles<sup>†</sup>  
and Academia Sinica<sup>‡</sup>*

Model averaging has long been proposed as a powerful alternative to model selection in regression analysis. However, how well it performs in high-dimensional regression is still poorly understood. Recently, Ando and Li [*J. Amer. Statist. Assoc.* **109** (2014) 254–265] introduced a new method of model averaging that allows the number of predictors to increase as the sample size increases. One notable feature of Ando and Li’s method is the relaxation on the total model weights so that weak signals can be efficiently combined from high-dimensional linear models. It is natural to ask if Ando and Li’s method and results can be extended to nonlinear models. Because all candidate models should be treated as working models, the existence of a theoretical target of the quasi maximum likelihood estimator under model misspecification needs to be established first. In this paper, we consider generalized linear models as our candidate models. We establish a general result to show the existence of pseudo-true regression parameters under model misspecification. We derive proper conditions for the leave-one-out cross-validation weight selection to achieve asymptotic optimality. Technically, the pseudo true target parameters between working models are not linearly linked. To overcome the encountered difficulties, we employ a novel strategy of decomposing and bounding the bias and variance terms in our proof. We conduct simulations to illustrate the merits of our model averaging procedure over several existing methods, including the lasso and group lasso methods, the Akaike and Bayesian information criterion model-averaging methods and some other state-of-the-art regularization methods.

## REFERENCES

- AKAIKE, H. (1978). On the likelihood of a time series model. *J. R. Stat. Soc., Ser. D Stat.* **27** 217–235.
- AKAIKE, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* **66** 237–242. [MR0548189](#)
- ANDO, T. (2009). Bayesian portfolio selection using multifactor model. *Int. J. Forecast.* **25** 550–566.
- ANDO, T. and LI, K.-C. (2014). A model-averaging approach for high-dimensional regression. *J. Amer. Statist. Assoc.* **109** 254–265. [MR3180561](#)
- ANDO, T. and LI, K. (2017). Supplement to “A weight-relaxed model averaging approach for high-dimensional generalized linear models.” DOI:[10.1214/17-AOS1538SUPP](https://doi.org/10.1214/17-AOS1538SUPP).

---

*MSC2010 subject classifications.* Primary 62J12; secondary 62F99.

*Key words and phrases.* Asymptotic optimality, high-dimensional regression models, model averaging, model misspecification.

- ANDO, T. and TSAY, R. (2010). Predictive likelihood for Bayesian model selection and averaging. *Int. J. Forecast.* **26** 744–763.
- 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](#)
- BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. K. (2010). Variable selection in high-dimensional linear models: Partially faithful distributions and the PC-simple algorithm. *Biometrika* **97** 261–278.
- CHARKHI, A., CLAESKENS, G. and HANSEN, B. E. (2016). Minimum mean squared error model averaging in likelihood models. *Statist. Sinica* **26** 809–840.
- CHUNG, T. S., RUST, R. T. and WEDEL, M. (2009). My mobile music: An adaptive personalization system for digital audio players. *Marketing Sci.* **28** 52–68.
- CLAESKENS, G. and HJORT, N. L. (2003). The focused information criterion. *J. Amer. Statist. Assoc.* **98** 900–945. [MR2041482](#)
- EKLUND, J. and KARLSSON, S. (2007). Forecast combination and model averaging using predictive measures. *Econometric Rev.* **26** 329–363. [MR2364365](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.
- FLYNN, C. J., HURVICH, C. M. and SIMONOFF, J. S. (2013). Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. *J. Amer. Statist. Assoc.* **108** 1031–1043.
- HANSEN, B. E. (2007). Least squares model averaging. *Econometrica* **75** 1175–1189. [MR2333497](#)
- HANSEN, B. E. and RACINE, J. S. (2012). Jackknife model averaging. *J. Econometrics* **167** 38–46. [MR2885437](#)
- HJORT, N. L. and CLAESKENS, G. (2003). Frequentist model average estimators. *J. Amer. Statist. Assoc.* **98** 879–899. [MR2041481](#)
- HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–417. [MR1765176](#)
- KASS, R. and RAFTERY, A. (1995). Bayes factors and model uncertainty. *J. Amer. Statist. Assoc.* **90** 773–795.
- LEE, Y. S. (2014). Management of a periodic-review inventory system using Bayesian model averaging when new marketing efforts are made. *Int. J. Production Econ.* **158** 278–289.
- LI, K.-C. (1986). Asymptotic optimality of  $C_L$  and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.* **14** 1011–1112.
- LI, K.-C. (1987). Asymptotic optimality for  $C_p$ ,  $C_L$ , cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15** 958–975.
- LV, J. and LIU, J. S. (2014). Model selection principles in misspecified models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 141–167.
- MADIGAN, D. and RAFTERY, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *J. Amer. Statist. Assoc.* **89** 1535–1546.
- MIN, K.-C. and ZELLNER, A. (1992). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *J. Econometrics* **56** 89–118.
- MONTGOMERY, J. M. and NYHAN, B. (2010). Bayesian model averaging: Theoretical developments and practical applications. *Polit. Anal.* **18** 245–270.
- MORO, S., LAUREANO, R. and CORTEZ, P. (2011). Using data mining for bank direct marketing: An application of the CRISP-DM methodology. In *Proceedings of the European Simulation and Modelling Conference* 117–121.
- OUYSSSE, R. and KOHN, R. (2010). Bayesian variable selection and model averaging in the arbitrage pricing theory model. *Comput. Statist. Data Anal.* **54** 3249–3268.
- RAFTERY, A. E., MADIGAN, D. and HOETING, J. A. (1997). Bayesian model averaging for linear regression models. *J. Amer. Statist. Assoc.* **92** 179–191.

- STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288.
- WAN, A. T. K., ZHANG, X. and ZOU, G. (2010). Least squares model averaging by Mallows criterion. *J. Econometrics* **156** 277–283. [MR2609932](#)
- WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25.
- YEUNG, K. E., BUMGARNER, R. E. and RAFTERY, A. E. (2005). Bayesian model averaging: Development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* **21** 2394–2402.
- YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67.
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHANG, Y., LI, R. and TSAI, C.-L. (2010). Regularization parameter selections via generalized information criterion. *J. Amer. Statist. Assoc.* **105** 312–323. [MR2656055](#)



# STRUCTURAL SIMILARITY AND DIFFERENCE TESTING ON MULTIPLE SPARSE GAUSSIAN GRAPHICAL MODELS<sup>1</sup>

BY WEIDONG LIU

*Shanghai Jiao Tong University*

We present a new framework on inferring structural similarities and differences among multiple high-dimensional Gaussian graphical models (GGMs) corresponding to the same set of variables under distinct experimental conditions. The new framework adopts the partial correlation coefficients to characterize the potential changes of dependency strengths between two variables. A hierarchical method has been further developed to recover edges with different or similar dependency strengths across multiple GGMs. In particular, we first construct two-sample test statistics for testing the equality of partial correlation coefficients and conduct large-scale multiple tests to estimate the substructure of differential dependencies. After removing differential substructure from original GGMs, a follow-up multiple testing procedure is used to detect the substructure of similar dependencies among GGMs. In each step, false discovery rate is controlled asymptotically at a desired level. Power results are proved, which demonstrate that our method is more powerful on finding common edges than the common approach that separately estimates GGMs. The performance of the proposed hierarchical method is illustrated on simulated datasets.

## REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. [MR1990662](#)
- BELILOVSKY, E., VAROQUAUX, G. and BLASCHKO, M. B. (2015). Hypothesis testing for differences in Gaussian graphical models: Applications to brain connectivity Technical report. <http://arxiv.org/abs/1512.08643>.
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300.
- CAI, T., LIU, W. and LUO, X. (2011). A constrained  $\ell_1$  minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35** 2313–2351.
- CHIQUET, J., GRANDVALET, Y. and AMBROISE, C. (2011). Inferring multiple graphical structures. *Stat. Comput.* **21** 537–553.
- CHU, J., LAZARUS, R., CAREY, V. J. and RABY, B. A. (2011). Quantifying differential gene connectivity between disease states for objective identification of disease-relevant genes. *BMC Syst. Biol.* **5** 89.

---

*MSC2010 subject classifications.* 62H12, 62H15.

*Key words and phrases.* Common substructure, false discovery rate, Gaussian graphical model, high dimensional, structural similarity, structural difference.

- COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation. Monographs on Statistics and Applied Probability* **67**. Chapman & Hall, London. [MR1456990](#)
- D'ASPREMONT, A., BANERJEE, O. and EL GHAOU, L. (2008). First-order methods for sparse covariance selection. *SIAM J. Matrix Anal. Appl.* **30** 56–66. [MR2399568](#)
- DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. [MR3164871](#)
- DE LA FUENTE, A. (2010). From “differential expression” to “differential networking”-identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26** 326–333.
- EFRON, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist. Assoc.* **102** 93–103. [MR2293302](#)
- FAN, J., FENG, Y. and WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* **3** 521–541. [MR2750671](#)
- FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostat.* **9** 432–441.
- GILL, R., DATTA, S. and DATTA, S. (2010). A statistical framework for differential network analysis from microarray data. *BMC Bioinformatics* **11** 95.
- GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional nonparanormal graphical models Technical report. <http://arxiv.org/abs/1502.02347>.
- GUO, J., LEVINA, E., MICHAELIDIS, G. and ZHU, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98** 1–15. [MR2804206](#)
- HARA, S. and WASHIO, T. (2013). Learning a common substructure of multiple graphical Gaussian models. *Neural Netw.* **38** 23–38.
- HONORIO, J. and SAMARAS, D. (2010). Multi-task learning of Gaussian graphical models. In *Proceedings of the 27th International Conference on Machine Learning*. Haifa, Israel.
- LI, H. and GUI, J. (2006). Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostat.* **7** 302–317.
- LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. [MR3161453](#)
- LIU, W. (2016). Supplement to “Structural similarity and difference testing on multiple sparse Gaussian graphical models”. DOI:[10.1214/17-AOS1539SUPP](https://doi.org/10.1214/17-AOS1539SUPP).
- LIU, W. and SHAO, Q. M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.* **42** 2003–2025.
- LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012). High-dimensional semi-parametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326. [MR3059084](#)
- MA, S., GONG, Q. and BOHNERT, H. J. (2007). An arabidopsis gene network based on the graphical Gaussian model. *Genome Res.* **17** 1614–1625.
- MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- NEGAHBAN, S. N. and WAINWRIGHT, M. J. (2011). Simultaneous support recovery in high dimensions: Benefits and perils of block  $\ell_1/\ell_\infty$ -regularization. *IEEE Trans. Inform. Theory* **57** 3841–3863. [MR2817058](#)
- OBOZINSKI, G., WAINWRIGHT, M. and JORDAN, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *Ann. Statist.* **39** 1–47.
- RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing  $\ell_1$ -penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. [MR2836766](#)
- REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. [MR3346695](#)

- ROTHMAN, A., BICKEL, P., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515.
- SCHÄFER, J. and STRIMMER, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21** 754–764.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. [MR1379242](#)
- XUE, L. and ZOU, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist.* **40** 2541–2571. [MR3097612](#)
- YANG, S., LU, Z., SHEN, X., WONKA, P. and YE, J. (2015). Fused multiple graphical lasso. *SIAM J. Optim.* **25** 916–943. [MR3343365](#)
- YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. [MR2719856](#)
- YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35.
- ZHANG, C. (2010). Estimation of large inverse matrices and graphical model selection. Technical Report. Rutgers University, Department of Statistics and Biostatistics.
- ZHANG, B. and WANG, Y. (2010). Learning structural changes of Gaussian graphical models in controlled experiments. In *Proc. of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)* 701–708, Avalon, CA.
- ZHAO, S. D., CAI, T. T. and LI, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268.
- ZHU, D., HERO, A. O., QIN, Z. S. and SWAROOP, A. (2005). High throughput screening of co-expressed gene pairs with controlled false discovery rate (FDR) and minimum acceptable strength (MAS). *J. Comput. Biol.* **12** 1029–1045.
- ZOLOTAREV, M., (1961). Concerning a certain probability problem. *Theory Probab. Appl.* **6** 201–204.

## ESTIMATING A PROBABILITY MASS FUNCTION WITH UNKNOWN LABELS<sup>1</sup>

BY DRAGI ANEVSKI<sup>2</sup>, RICHARD D. GILL AND STEFAN ZOHREN<sup>3</sup>

*Lund University, Leiden University and University of Oxford*

In the context of a species sampling problem, we discuss a nonparametric maximum likelihood estimator for the underlying probability mass function. The estimator is known in the computer science literature as the high profile estimator. We prove strong consistency and derive the rates of convergence, for an extended model version of the estimator. We also study a sieved estimator for which similar consistency results are derived. Numerical computation of the sieved estimator is of great interest for practical problems, such as forensic DNA analysis, and we present a computational algorithm based on the stochastic approximation of the expectation maximisation algorithm. As an interesting byproduct of the numerical analyses, we introduce an algorithm for bounded isotonic regression for which we also prove convergence.

## REFERENCES

- [1] ACHARYA, J., ORLITSKY, A. and PAN, S. (2009). The maximum likelihood probability of unique-singleton, ternary, and length-7 patterns. In *IEEE International Symposium on Information Theory* 1135–1139.
- [2] ANEVSKI, D. and FOUGÈRES, A.-L. (2007). Limit properties of the monotone rearrangement for density and regression function estimation. Lund University. Available at [arXiv:0710.4617v1](https://arxiv.org/abs/0710.4617v1).
- [3] ANEVSKI, D., GILL, R. D. and ZOHREN, S. (2017). Supplement to “Estimating a probability mass function with unknown labels.” DOI:[10.1214/17-AOS1542SUPP](https://doi.org/10.1214/17-AOS1542SUPP).
- [4] BALABDAOUI, F., RUFIBACH, K. and SANTAMBROGIO, F. (2010). Least-squares estimation of two-ordered monotone regression curves. *J. Nonparametr. Stat.* **22** 1019–1037. [MR2738880](https://arxiv.org/abs/1005.0001)
- [5] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642–669. [MR0083864](https://arxiv.org/abs/1005.0001)
- [6] EFRON, B. and THISTED, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.
- [7] ESTY, W. W. (1982). Confidence intervals for the coverage of low coverage samples. *Ann. Statist.* **10** 190–196. [MR0642730](https://arxiv.org/abs/1005.0001)
- [8] ESTY, W. W. (1983). A normal limit law for a nonparametric estimator of the coverage of a random sample. *Ann. Statist.* **11** 905–912. [MR0707940](https://arxiv.org/abs/1005.0001)
- [9] FISHER, R. A., CORBET, A. S. and WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**.

---

*MSC2010 subject classifications.* 62G05, 62G20, 65C60, 62P10.

*Key words and phrases.* NPML, high profile, probability mass function, strong consistency, sieve, ordered, monotone rearrangement, nonparametric, SA-EM, rates.

- [10] GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264. [MR0061330](#)
- [11] GOOD, I. J. and TOULMIN, G. H. (1956). The population frequencies of species and the estimation of population parameters. *Biometrika* **43** 45–63.
- [12] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge University Press, Cambridge.
- [13] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electron. J. Stat.* **3** 1567–1605. [MR2578839](#)
- [14] LIEB, E. H. and LOSS, M. (1997). *Analysis. Graduate Studies in Mathematics* **14**. Amer. Math. Soc., Providence, RI. [MR1415616](#)
- [15] MAO, C. X. and LINDSAY, B. G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika* **89** 669–681. [MR1929171](#)
- [16] MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283.
- [17] ORLITSKY, A. and PAN, S. (2009). The maximum likelihood probability of skewed patterns. In *IEEE International Symposium on Information Theory*.
- [18] ORLITSKY, A., SAJAMA, S., SANTHANAM, N. P., VISWANATHAN, K. and ZHANG, J. (2004). Algorithms for modeling distributions over large alphabets. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on Information Theory* 304.
- [19] ORLITSKY, A., SAJAMA, S., SANTHANAM, N. P., VISWANATHAN, K. and ZHANG, J. (2004). On modeling profiles instead of values. In *Proceeding UAI'04 Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* 426–435.
- [20] ORLITSKY, A., SAJAMA, S., SANTHANAM, N. P., VISWANATHAN, K. and ZHANG, J. (2005). Convergence of profile based estimators. In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on Information Theory* 1843–1847.
- [21] ORLITSKY, A., SANTHANAM, N. P., VISWANATHAN, K. and ZHANG, J. (2004). On modeling profiles instead of values. In *Proceedings of the Twentieth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-04)* 426–435. AUAI Press, Arlington, VA.
- [22] RAMANUJAN, S. and HARDY, G. H. (1918). Asymptotic formulae in combinatorial analysis. *Proc. Lond. Math. Soc.* **17** 75–115.
- [23] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- [24] VAN EEDEN, C. (1957). Maximum likelihood estimation of partially or completely ordered parameters. ii. *Indag. Math.* **60** 201–211.
- [25] VONTOBEL, P. O. (2012). The Bethe permanent of a non-negative matrix. *IEEE Trans. Inform. Theory* **59** 1866–1901.
- [26] VONTOBEL, P. O. (2014). The Bethe and Sinkhorn approximations of the pattern maximum likelihood estimate and their connections to the Valiant–Valiant estimate. In *Proceedings of Information Theory and Applications Workshop (ITA), 9–14 Feb.*
- [27] ZHANG, C.-H. and ZHANG, Z. (2009). Asymptotic normality of a nonparametric estimator of sample coverage. *Ann. Statist.* **37** 2582–2595. [MR2543704](#)

# OPTIMAL SEQUENTIAL DETECTION IN MULTI-STREAM DATA

BY HOCK PENG CHAN<sup>1</sup>

*National University of Singapore*

Consider a large number of detectors each generating a data stream. The task is to detect online, distribution changes in a small fraction of the data streams. Previous approaches to this problem include the use of mixture likelihood ratios and sum of CUSUMs. We provide here extensions and modifications of these approaches that are optimal in detecting normal mean shifts. We show how the (optimal) detection delay depends on the fraction of data streams undergoing distribution changes as the number of detectors goes to infinity. There are three detection domains. In the first domain for moderately large fractions, immediate detection is possible. In the second domain for smaller fractions, the detection delay grows logarithmically with the number of detectors, with an asymptotic constant extending those in sparse normal mixture detection. In the third domain for even smaller fractions, the detection delay lies in the framework of the classical detection delay formula of Lorden. We show that the optimal detection delay is achieved by the sum of detectability score transformations of either the partial scores or CUSUM scores of the data streams.

## REFERENCES

- [1] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. [MR2246369](#)
- [2] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.* **34** 326–349. [MR2275244](#)
- [3] CHAN, H. P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37** 3985–4010. [MR2572450](#)
- [4] CHAN, H. P. and WALTHER, G. (2015). Optimal detection of multi-sample aligned sparse signals. *Ann. Statist.* **43** 1865–1895. [MR3375870](#)
- [5] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#)
- [6] INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#)
- [7] INGSTER, Y. I. (1998). Minimax detection of a signal for  $l^n$ -balls. *Math. Methods Statist.* **7** 401–428. [MR1680087](#)
- [8] JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. [MR3034330](#)

---

*MSC2010 subject classifications.* 62G10, 62L10.

*Key words and phrases.* Average run length, CUSUM, detectability score, detection delay, mixture likelihood ratio, sparse detection, stopping rule.

- [9] LAI, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 613–658. [MR1354072](#)
- [10] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. [MR0309251](#)
- [11] MEI, Y. (2006). Sequential change-point detection when unknown parameters are present in the pre-change distribution. *Ann. Statist.* **34** 92–122. [MR2275236](#)
- [12] MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. [MR2650748](#)
- [13] MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387. [MR0868306](#)
- [14] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. [MR0773162](#)
- [15] POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779. [MR0888438](#)
- [16] SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. [MR0799155](#)
- [17] SIEGMUND, D., YAKIR, B. and ZHANG, N. R. (2011). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Stat.* **5** 645–668. [MR2840169](#)
- [18] TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2008). Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Anal.* **27** 441–475. [MR2460208](#)
- [19] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. [MR3099117](#)

THE ANNALS  
*of*  
STATISTICS

*AN OFFICIAL JOURNAL OF THE*  
INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME 45

2017



## CONTENTS OF VOLUME 45

### Articles

|                                                                                                                                                                       |           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| AI, MINGYAO, LI, KANG AND ZHENG, WEI. Identification of universally optimal circular designs for the interference model . . .                                         | 1462–1487 |
| AIROLDI, EDOARDO M. AND TOULIS, PANOS. Asymptotic and finite-sample properties of estimators based on stochastic gradients . . . . .                                  | 1694–1727 |
| ANDO, TOMOHIRO AND LI, KER-CHAU. A weight-relaxed model averaging approach for high-dimensional generalized linear models . . . . .                                   | 2654–2679 |
| ANEVSKI, DRAGI, GILL, RICHARD D. AND ZOHREN, STEFAN. Estimating a probability mass function with unknown labels . .                                                   | 2708–2735 |
| ARIAS-CASTRO, ERY AND VERZELEN, NICOLAS. Detection and feature selection in sparse mixture models . . . . .                                                           | 1920–1950 |
| ASTON, JOHN A. D., PIGOLI, DAVIDE AND TAVAKOLI, SHAHIN. Tests for separability in nonparametric covariance operators of random surfaces . . . . .                     | 1431–1461 |
| ATCHADÉ, YVES A. On the contraction properties of some high-dimensional quasi-posterior distributions . . . . .                                                       | 2248–2273 |
| BAI, SHUYANG AND TAQQU, MURAD S. On the validity of resampling methods under long memory . . . . .                                                                    | 2365–2399 |
| BALAKRISHNAN, SIVARAMAN, WAINWRIGHT, MARTIN J. AND YU, BIN. Statistical guarantees for the EM algorithm: From population to sample-based analysis . . . . .           | 77–120    |
| BARANIUK, RICHARD G., MOUSAVI, ALI AND MALEKI, ARIAN. Consistent parameter estimation for LASSO and approximate message passing . . . . .                             | 2427–2454 |
| BASU, KINJAL AND MUKHERJEE, RAJARSHI. Asymptotic normality of scrambled geometric net quadrature . . . . .                                                            | 1759–1788 |
| BELITSER, EDUARD. On coverage and local radial rates of credible sets . . . . .                                                                                       | 1124–1151 |
| BELLONI, ALEXANDRE AND OLIVEIRA, ROBERTO I. Approximate group context tree . . . . .                                                                                  | 355–385   |
| BERMANN, GEORGINA, BORNKAMP, BJÖRN, FELLER, CHRYSTEL, SCHORNING, KIRSTEN AND DETTE, HOLGER. Optimal designs for dose response curves with common parameters . . . . . | 2102–2132 |
| BHATTACHARYA, ANIRBAN, DUNSON, DAVID B. AND JOHNDROW, JAMES E. Tensor decompositions and sparse log-linear models . . . . .                                           | 1–38      |
| BIBINGER, MARKUS, JIRAK, MORITZ AND VETTER, MATHIAS. Nonparametric change-point analysis of volatility . . . . .                                                      | 1542–1578 |

|                                                                                                                                                                                   |           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| BICKEL, PETER J. AND WANG, Y. X. RACHEL. Likelihood-based model selection for stochastic block models . . . . .                                                                   | 500–528   |
| BOGDAN, MAŁGORZATA, CANDÈS, EMMANUEL AND SU, WEIJIE. False discoveries occur early on the Lasso path . . . . .                                                                    | 2133–2150 |
| BOISTARD, HÉLÈNE, LOPUHAÄ, HENDRIK P. AND RUIZ-GAZEN, ANNE. Functional central limit theorems for single-stage sampling designs . . . . .                                         | 1728–1758 |
| BORNKAMP, BJÖRN, FELLER, CHRYSTEL, SCHORNING, KIRSTEN, DETTE, HOLGER AND BERMAN, GEORGINA. Optimal designs for dose response curves with common parameters . . . . .              | 2102–2132 |
| BULL, ADAM D. Semimartingale detection and goodness-of-fit tests . . . . .                                                                                                        | 1254–1283 |
| CAI, T. TONY AND GUO, ZIJIAN. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity . . . . .                                                 | 615–646   |
| CAI, T. TONY, LIANG, TENG YUAN AND RAKHLIN, ALEXANDER. Computational and statistical boundaries for submatrix localization in a large noisy matrix . . . . .                      | 1403–1430 |
| CANDÈS, EMMANUEL, SU, WEIJIE AND BOGDAN, MAŁGORZATA. False discoveries occur early on the Lasso path . . . . .                                                                    | 2133–2150 |
| CARDOT, HERVÉ, CÉNAC, PEGGY AND GODICHON-BAGGIONI, ANTOINE. Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls . . . . .                 | 591–614   |
| CÉNAC, PEGGY, GODICHON-BAGGIONI, ANTOINE AND CARDOT, HERVÉ. Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls . . . . .                 | 591–614   |
| CHAKRABORTY, ANIRVAN AND CHAUDHURI, PROBAL. Tests for high-dimensional data based on means, spatial signs and spatial ranks . . . . .                                             | 771–799   |
| CHAMBAZ, ANTOINE, ZHENG, WENJING AND VAN DER LAAN, MARK J. Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward . . . . . | 2537–2564 |
| CHAN, HOCK PENG. Optimal sequential detection in multi-stream data . . . . .                                                                                                      | 2736–2763 |
| CHAUDHURI, PROBAL AND CHAKRABORTY, ANIRVAN. Tests for high-dimensional data based on means, spatial signs and spatial ranks . . . . .                                             | 771–799   |
| CHENG, DAN AND SCHWARTZMAN, ARMIN. Multiple testing of local maxima for detection of peaks in random fields . . . . .                                                             | 529–556   |
| CHERNOZHUKOV, VICTOR, GALICHON, ALFRED, HALLIN, MARC AND HENRY, MARC. Monge–Kantorovich depth, quantiles, ranks and signs . . . . .                                               | 223–256   |

|                                                                                                                                                                        |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| CHERNOZHUKOV, VICTOR, HANSEN, CHRISTIAN AND LIAO, YUAN. A lava attack on the recovery of sums of dense and sparse signals .....                                        | 39–76     |
| CHOI, DAVID. Co-clustering of nonsmooth graphons.....                                                                                                                  | 1488–1515 |
| CHOI, YUNJIN, TAYLOR, JONATHAN AND TIBSHIRANI, ROBERT. Selecting the number of principal components: Estimation of the true rank of a noisy matrix .....               | 2590–2617 |
| COLLIER, OLIVIER, COMMINGES, LAËTITIA AND TSYBAKOV, ALEXANDRE B. Minimax estimation of linear and quadratic functionals on sparsity classes .....                      | 923–958   |
| COMMINGES, LAËTITIA, TSYBAKOV, ALEXANDRE B. AND COLLIER, OLIVIER. Minimax estimation of linear and quadratic functionals on sparsity classes .....                     | 923–958   |
| CONSTANTINOU, PANAYIOTA AND DAWID, A. PHILIP. Extended conditional independence and applications in causal inference .                                                 | 2618–2653 |
| CUTTING, CHRISTINE, PAINDAVEINE, DAVY AND VERDEBOUT, THOMAS. Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives ..... | 1024–1058 |
| DATTA, ABHIRUP AND ZOU, HUI. CoCoLasso for high-dimensional error-in-variables regression .....                                                                        | 2400–2426 |
| DAWID, A. PHILIP AND CONSTANTINOU, PANAYIOTA. Extended conditional independence and applications in causal inference .                                                 | 2618–2653 |
| DEREUDRE, DAVID AND LAVANCIER, FRÉDÉRIC. Consistency of likelihood estimation for Gibbs point processes.....                                                           | 744–770   |
| DETTE, HOLGER, BERMAN, GEORGINA, BORNKAMP, BJÖRN, FELLER, CHRYSTEL AND SCHORNING, KIRSTEN. Optimal designs for dose response curves with common parameters ....        | 2102–2132 |
| DETTE, HOLGER, KONSTANTINOU, MARIA AND ZHIGLJAVSKY, ANATOLY. A new approach to optimal designs for correlated observations .....                                       | 1579–1608 |
| DICKER, LEE H. AND ERDOGDU, MURAT A. Flexible results for quadratic forms with applications to variance components estimation .....                                    | 386–414   |
| DOBRIBAN, EDGAR. Sharp detection in PCA under correlations: All eigenvalues matter .....                                                                               | 1810–1833 |
| DUKKIPATI, AMBEDKAR AND GHOSH DASTIDAR, DEBARGHYA. Consistency of spectral hypergraph partitioning under planted partition model.....                                  | 289–315   |
| DUNSON, DAVID B., JOHNDROW, JAMES E. AND BHATTACHARYA, ANIRBAN. Tensor decompositions and sparse log-linear models .....                                               | 1–38      |

|                                                                                                                                                                            |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| ERDOGDU, MURAT A. AND DICKER, LEE H. Flexible results for quadratic forms with applications to variance components estimation .....                                        | 386–414   |
| ERNST, PHILIP A., SHEPP, LARRY A. AND WYNER, ABRAHAM J. Yule’s “nonsense correlation” solved! .....                                                                        | 1789–1809 |
| FALLAT, SHAUN, LAURITZEN, STEFFEN, SADEGHI, KAYVAN, UHLER, CAROLINE, WERMUTH, NANNY AND ZWIERNIK, PIOTR. Total positivity in Markov structures .....                       | 1152–1184 |
| FAN, JIANQING AND WANG, WEICHEN. Asymptotics of empirical eigenstructure for high dimensional spiked covariance .....                                                      | 1342–1374 |
| FAN, YINGYING, LV, JINCHI, KONG, YINFEI AND LI, DAOJI. Interaction pursuit in high-dimensional multi-response regression via distance correlation .....                    | 897–922   |
| FELLER, CHRYSTEL, SCHORNING, KIRSTEN, DETTE, HOLGER, BERMANN, GEORGINA AND BORNKAMP, BJÖRN. Optimal designs for dose response curves with common parameters ....           | 2102–2132 |
| FUKUMIZU, KENJI, LIU, SONG, SUZUKI, TAIJI, RELATOR, RAISSA, SESE, JUN AND SUGIYAMA, MASASHI. Support consistency of direct sparse-change learning in Markov networks ..... | 959–990   |
| GALICHON, ALFRED, HALLIN, MARC, HENRY, MARC AND CHERNOZHUKOV, VICTOR. Monge–Kantorovich depth, quantiles, ranks and signs .....                                            | 223–256   |
| GAO, CHAO, MA, ZONGMING AND ZHOU, HARRISON H. Sparse CCA: Adaptive estimation and computational barriers .....                                                             | 2074–2101 |
| GHOSAL, SUBHASHIS AND LI, MENG. Bayesian detection of image boundaries .....                                                                                               | 2190–2217 |
| GHOSHDASTIDAR, DEBARGHYA AND DUKKIPATI, AMBEDKAR. Consistency of spectral hypergraph partitioning under planted partition model .....                                      | 289–315   |
| GILL, RICHARD D., ZOHREN, STEFAN AND ANEVSKI, DRAGI. Estimating a probability mass function with unknown labels ...                                                        | 2708–2735 |
| GODICHON-BAGGIONI, ANTOINE, CARDOT, HERVÉ AND CÉNAC, PEGGY. Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls .....              | 591–614   |
| GRIGAS, PAUL, MAZUMDER, RAHUL AND M. FREUND, ROBERT. A new perspective on boosting in linear regression via subgradient optimization and relatives .....                   | 2328–2364 |
| GUO, ZIJIAN AND CAI, T. TONY. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity ...                                                | 615–646   |
| HALLIN, MARC, HENRY, MARC, CHERNOZHUKOV, VICTOR AND GALICHON, ALFRED. Monge–Kantorovich depth, quantiles, ranks and signs .....                                            | 223–256   |

|                                                                                                                                                      |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| HAN, DONG, TSUNG, FUGEE AND XIAN, JINGUO. On the optimality of Bayesian change-point detection .....                                                 | 1375–1402 |
| HANG, HANYUAN AND STEINWART, INGO. A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning ..... | 708–743   |
| HANSEN, CHRISTIAN, LIAO, YUAN AND CHERNOZHUKOV, VICTOR. A lava attack on the recovery of sums of dense and sparse signals .....                      | 39–76     |
| HASTIE, TREVOR, OWEN, ART B., WANG, JINGSHU AND ZHAO, QINGYUAN. Confounder adjustment in multiple hypothesis testing .....                           | 1863–1894 |
| HENRY, MARC, CHERNOZHUKOV, VICTOR, GALICHON, ALFRED AND HALLIN, MARC. Monge–Kantorovich depth, quantiles, ranks and signs.....                       | 223–256   |
| HU, RUI AND WIENS, DOUGLAS P. Robust discrimination designs over Hellinger neighbourhoods .....                                                      | 1638–1663 |
| JAMES, LANCELOT F. Bayesian Poisson calculus for latent feature modeling via generalized Indian Buffet Process priors.....                           | 2016–2045 |
| JIANG, TIEFENG, LEDER, KEVIN AND XU, GONGJUN. Rare-event analysis for extremal eigenvalues of white Wishart matrices ...                             | 1609–1637 |
| JIN, JIASHUN, KE, ZHENG TRACY AND WANG, WANJIE. Phase transitions for high dimensional clustering and related problems                               | 2151–2189 |
| JIRAK, MORITZ, VETTER, MATHIAS AND BIBINGER, MARKUS. Nonparametric change-point analysis of volatility .....                                         | 1542–1578 |
| JOHNDROW, JAMES E., BHATTACHARYA, ANIRBAN AND DUNSON, DAVID B. Tensor decompositions and sparse log-linear models .....                              | 1–38      |
| KAO, MING-HUNG, LIN, YUAN-LUNG AND PHOA, FREDERICK KIN HING. Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays .....     | 2483–2510 |
| KE, ZHENG TRACY, WANG, WANJIE AND JIN, JIASHUN. Phase transitions for high dimensional clustering and related problems                               | 2151–2189 |
| KHARE, KSHITIJ, PAL, SUBHADIP AND SU, ZHIHUA. A Bayesian approach for envelope models .....                                                          | 196–222   |
| KLOPP, OLGA, TSYBAKOV, ALEXANDRE B. AND VERZELEN, NICOLAS. Oracle inequalities for network models and sparse graphon estimation .....                | 316–354   |
| KOLTCHINSKII, VLADIMIR AND LOUNICI, KARIM. Normal approximation and concentration of spectral projectors of sample covariance.....                   | 121–157   |

|                                                                                                                                                                                     |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| KONG, WEIHAO AND VALIANT, GREGORY. Spectrum estimation from samples .....                                                                                                           | 2218–2247 |
| KONG, YINFEI, LI, DAOJI, FAN, YINGYING AND LV, JINCHI. Interaction pursuit in high-dimensional multi-response regression via distance correlation .....                             | 897–922   |
| KONSTANTINOU, MARIA, ZHIGLJAVSKY, ANATOLY AND DETTE, HOLGER. A new approach to optimal designs for correlated observations .....                                                    | 1579–1608 |
| LAN, WEI, SONG, PETER X. K., TSAI, CHIH-LING AND ZHONG, PING-SHOU. Tests for covariance structures with high-dimensional repeated measurements .....                                | 1185–1213 |
| LAURITZEN, STEFFEN, SADEGHI, KAYVAN, UHLER, CAROLINE, WERMUTH, NANNY, ZWIERNIK, PIOTR AND FALLAT, SHAUN. Total positivity in Markov structures .....                                | 1152–1184 |
| LAVANCIER, FRÉDÉRIC AND DEREUDRE, DAVID. Consistency of likelihood estimation for Gibbs point processes .....                                                                       | 744–770   |
| LEDER, KEVIN, XU, GONGJUN AND JIANG, TIEFENG. Rare-event analysis for extremal eigenvalues of white Wishart matrices ...                                                            | 1609–1637 |
| LEE, YOUNG K., MAMMEN, ENNO, NIELSEN, JENS P. AND PARK, BYEONG U. Operational time and in-sample density forecasting .....                                                          | 1312–1341 |
| LI, BING AND SONG, JUN. Nonlinear sufficient dimension reduction for functional data .....                                                                                          | 1059–1095 |
| LI, DAOJI, FAN, YINGYING, LV, JINCHI AND KONG, YINFEI. Interaction pursuit in high-dimensional multi-response regression via distance correlation .....                             | 897–922   |
| LI, GUODONG, LIU, YUEWEN, WANG, HANSHENG, ZHU, XUEN-ING AND PAN, RUI. Network vector autoregression .....                                                                           | 1096–1123 |
| LI, JUN AND ZHONG, PING-SHOU. A rate optimal procedure for recovering sparse differences between high-dimensional means under dependence .....                                      | 557–590   |
| LI, KANG, ZHENG, WEI AND AI, MINGYAO. Identification of universally optimal circular designs for the interference model ...                                                         | 1462–1487 |
| LI, KER-CHAU AND ANDO, TOMOHIRO. A weight-relaxed model averaging approach for high-dimensional generalized linear models .....                                                     | 2654–2679 |
| LI, LINGLING, MUKHERJEE, RAJARSHI, TCHETGEN, ERIC TCHETGEN, VAN DER VAART, AAD AND ROBINS, JAMES M. Minimax estimation of a functional on a structured high-dimensional model ..... | 1951–1987 |

|                                                                                                                                                                            |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| LI, MENG AND GHOSAL, SUBHASHIS. Bayesian detection of image boundaries .....                                                                                               | 2190–2217 |
| LI, ZENG, WANG, QINWEN AND YAO, JIANFENG. Identifying the number of factors from singular values of a large sample autocovariance matrix .....                             | 257–288   |
| LIANG, TENGYUAN, RAKHLIN, ALEXANDER AND CAI, T. TONY. Computational and statistical boundaries for submatrix localization in a large noisy matrix .....                    | 1403–1430 |
| LIAO, YUAN, CHERNOZHUKOV, VICTOR AND HANSEN, CHRISTIAN. A lava attack on the recovery of sums of dense and sparse signals .....                                            | 39–76     |
| LIN, YUAN-LUNG, PHOA, FREDERICK KIN HING AND KAO, MING-HUNG. Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays .....                           | 2483–2510 |
| LIU, HAN AND NING, YANG. A general theory of hypothesis tests and confidence regions for sparse high dimensional models ....                                               | 158–195   |
| LIU, HAN, NING, YANG AND ZHAO, TIANQI. A likelihood ratio framework for high-dimensional semiparametric regression ....                                                    | 2299–2327 |
| LIU, SONG, SUZUKI, TAIJI, RELATOR, RAISSA, SESE, JUN, SUGIYAMA, MASASHI AND FUKUMIZU, KENJI. Support consistency of direct sparse-change learning in Markov networks ..... | 959–990   |
| LIU, WEIDONG. Structural similarity and difference testing on multiple sparse Gaussian graphical models .....                                                              | 2680–2707 |
| LIU, YUEWEN, WANG, HANSHENG, ZHU, XUENING, PAN, RUI AND LI, GUODONG. Network vector autoregression .....                                                                   | 1096–1123 |
| LOH, PO-LING. Statistical consistency and asymptotic normality for high-dimensional robust $M$ -estimators .....                                                           | 866–896   |
| LOH, PO-LING AND WAINWRIGHT, MARTIN J. Support recovery without incoherence: A case for nonconvex regularization ....                                                      | 2455–2482 |
| LOK, JUDITH J. Mimicking counterfactual outcomes to estimate causal effects .....                                                                                          | 461–499   |
| LOPUHAÄ, HENDRIK P., RUIZ-GAZEN, ANNE AND BOISTARD, HÉLÈNE. Functional central limit theorems for single-stage sampling designs .....                                      | 1728–1758 |
| LOUNICI, KARIM AND KOLTCHINSKII, VLADIMIR. Normal approximation and concentration of spectral projectors of sample covariance .....                                        | 121–157   |
| LV, JINCHI, KONG, YINFEI, LI, DAOJI AND FAN, YINGYING. Interaction pursuit in high-dimensional multi-response regression via distance correlation .....                    | 897–922   |

|                                                                                                                                                                                                     |           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| M. FREUND, ROBERT, GRIGAS, PAUL AND MAZUMDER, RAHUL.<br>A new perspective on boosting in linear regression via subgradient optimization and relatives . . . . .                                     | 2328–2364 |
| MA, ZONGMING, ZHOU, HARRISON H. AND GAO, CHAO. Sparse<br>CCA: Adaptive estimation and computational barriers . . . . .                                                                              | 2074–2101 |
| MAATHUIS, MARLOES H., RICHARDSON, THOMAS S. AND NANDY,<br>PREETAM. Estimating the effect of joint interventions from<br>observational data in sparse high-dimensional settings . . . . .            | 647–674   |
| MALEKI, ARIAN, BARANIUK, RICHARD G. AND MOUSAVI, ALI.<br>Consistent parameter estimation for LASSO and approximate<br>message passing . . . . .                                                     | 2427–2454 |
| MAMMEN, ENNO, NIELSEN, JENS P., PARK, BYEONG U. AND<br>LEE, YOUNG K. Operational time and in-sample density fore-<br>casting . . . . .                                                              | 1312–1341 |
| MAZUMDER, RAHUL, M. FREUND, ROBERT AND GRIGAS, PAUL.<br>A new perspective on boosting in linear regression via subgradient<br>optimization and relatives . . . . .                                  | 2328–2364 |
| MENDELSON, SHAHAR. “Local” vs. “global” parameters—<br>breaking the Gaussian complexity barrier . . . . .                                                                                           | 1835–1862 |
| METELKINA, ASYA AND PRONZATO, LUC. Information-regret<br>compromise in covariate-adaptive treatment allocation . . . . .                                                                            | 2046–2073 |
| MOUSAVI, ALI, MALEKI, ARIAN AND BARANIUK, RICHARD G.<br>Consistent parameter estimation for LASSO and approximate<br>message passing . . . . .                                                      | 2427–2454 |
| MUKHERJEE, RAJARSHI AND BASU, KINJAL. Asymptotic nor-<br>mality of scrambled geometric net quadrature . . . . .                                                                                     | 1759–1788 |
| MUKHERJEE, RAJARSHI, TCHETGEN, ERIC TCHETGEN, VAN DER<br>VAART, AAD, ROBINS, JAMES M. AND LI, LINGLING. Min-<br>imax estimation of a functional on a structured high-dimensional<br>model . . . . . | 1951–1987 |
| NANDY, PREETAM, MAATHUIS, MARLOES H. AND RICHARDSON,<br>THOMAS S. Estimating the effect of joint interventions from<br>observational data in sparse high-dimensional settings . . . . .             | 647–674   |
| NICKL, RICHARD AND SÖHL, JAKOB. Nonparametric Bayesian<br>posterior contraction rates for discretely observed scalar diffu-<br>sions . . . . .                                                      | 1664–1693 |
| NIELSEN, JENS P., PARK, BYEONG U., LEE, YOUNG K. AND<br>MAMMEN, ENNO. Operational time and in-sample density<br>forecasting . . . . .                                                               | 1312–1341 |
| NING, YANG AND LIU, HAN. A general theory of hypothesis tests<br>and confidence regions for sparse high dimensional models . . . .                                                                  | 158–195   |



|                                                                                                                                                                            |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| NING, YANG, ZHAO, TIANQI AND LIU, HAN. A likelihood ratio framework for high-dimensional semiparametric regression . . .                                                   | 2299–2327 |
| OLIVEIRA, ROBERTO I. AND BELLONI, ALEXANDRE. Approximate group context tree . . . . .                                                                                      | 355–385   |
| OVERGAARD, MORTEN, PARNER, ERIK THORLUND AND PEDERSEN, JAN. Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations . . . . .        | 1988–2015 |
| OWEN, ART B., WANG, JINGSHU, ZHAO, QINGYUAN AND HASTIE, TREVOR. Confounder adjustment in multiple hypothesis testing . . . . .                                             | 1863–1894 |
| PAINDAVEINE, DAVY AND VERDEBOUT, THOMAS. Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities . . . . .        | 800–832   |
| PAINDAVEINE, DAVY, VERDEBOUT, THOMAS AND CUTTING, CHRISTINE. Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives . . . . . | 1024–1058 |
| PAL, SUBHADIP, SU, ZHIHUA AND KHARE, KSHITIJ. A Bayesian approach for envelope models . . . . .                                                                            | 196–222   |
| PAN, RUI, LI, GUODONG, LIU, YUEWEN, WANG, HANSHENG AND ZHU, XUENING. Network vector autoregression . . . . .                                                               | 1096–1123 |
| PARK, BYEONG U., LEE, YOUNG K., MAMMEN, ENNO AND NIELSEN, JENS P. Operational time and in-sample density forecasting . . . . .                                             | 1312–1341 |
| PARNER, ERIK THORLUND, PEDERSEN, JAN AND OVERGAARD, MORTEN. Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations . . . . .        | 1988–2015 |
| PEDERSEN, JAN, OVERGAARD, MORTEN AND PARNER, ERIK THORLUND. Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations . . . . .        | 1988–2015 |
| PENSKY, MARIANNA. Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity . .                                             | 1516–1541 |
| PHOA, FREDERICK KIN HING, KAO, MING-HUNG AND LIN, YUAN-LUNG. Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays . . . . .                       | 2483–2510 |
| PIGOLI, DAVIDE, TAVAKOLI, SHAHIN AND ASTON, JOHN A. D. Tests for separability in nonparametric covariance operators of random surfaces . . . . .                           | 1431–1461 |
| PILANCI, MERT, WAINWRIGHT, MARTIN J. AND YANG, YUN. Randomized sketches for kernels: Fast and optimal nonparametric regression . . . . .                                   | 991–1023  |

|                                                                                                                                                                                          |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| PRONZATO, LUC AND METELKINA, ASYA. Information-regret compromise in covariate-adaptive treatment allocation . . . . .                                                                    | 2046–2073 |
| QU, ANNIE AND SHI, PEIBEI. Weak signal identification and inference in penalized model selection . . . . .                                                                               | 1214–1253 |
| RAKHLIN, ALEXANDER, CAI, T. TONY AND LIANG, TENGYUAN. Computational and statistical boundaries for submatrix localization in a large noisy matrix . . . . .                              | 1403–1430 |
| RAY, KOLYAN. Adaptive Bernstein–von Mises theorems in Gaussian white noise . . . . .                                                                                                     | 2511–2536 |
| RELATOR, RAISSA, SESE, JUN, SUGIYAMA, MASASHI, FUKUMIZU, KENJI, LIU, SONG AND SUZUKI, TAIJI. Support consistency of direct sparse-change learning in Markov networks . . . . .           | 959–990   |
| RICHARDSON, THOMAS S., NANDY, PREETAM AND MAATHUIS, MARLOES H. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings . . . . .        | 647–674   |
| ROBINS, JAMES M., LI, LINGLING, MUKHERJEE, RAJARSHI, TCHETGEN, ERIC TCHETGEN AND VAN DER VAART, AAD. Minimax estimation of a functional on a structured high-dimensional model . . . . . | 1951–1987 |
| ROUSSEAU, JUDITH AND SZABO, BOTOND. Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator . . . . .                                 | 833–865   |
| RUIZ-GAZEN, ANNE, BOISTARD, HÉLÈNE AND LOPUHAÄ, HENDRIK P. Functional central limit theorems for single-stage sampling designs . . . . .                                                 | 1728–1758 |
| SADEGHI, KAYVAN, UHLER, CAROLINE, WERMUTH, NANNY, ZWIERNIK, PIOTR, FALLAT, SHAUN AND LAURITZEN, STEFFEN. Total positivity in Markov structures . . . . .                                 | 1152–1184 |
| SCHORNING, KIRSTEN, DETTE, HOLGER, BERMAN, GEORGINA, BORNKAMP, BJÖRN AND FELLER, CHRYSTEL. Optimal designs for dose response curves with common parameters . . . . .                     | 2102–2132 |
| SCHWARTZMAN, ARMIN AND CHENG, DAN. Multiple testing of local maxima for detection of peaks in random fields . . . . .                                                                    | 529–556   |
| SESE, JUN, SUGIYAMA, MASASHI, FUKUMIZU, KENJI, LIU, SONG, SUZUKI, TAIJI AND RELATOR, RAISSA. Support consistency of direct sparse-change learning in Markov networks . . . . .           | 959–990   |
| SHEPP, LARRY A., WYNER, ABRAHAM J. AND ERNST, PHILIP A. Yule’s “nonsense correlation” solved! . . . . .                                                                                  | 1789–1809 |
| SHI, PEIBEI AND QU, ANNIE. Weak signal identification and inference in penalized model selection . . . . .                                                                               | 1214–1253 |

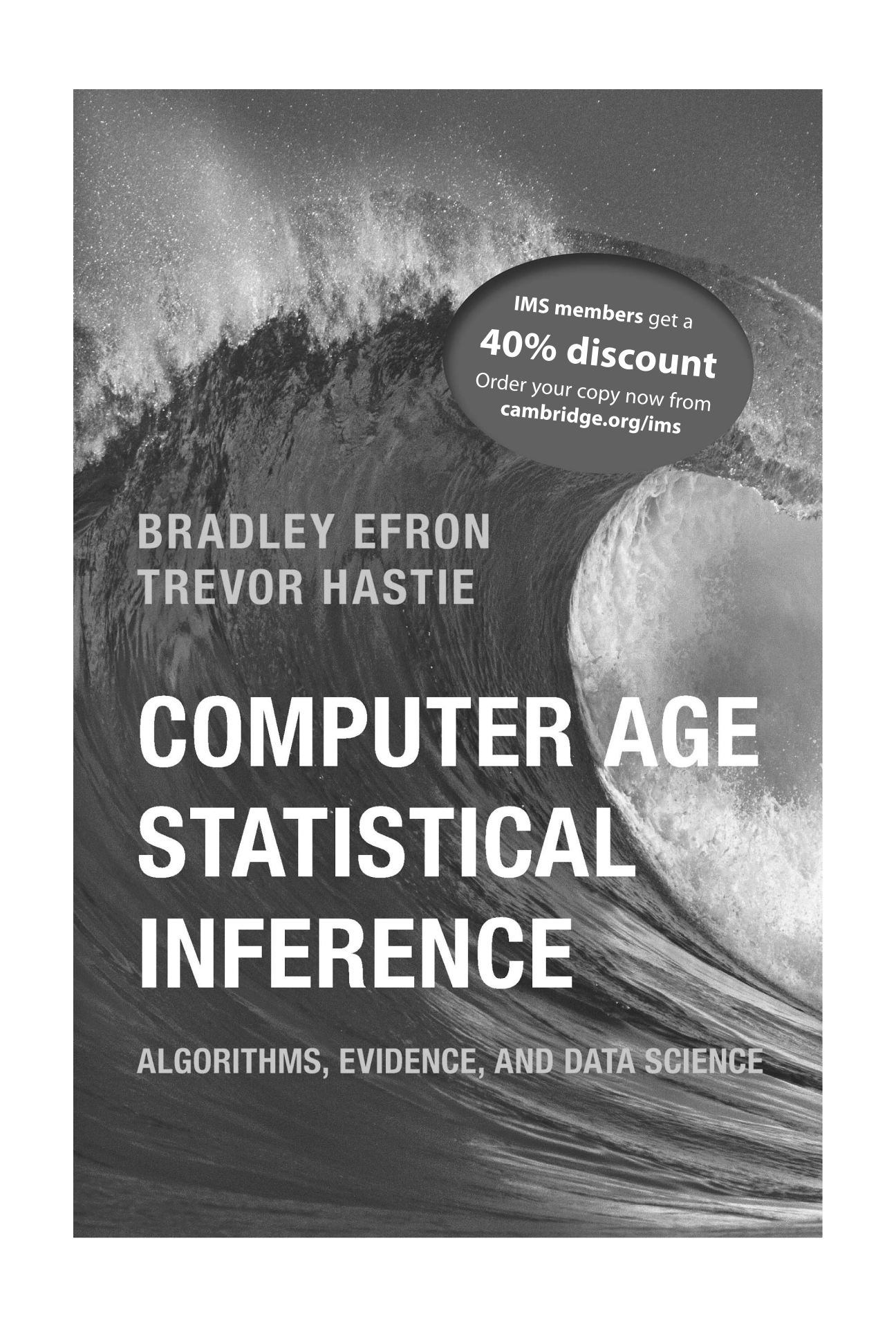
|                                                                                                                                                                                      |           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| SÖHL, JAKOB AND NICKL, RICHARD. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions .....                                                   | 1664–1693 |
| SONG, JUN AND LI, BING. Nonlinear sufficient dimension reduction for functional data .....                                                                                           | 1059–1095 |
| SONG, PETER X. K., TSAI, CHIH-LING, ZHONG, PING-SHOU AND LAN, WEI. Tests for covariance structures with high-dimensional repeated measurements .....                                 | 1185–1213 |
| STEINWART, INGO AND HANG, HANYUAN. A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning .....                                 | 708–743   |
| SU, WEIJIE, BOGDAN, MAŁGORZATA AND CANDÈS, EMMANUEL. False discoveries occur early on the Lasso path .....                                                                           | 2133–2150 |
| SU, ZHIHUA, KHARE, KSHITIJ AND PAL, SUBHADIP. A Bayesian approach for envelope models .....                                                                                          | 196–222   |
| SUGIYAMA, MASASHI, FUKUMIZU, KENJI, LIU, SONG, SUZUKI, TAJI, RELATOR, RAISSA AND SESE, JUN. Support consistency of direct sparse-change learning in Markov networks ....             | 959–990   |
| SUZUKI, TAJI, RELATOR, RAISSA, SESE, JUN, SUGIYAMA, MASASHI, FUKUMIZU, KENJI AND LIU, SONG. Support consistency of direct sparse-change learning in Markov networks .....            | 959–990   |
| SZABO, BOTOND AND ROUSSEAU, JUDITH. Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator .....                                 | 833–865   |
| TANG, CHUAN-FA, WANG, DEWEI AND TEBBS, JOSHUA M. Nonparametric goodness-of-fit tests for uniform stochastic ordering .....                                                           | 2565–2589 |
| TAQQU, MURAD S. AND BAI, SHUYANG. On the validity of resampling methods under long memory .....                                                                                      | 2365–2399 |
| TAVAKOLI, SHAHIN, ASTON, JOHN A. D. AND PIGOLI, DAVIDE. Tests for separability in nonparametric covariance operators of random surfaces .....                                        | 1431–1461 |
| TAYLOR, JONATHAN, TIBSHIRANI, ROBERT AND CHOI, YUNJIN. Selecting the number of principal components: Estimation of the true rank of a noisy matrix .....                             | 2590–2617 |
| TCHETGEN, ERIC TCHETGEN, VAN DER VAART, AAD, ROBINS, JAMES M., LI, LINGLING AND MUKHERJEE, RAJARSHI. Minimax estimation of a functional on a structured high-dimensional model ..... | 1951–1987 |

|                                                                                                                                                                                      |           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| TEBBS, JOSHUA M., TANG, CHUAN-FA AND WANG, DEWEI.<br>Nonparametric goodness-of-fit tests for uniform stochastic ordering .....                                                       | 2565–2589 |
| TIBSHIRANI, ROBERT, CHOI, YUNJIN AND TAYLOR, JONATHAN.<br>Selecting the number of principal components: Estimation of the true rank of a noisy matrix .....                          | 2590–2617 |
| TODOROV, VIKTOR. Testing for time-varying jump activity for pure jump semimartingales .....                                                                                          | 1284–1311 |
| TOULIS, PANOS AND AIROLDI, EDOARDO M. Asymptotic and finite-sample properties of estimators based on stochastic gradients .....                                                      | 1694–1727 |
| TSAI, CHIH-LING, ZHONG, PING-SHOU, LAN, WEI AND SONG, PETER X. K. Tests for covariance structures with high-dimensional repeated measurements .....                                  | 1185–1213 |
| TSUNG, FUGEE, XIAN, JINGUO AND HAN, DONG. On the optimality of Bayesian change-point detection .....                                                                                 | 1375–1402 |
| TSYBAKOV, ALEXANDRE B., COLLIER, OLIVIER AND COM-MINGES, LAËTITIA. Minimax estimation of linear and quadratic functionals on sparsity classes .....                                  | 923–958   |
| TSYBAKOV, ALEXANDRE B., VERZELEN, NICOLAS AND KLOPP, OLGA. Oracle inequalities for network models and sparse graphon estimation .....                                                | 316–354   |
| UHLER, CAROLINE, WERMUTH, NANNY, ZWIERNIK, PIOTR, FAL-LAT, SHAUN, LAURITZEN, STEFFEN AND SADEGHI, KAY-VAN. Total positivity in Markov structures .....                               | 1152–1184 |
| VALIANT, GREGORY AND KONG, WEIHAO. Spectrum estimation from samples .....                                                                                                            | 2218–2247 |
| VAN DER LAAN, MARK J., CHAMBAZ, ANTOINE AND ZHENG, WENJING. Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward .....       | 2537–2564 |
| VAN DER VAART, AAD, ROBINS, JAMES M., LI, LINGLING, MUKHERJEE, RAJARSHI AND TCHETGEN, ERIC TCHETGEN. Minimax estimation of a functional on a structured high-dimensional model ..... | 1951–1987 |
| VERDEBOUT, THOMAS, CUTTING, CHRISTINE AND PAINDAVEINE, DAVY. Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives .....               | 1024–1058 |
| VERDEBOUT, THOMAS AND PAINDAVEINE, DAVY. Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities .....                      | 800–832   |
| VERZELEN, NICOLAS AND ARIAS-CASTRO, ERY. Detection and feature selection in sparse mixture models .....                                                                              | 1920–1950 |

|                                                                                                                                                         |           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| VERZELEN, NICOLAS, KLOPP, OLGA AND TSYBAKOV, ALEXANDRE B. Oracle inequalities for network models and sparse graphon estimation .....                    | 316–354   |
| VETTER, MATHIAS, BIBINGER, MARKUS AND JIRAK, MORITZ. Nonparametric change-point analysis of volatility .....                                            | 1542–1578 |
| WAINWRIGHT, MARTIN J. AND LOH, PO-LING. Support recovery without incoherence: A case for nonconvex regularization .....                                 | 2455–2482 |
| WAINWRIGHT, MARTIN J., YANG, YUN AND PILANCI, MERT. Randomized sketches for kernels: Fast and optimal nonparametric regression .....                    | 991–1023  |
| WAINWRIGHT, MARTIN J., YU, BIN AND BALAKRISHNAN, SIVARAMAN. Statistical guarantees for the EM algorithm: From population to sample-based analysis ..... | 77–120    |
| WANG, DEWEI, TEBBS, JOSHUA M. AND TANG, CHUAN-FA. Nonparametric goodness-of-fit tests for uniform stochastic ordering .....                             | 2565–2589 |
| WANG, HANSHENG, ZHU, XUENING, PAN, RUI, LI, GUODONG AND LIU, YUEWEN. Network vector autoregression .....                                                | 1096–1123 |
| WANG, JINGSHU, ZHAO, QINGYUAN, HASTIE, TREVOR AND OWEN, ART B. Confounder adjustment in multiple hypothesis testing .....                               | 1863–1894 |
| WANG, QINWEN AND YAO, JIANFENG. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application .....                                  | 415–460   |
| WANG, QINWEN, YAO, JIANFENG AND LI, ZENG. Identifying the number of factors from singular values of a large sample autocovariance matrix .....          | 257–288   |
| WANG, WANJIE, JIN, JIASHUN AND KE, ZHENG TRACY. Phase transitions for high dimensional clustering and related problems                                  | 2151–2189 |
| WANG, WEICHEN AND FAN, JIANQING. Asymptotics of empirical eigenstructure for high dimensional spiked covariance .....                                   | 1342–1374 |
| WANG, Y. X. RACHEL AND BICKEL, PETER J. Likelihood-based model selection for stochastic block models .....                                              | 500–528   |
| WERMUTH, NANNY, ZWIERNIK, PIOTR, FALLAT, SHAUN, LAURITZEN, STEFFEN, SADEGHI, KAYVAN AND UHLER, CAROLINE. Total positivity in Markov structures .....    | 1152–1184 |
| WIENS, DOUGLAS P. AND HU, RUI. Robust discrimination designs over Hellinger neighbourhoods .....                                                        | 1638–1663 |
| WU, WEI BIAO AND ZHANG, DANNA. Gaussian approximation for high dimensional time series .....                                                            | 1895–1919 |
| WYNER, ABRAHAM J., ERNST, PHILIP A. AND SHEPP, LARRY A. Yule’s “nonsense correlation” solved! .....                                                     | 1789–1809 |

|                                                                                                                                                                                    |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| XIAN, JINGUO, HAN, DONG AND TSUNG, FUGEE. On the optimality of Bayesian change-point detection . . . . .                                                                           | 1375–1402 |
| XU, GONGJUN. Identifiability of restricted latent class models with binary responses . . . . .                                                                                     | 675–707   |
| XU, GONGJUN, JIANG, TIEFENG AND LEDER, KEVIN. Rare-event analysis for extremal eigenvalues of white Wishart matrices . . .                                                         | 1609–1637 |
| YANG, YUN, PILANCI, MERT AND WAINWRIGHT, MARTIN J. Randomized sketches for kernels: Fast and optimal nonparametric regression . . . . .                                            | 991–1023  |
| YAO, JIANFENG, LI, ZENG AND WANG, QINWEN. Identifying the number of factors from singular values of a large sample autocovariance matrix . . . . .                                 | 257–288   |
| YAO, JIANFENG AND WANG, QINWEN. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application . . . .                                                           | 415–460   |
| YU, BIN, BALAKRISHNAN, SIVARAMAN AND WAINWRIGHT, MARTIN J. Statistical guarantees for the EM algorithm: From population to sample-based analysis . . . . .                         | 77–120    |
| ZHANG, DANNA AND WU, WEI BIAO. Gaussian approximation for high dimensional time series . . . . .                                                                                   | 1895–1919 |
| ZHAO, QINGYUAN, HASTIE, TREVOR, OWEN, ART B. AND WANG, JINGSHU. Confounder adjustment in multiple hypothesis testing . . . . .                                                     | 1863–1894 |
| ZHAO, TIANQI, LIU, HAN AND NING, YANG. A likelihood ratio framework for high-dimensional semiparametric regression . . .                                                           | 2299–2327 |
| ZHENG, WEI, AI, MINGYAO AND LI, KANG. Identification of universally optimal circular designs for the interference model . . .                                                      | 1462–1487 |
| ZHENG, WENJING, VAN DER LAAN, MARK J. AND CHAMBAZ, ANTOINE. Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward . . . . . | 2537–2564 |
| ZHIGLJAVSKY, ANATOLY, DETTE, HOLGER AND KONSTANTINOU, MARIA. A new approach to optimal designs for correlated observations . . . . .                                               | 1579–1608 |
| ZHONG, PING-SHOU, LAN, WEI, SONG, PETER X. K. AND TSAI, CHIH-LING. Tests for covariance structures with high-dimensional repeated measurements . . . . .                           | 1185–1213 |
| ZHONG, PING-SHOU AND LI, JUN. A rate optimal procedure for recovering sparse differences between high-dimensional means under dependence . . . . .                                 | 557–590   |
| ZHOU, HARRISON H., GAO, CHAO AND MA, ZONGMING. Sparse CCA: Adaptive estimation and computational barriers . . . . .                                                                | 2074–2101 |

|                                                                                                                                                                |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| ZHU, XUENING, PAN, RUI, LI, GUODONG, LIU, YUEWEN AND<br>WANG, HANSHENG. Network vector autoregression . . . . .                                                | 1096–1123 |
| ZHU, YING. Nonasymptotic analysis of semiparametric regression<br>models with high-dimensional parametric coefficients . . . . .                               | 2274–2298 |
| ZOHREN, STEFAN, ANEVSKI, DRAGI AND GILL, RICHARD D.<br>Estimating a probability mass function with unknown labels . . .                                        | 2708–2735 |
| ZOU, HUI AND DATTA, ABHIRUP. CoCoLasso for high-<br>dimensional error-in-variables regression . . . . .                                                        | 2400–2426 |
| ZWIERNIK, PIOTR, FALLAT, SHAUN, LAURITZEN, STEFFEN,<br>SADEGHI, KAYVAN, UHLER, CAROLINE AND WERMUTH,<br>NANNY. Total positivity in Markov structures . . . . . | 1152–1184 |



IMS members get a  
**40% discount**

Order your copy now from  
[cambridge.org/ims](http://cambridge.org/ims)

BRADLEY EFRON  
TREVOR HASTIE

# COMPUTER AGE STATISTICAL INFERENCE

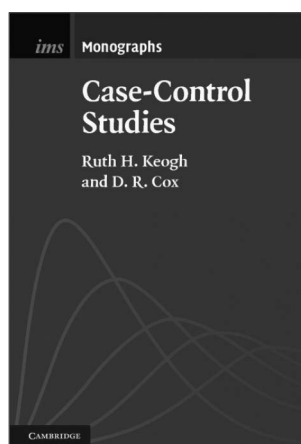
ALGORITHMS, EVIDENCE, AND DATA SCIENCE





*The Institute of Mathematical Statistics presents*

# IMS MONOGRAPHS



## **Case-Control Studies**

Ruth H. Keogh  
and D. R. Cox

The case-control approach is a powerful method for investigating factors that may explain a particular event. It is extensively used in epidemiology to study disease incidence, one of the best-known examples being Bradford Hill and Doll's investigation of the possible connection between cigarette smoking and lung cancer. More recently, case-control studies have been increasingly used in other fields, including sociology and econometrics.

With a particular focus on statistical analysis, this book is ideal for applied and theoretical statisticians wanting an up-to-date introduction to the field. It covers the fundamentals of case-control study design and analysis as well as more recent developments, including two-stage studies, case-only studies and methods for case-control sampling in time. The latter have important applications in large prospective cohorts which require case-control sampling designs to make efficient use of resources. More theoretical background is provided in an appendix for those new to the field.

**IMS member? Claim  
your 40% discount:  
[www.cambridge.org/ims](http://www.cambridge.org/ims)**

**Hardback price  
US\$48.00  
(non-member price  
\$80.00)**

---

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.