

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

- Approximating faces of marginal polytopes in discrete hierarchical models
NANWEI WANG, JOHANNES RAUH AND HÉLÈNE MASSAM 1203
- CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality T. TONY CAI, JING MA AND LINJUN ZHANG 1234
- Exponential ergodicity of the bouncy particle sampler GEORGE DELIGIANNIDIS, ALEXANDRE BOUCHARD-CÔTÉ AND ARNAUD DOUCET 1268
- The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data
JORIS BIERKENS, PAUL FEARNHEAD AND GARETH ROBERTS 1288
- Estimation of large covariance and precision matrices from temporally dependent observations HAI SHU AND BIN NAN 1321
- Bootstrap tuning in Gaussian ordered model selection
VLADIMIR SPOKOINY AND NIKLAS WILLRICH 1351
- Sequential change-point detection based on nearest neighbors HAO CHEN 1381
- Prediction when fitting simple models to high-dimensional data
LUKAS STEINBERGER AND HANNES LEEB 1408
- Two-sample and ANOVA tests for high dimensional means
SONG XI CHEN, JUN LI AND PING-SHOU ZHONG 1443
- Valid confidence intervals for post-model-selection predictors
FRANÇOIS BACHOC, HANNES LEEB AND BENEDIKT M. PÖTSCHER 1475
- A robust and efficient approach to causal inference based on sparse sufficient dimension reduction SHUJIE MA, LIPING ZHU, ZHIWEI ZHANG, CHIH-LING TSAI AND RAYMOND J. CARROLL 1505
- The maximum likelihood threshold of a path diagram MATHIAS DRTON, CHRISTOPHER FOX, ANDREAS KÄUFL AND GUILLAUME POULIOT 1536
- Convex regularization for high-dimensional multiresponse tensor regression
GARVESH RASKUTTI, MING YUAN AND HAN CHEN 1554
- Large sample theory for merged data from multiple sources TAKUMI SAEGUSA 1585
- Khinchine's theorem and Edgeworth approximations for weighted sums
SERGEY G. BOBKOV 1616
- Distributed inference for quantile regression processes
STANISLAV VOLGUSHEV, SHIH-KANG CHAO AND GUANG CHENG 1634
- Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data YUTA KOIKE 1663
- Causal Dantzig: Fast inference in linear structural equation models with hidden variables under additive interventions DOMINIK ROTHENHÄUSLER, PETER BÜHLMANN AND NICOLAI MEINSHAUSEN 1688
- Nonpenalized variable selection in high-dimensional linear model settings via generalized fiducial inference JONATHAN P. WILLIAMS AND JAN HANNIG 1723
- Super-resolution estimation of cyclic arrival rates
NINGYUAN CHEN, DONALD K. K. LEE AND SAHAND N. NEGAHBAN 1754
- Sequential multiple testing with generalized error control: An asymptotic optimality theory YANGLEI SONG AND GEORGIOS FELLOURIS 1776

THE ANNALS OF STATISTICS

Vol. 47, No. 3, pp. 1203–1803 June 2019

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: Xiao-Li Meng, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

President-Elect: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

Past President: Alison Etheridge, Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

The Annals of Statistics. *Editors:* Richard J. Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK. Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027, USA

The Annals of Applied Statistics. *Editor-in-Chief:* Karen Kafadar, Department of Statistics, University of Virginia, Heidelberg Institute for Theoretical Studies, Charlottesville, VA 22904-4135, USA

The Annals of Probability. *Editor:* Amir Dembo, Department of Statistics and Department of Mathematics, Stanford University, Stanford, California 94305, USA

The Annals of Applied Probability. *Editors:* François Delarue, Laboratoire J. A. Dieudonné, Université de Nice Sophia-Antipolis, France-06108 Nice Cedex 2. Peter Friz, Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany and Weierstrass-Institut für Angewandte Analysis und Stochastik, 10117 Berlin, Germany

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

The IMS Bulletin. *Editor:* Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

The Annals of Statistics [ISSN 0090-5364 (print); ISSN 2168-8966 (online)], Volume 47, Number 3, June 2019. 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.

APPROXIMATING FACES OF MARGINAL POLYTOPES IN DISCRETE HIERARCHICAL MODELS

BY NANWEI WANG*, JOHANNES RAUH*,[†] AND HÉLÈNE MASSAM*,¹

*York University** and *Max Planck Institute for Mathematics in the Sciences*[†]

The existence of the maximum likelihood estimate in a hierarchical log-linear model is crucial to the reliability of inference for this model. Determining whether the estimate exists is equivalent to finding whether the sufficient statistics vector t belongs to the boundary of the marginal polytope of the model. The dimension of the smallest face F_t containing t determines the dimension of the reduced model which should be considered for correct inference. For higher-dimensional problems, it is not possible to compute F_t exactly. Massam and Wang (2015) found an outer approximation to F_t using a collection of submodels of the original model. This paper refines the methodology to find an outer approximation and devises a new methodology to find an inner approximation. The inner approximation is given not in terms of a face of the marginal polytope, but in terms of a subset of the vertices of F_t .

Knowing F_t exactly indicates which cell probabilities have maximum likelihood estimates equal to 0. When F_t cannot be obtained exactly, we can use, first, the outer approximation F_2 to reduce the dimension of the problem and then the inner approximation F_1 to obtain correct estimates of cell probabilities corresponding to elements of F_1 and improve the estimates of the remaining probabilities corresponding to elements in $F_2 \setminus F_1$. Using both real-world and simulated data, we illustrate our results, and show that our methodology scales to high dimensions.

REFERENCES

- BANERJEE, O., EL GHAOU, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. [MR2417243](#)
- BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. [MR0489333](#)
- CSISZÁR, I. and MATÚŠ, F. (2008). Generalized maximum likelihood estimates for exponential families. *Probab. Theory Related Fields* **141** 213–246. [MR2372970](#)
- CSISZÁR, I. and SHIELDS, P. (2004). *Information Theory and Statistics: A Tutorial*, 1st ed. Now Publishers, Hanover, MA.
- DEZA, M. M. and LAURENT, M. (2010). *Geometry of Cuts and Metrics. Algorithms and Combinatorics* **15**. Springer, Heidelberg. [MR2841334](#)
- DOBRA, A., EROSHEVA, E. A. and FIENBERG, S. E. (2004). Disclosure limitation methods based on bounds for large contingency tables with applications to disability. In *Statistical Data Mining and Knowledge Discovery* 93–116. Chapman & Hall, Boca Raton, FL. [MR2048950](#)

MSC2010 subject classifications. 62H12, 62F10, 62F99.

Key words and phrases. Existence of the maximum likelihood estimate, marginal polytope, faces, facial sets, extended maximum likelihood estimate.

- DOBRA, A. and LENKOSKI, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *Ann. Appl. Stat.* **5** 969–993. [MR2840183](#)
- ERIKSSON, N., FIENBERG, S. E., RINALDO, A. and SULLIVANT, S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *J. Symbolic Comput.* **41** 222–233. [MR2197157](#)
- FIENBERG, S. E. (1980). *The Analysis of Cross-Classified Categorical Data*, 2nd ed. MIT Press, Cambridge, MA. [MR0623082](#)
- FIENBERG, S. E. and RINALDO, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *J. Statist. Plann. Inference* **137** 3430–3445. [MR2363267](#)
- FIENBERG, S. E. and RINALDO, A. (2012). Maximum likelihood estimation in log-linear models. *Ann. Statist.* **40** 996–1023. [MR2985941](#)
- GAWRILOW, E. and JOSWIG, M. (2000). Polymake: A framework for analyzing convex polytopes. In *Polytopes—Combinatorics and Computation (Oberwolfach, 1997)*. *DMV Sem.* **29** 43–73. Birkhäuser, Basel. [MR1785292](#)
- GEYER, C. J. (2009). Likelihood inference in exponential families and directions of recession. *Electron. J. Stat.* **3** 259–289. [MR2495839](#)
- HABERMAN, S. J. (1974). *The Analysis of Frequency Data*. The Univ. Chicago Press, Chicago, IL. [MR0408098](#)
- KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β -model and synthetic graphs. *Ann. Statist.* **44** 87–112. [MR3449763](#)
- LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. [MR1419991](#)
- LETAC, G. and MASSAM, H. (2012). Bayes factors and the geometry of discrete hierarchical log-linear models. *Ann. Statist.* **40** 861–890. [MR2985936](#)
- LIU, Q. and IHLER, A. (2012). Distributed parameter estimation via pseudo-likelihood. *Int. Conf. Mach. Learn. (ICML)*.
- MASSAM, H. and WANG, N. (2015). A local approach to estimation in discrete loglinear models. Preprint. Available at [arXiv:1504.05434](#).
- MASSAM, H. and WANG, N. (2018). Local conditional and marginal approach to parameter estimation in discrete graphical models. *J. Multivariate Anal.* **164** 1–21. [MR3738130](#)
- RAUH, J., KAHLE, T. and AY, N. (2011). Support sets in exponential families and oriented matroid theory. *Internat. J. Approx. Reason.* **52** 613–626. [MR2787021](#)
- RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1 -regularized logistic regression. *Ann. Statist.* **38** 1287–1319. [MR2662343](#)
- SCHMIDT, M. (2005). minFunc: Unconstrained differentiable multivariate optimization in Matlab. <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>.
- VLACH, M. (1986). Conditions for the existence of solutions of the three-dimensional planar transportation problem. *Discrete Appl. Math.* **13** 61–78. [MR0829339](#)
- WANG, N., RAUH J. and MASSAM, H. (2019). Supplement to “Approximating faces of marginal polytopes in discrete hierarchical models.” DOI:[10.1214/18-AOS1710SUPP](#).
- ZIEGLER, G. M. (1995). *Lectures on Polytopes*. *Graduate Texts in Mathematics* **152**. Springer, New York. [MR1311028](#)

CHIME: CLUSTERING OF HIGH-DIMENSIONAL GAUSSIAN MIXTURES WITH EM ALGORITHM AND ITS OPTIMALITY¹

BY T. TONY CAI, JING MA AND LINJUN ZHANG

University of Pennsylvania

Unsupervised learning is an important problem in statistics and machine learning with a wide range of applications. In this paper, we study clustering of high-dimensional Gaussian mixtures and propose a procedure, called CHIME, that is based on the EM algorithm and a direct estimation method for the sparse discriminant vector. Both theoretical and numerical properties of CHIME are investigated. We establish the optimal rate of convergence for the excess misclustering error and show that CHIME is minimax rate optimal. In addition, the optimality of the proposed estimator of the discriminant vector is also established. Simulation studies show that CHIME outperforms the existing methods under a variety of settings. The proposed CHIME procedure is also illustrated in an analysis of a glioblastoma gene expression data set and shown to have superior performance.

Clustering of Gaussian mixtures in the conventional low-dimensional setting is also considered. The technical tools developed for the high-dimensional setting are used to establish the optimality of the clustering procedure that is based on the classical EM algorithm.

REFERENCES

- [1] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- [2] AZIZYAN, M., SINGH, A. and WASSERMAN, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. In *NIPS* 2139–2147.
- [3] AZIZYAN, M., SINGH, A. and WASSERMAN, L. A. (2015). Efficient sparse clustering of high-dimensional non-spherical Gaussian mixtures. In *AISTATS*.
- [4] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. [MR3611487](#)
- [5] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. [MR2485008](#)
- [6] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York. [MR2247587](#)
- [7] BOUYEYRON, C. and BRUNET-SAUMARD, C. (2014). Model-based clustering of high-dimensional data: A review. *Comput. Statist. Data Anal.* **71** 52–78. [MR3131954](#)
- [8] BRADLEY, P. S., FAYYAD, U. M. and MANGASARIAN, O. L. (1999). Mathematical programming for data mining: Formulations and challenges. *INFORMS J. Comput.* **11** 217–238. [MR1704656](#)

MSC2010 subject classifications. Primary 62G15; secondary 62C20, 62H35.

Key words and phrases. High-dimensional data, unsupervised learning, Gaussian mixture model, EM algorithm, misclustering error, Minimax optimality.

- [9] CAI, T. and LIU, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *J. Amer. Statist. Assoc.* **106** 1566–1577. [MR2896857](#)
- [10] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- [11] CAI, T. T., MA, J. and ZHANG, L. (2019). Supplement to “CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality.” DOI:[10.1214/18-AOS1711SUPP](#).
- [12] CAI, T. T. and ZHANG, L. (2018). High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Statist. Sinica* **28** 963–993. [MR3791096](#)
- [13] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. [MR0501537](#)
- [14] DUDA, R. O. and HART, P. E. (1973). *Pattern Classification and Scene Analysis*, Vol. 3. Wiley, New York.
- [15] EVERITT, B. S. (1981). *Finite Mixture Distributions*. Wiley, New York.
- [16] FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. [MR1951635](#)
- [17] GE, R., HUANG, Q. and KAKADE, S. M. (2015). Learning mixtures of Gaussians in high dimensions [extended abstract]. In *STOC’15—Proceedings of the 2015 ACM Symposium on Theory of Computing* 761–770. ACM, New York. [MR3388256](#)
- [18] HARDT, M. and PRICE, E. (2015). Tight bounds for learning a mixture of two Gaussians [extended abstract]. In *STOC’15—Proceedings of the 2015 ACM Symposium on Theory of Computing* 753–760. ACM, New York. [MR3388255](#)
- [19] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. [MR2722294](#)
- [20] JIN, C., ZHANG, Y., BALAKRISHNAN, S., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). Local maxima in the likelihood of Gaussian mixture models: Structural results and algorithmic consequences. In *NIPS* 4116–4124.
- [21] JIN, J., KE, Z. T. and WANG, W. (2017). Phase transitions for high dimensional clustering and related problems. *Ann. Statist.* **45** 2151–2189. [MR3718165](#)
- [22] JIN, J. and WANG, W. (2016). Influential features PCA for high dimensional clustering. *Ann. Statist.* **44** 2323–2359. [MR3576543](#)
- [23] LINDSAY, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics **5**, Inst. Math. Statist., Hayward, CA; Amer. Statist. Assoc., Alexandria, VA.
- [24] MAI, Q., YANG, Y. and ZOU, H. (2018). Multiclass sparse discriminant analysis. *Statist. Sinica*. To appear.
- [25] MAI, Q., ZOU, H. and YUAN, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42. [MR2899661](#)
- [26] MOITRA, A. and VALIANT, G. (2010). Settling the polynomial learnability of mixtures of Gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010* 93–102. IEEE Comput. Soc., Los Alamitos, CA. [MR3024779](#)
- [27] NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2015). A unified theory of confidence regions and testing for high dimensional estimating equations. Preprint. Available at [arXiv:1510.08986](#).
- [28] PEARSON, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. Ser. A* **185** 71–110.
- [29] REDNER, R. A. and WALKER, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** 195–239. [MR0738930](#)
- [30] REYNOLDS, D. (2015). Gaussian mixture models. In *Encyclopedia of Biometrics* 827–832. Springer, New York.

- [31] SCOTT, A. J. and SYMONS, M. J. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics* **27** 387–397.
- [32] TIAN, L. and GU, Q. (2017). Communication-efficient distributed sparse linear discriminant analysis. In *AISTATS. Proceedings of Machine Learning Research* **54** 1178–1187.
- [33] TIBSHIRANI, R. and WALTHER, G. (2005). Cluster validation by prediction strength. *J. Comput. Graph. Statist.* **14** 511–528. [MR2170199](#)
- [34] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. Revised and extended from the 2004 French original, translated by Vladimir Zaiats. [MR2724359](#)
- [35] VERHAAK, R. G. W., HOADLEY, K. A., PURDOM, E., WANG, V., QI, Y., WILKERSON, M. D., MILLER, C. R., DING, L., GOLUB, T., MESIROV, J. P. et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17** 98–110.
- [36] WANG, Z., GU, Q., NING, Y. and LIU, H. (2015). High dimensional expectation-maximization algorithm: Statistical optimization and asymptotic normality. In *NIPS* 2521–2529.
- [37] WARD, J. H. JR. (1963). Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* **58** 236–244. [MR0148188](#)
- [38] WITTEN, D. M. and TIBSHIRANI, R. (2010). A framework for feature selection in clustering. *J. Amer. Statist. Assoc.* **105** 713–726. [MR2724855](#)
- [39] YI, X. and CARAMANIS, C. (2015). Regularized EM algorithms: A unified framework and statistical guarantees. In *NIPS* 1567–1575.
- [40] ZHOU, H., PAN, W. and SHEN, X. (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electron. J. Stat.* **3** 1473–1496. [MR2578834](#)

EXPONENTIAL ERGODICITY OF THE BOUNCY PARTICLE SAMPLER

BY GEORGE DELIGIANNIDIS*, ALEXANDRE BOUCHARD-CÔTÉ^{†,1} AND ARNAUD DOUCET*,²

*University of Oxford** and *University of British Columbia*[†]

Nonreversible Markov chain Monte Carlo schemes based on piecewise deterministic Markov processes have been recently introduced in applied probability, automatic control, physics and statistics. Although these algorithms demonstrate experimentally good performance and are accordingly increasingly used in a wide range of applications, geometric ergodicity results for such schemes have only been established so far under very restrictive assumptions. We give here verifiable conditions on the target distribution under which the Bouncy Particle Sampler algorithm introduced in [*Phys. Rev. E* **85** (2012) 026703, 1671–1691] is geometrically ergodic and we provide a central limit theorem for the associated ergodic averages. This holds essentially whenever the target satisfies a curvature condition and the growth of the negative logarithm of the target is at least linear and at most quadratic. For target distributions with thinner tails, we propose an original modification of this scheme that is geometrically ergodic. For targets with thicker tails, we extend the idea pioneered in [*Ann. Statist.* **40** (2012) 3050–3076] in a random walk Metropolis context. We establish geometric ergodicity of the Bouncy Particle Sampler with respect to an appropriate transformation of the target. Mapping the resulting process back to the original parameterization, we obtain a geometrically ergodic piecewise deterministic Markov process.

REFERENCES

- [1] BIERKENS, J., BOUCHARD-CÔTÉ, A., DOUCET, A., DUNCAN, A. B., FEARNHEAD, P., LIENART, T., ROBERTS, G. O. and VOLLMER, S. J. (2018). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Probab. Lett.* **136** 148–154.
- [2] BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825.
- [3] BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. O. (2016). The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. Available at [arXiv:1607.03188](https://arxiv.org/abs/1607.03188).
- [4] BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. [MR3655855](https://arxiv.org/abs/1703.05855)
- [5] BIERKENS, J., ROBERTS, G. O. and ZITT, P. A. (2017). Ergodicity of the Zig-Zag process. Available at [arXiv:1712.09875](https://arxiv.org/abs/1712.09875).

MSC2010 subject classifications. 60J25, 65C05, 65C40, 60F05.

Key words and phrases. Piecewise deterministic Markov process, Markov chain Monte Carlo, geometric ergodicity, change of variable, central limit theorem.

- [6] BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* DOI:[10.1080/01621459.2017.1294075](https://doi.org/10.1080/01621459.2017.1294075). Available at [arXiv:1510.02451](https://arxiv.org/abs/1510.02451).
- [7] BURKE, C. J. and ROSENBLATT, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Stat.* **29** 1112–1122. [MR0101557](#)
- [8] COSTA, O. L. V. (1990). Stationary distributions for piecewise-deterministic Markov processes. *J. Appl. Probab.* **27** 60–73. [MR1039184](#)
- [9] COSTA, O. L. V. and DUFOUR, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **47** 1053–1077. [MR2385873](#)
- [10] DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388. [MR0790622](#)
- [11] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. [MR1283589](#)
- [12] DEL MORAL, P. and PENEV, S. (2017). *Stochastic Processes: From Applications to Theory*. CRC Press, Boca Raton, FL. [MR3618157](#)
- [13] DELIGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2019). Supplement to “Exponential ergodicity of the bouncy particle sampler.” DOI:[10.1214/18-AOS1714SUPP](https://doi.org/10.1214/18-AOS1714SUPP).
- [14] DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. [MR0836973](#)
- [15] DOWN, D., MEYN, S. P. and TWEEDIE, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. [MR1379163](#)
- [16] FEARNHED, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2016). Piecewise deterministic Markov processes for continuous-time Monte Carlo. Available at [arXiv:1611.07873](https://arxiv.org/abs/1611.07873).
- [17] FÉTIQUE, N. (2017). Long-time behaviour of generalised Zig-Zag process. Available at [arXiv:1710.01087](https://arxiv.org/abs/1710.01087).
- [18] FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. [MR3542627](#)
- [19] GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. [MR1404536](#)
- [20] GRAMACY, R. B. and POLSON, N. G. (2012). Simulation-based regularized logistic regression. *Bayesian Anal.* **7** 567–589. [MR2981628](#)
- [21] ISOBE, M. and KRAUTH, W. (2015). Hard-sphere melting and crystallization with event-chain Monte Carlo. *J. Chem. Phys.* **143** 084509.
- [22] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. [MR1731030](#)
- [23] JARNER, S. F. and TWEEDIE, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9** 559–578. [MR1996270](#)
- [24] JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. [MR3097969](#)
- [25] JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. [MR1888447](#)
- [26] KAPFER, S. C. and KRAUTH, W. (2016). Cell-veto Monte Carlo algorithms for long-range systems. *Phys. Rev. E* **94** 3 031302(R).
- [27] LIVINGSTONE, S., BETANCOURT, M., BYRNE, S. and GIROLAMI, M. (2016). On the geometric ergodicity of Hamiltonian Monte Carlo. Available at [arXiv:1601.08057](https://arxiv.org/abs/1601.08057).
- [28] MESQUITA, A. R. and HESPANHA, J. P. (2010). Construction of Lyapunov functions for piecewise-deterministic Markov processes. In *Proc. 49th IEEE Conf. Decision and Control* 2408–2413.

- [29] MESQUITA, A. R. and HESPANHA, J. P. (2012). Jump control of probability densities with applications to autonomous vehicle motion. *IEEE Trans. Automat. Control* **57** 2588–2598. [MR2991659](#)
- [30] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. [MR1234295](#)
- [31] MICHEL, M., KAPFER, S. C. and KRAUTH, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.* **140** 054116.
- [32] MONMARCHÉ, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 357–398. [MR3487077](#)
- [33] NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. [MR1994729](#)
- [34] NISHIKAWA, Y. and HUKUSHIMA, K. (2016). Event-chain Monte Carlo algorithm for continuous spin systems and its application. *Inst. Phys. Conf. Ser.* **750** 012014.
- [35] PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2017). Stochastic bouncy particle sampler. In *Proc. 34th Int. Conf. Machine Learning* 2741–2750.
- [36] PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85** 026703, 1671–1691.
- [37] ROBERTS, G. O. and ROSENTHAL, J. S. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.* **1** no. 9, approx. 21 pp. [MR1423462](#)
- [38] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. [MR1440273](#)
- [39] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. [MR1399158](#)
- [40] SHERLOCK, C. and THIERY, A. (2017). A discrete bouncy particle sampler. Available at [arXiv:1707.05200](#).
- [41] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise deterministic Markov chain Monte Carlo. Available at [arXiv:1707.05296](#).
- [42] WU, C. and ROBERT, C. P. (2017). Generalized bouncy particle sampler. Available at [arXiv:1706.04781](#).

THE ZIG-ZAG PROCESS AND SUPER-EFFICIENT SAMPLING FOR BAYESIAN ANALYSIS OF BIG DATA¹

BY JORIS BIERKENS, PAUL FEARNHEAD AND GARETH ROBERTS

Delft University of Technology, Lancaster University and University of Warwick

Standard MCMC methods can scale poorly to big data settings due to the need to evaluate the likelihood at each iteration. There have been a number of approximate MCMC algorithms that use sub-sampling ideas to reduce this computational burden, but with the drawback that these algorithms no longer target the true posterior distribution. We introduce a new family of Monte Carlo methods based upon a multidimensional version of the Zig-Zag process of [Ann. Appl. Probab. **27** (2017) 846–882], a continuous-time piecewise deterministic Markov process. While traditional MCMC methods are reversible by construction (a property which is known to inhibit rapid convergence) the Zig-Zag process offers a flexible nonreversible alternative which we observe to often have favourable convergence properties. We show how the Zig-Zag process can be simulated without discretisation error, and give conditions for the process to be ergodic. Most importantly, we introduce a sub-sampling version of the Zig-Zag process that is an example of an *exact approximate scheme*, that is, the resulting approximate process still has the posterior as its stationary distribution. Furthermore, if we use a control-variate idea to reduce the variance of our unbiased estimator, then the Zig-Zag process can be super-efficient: after an initial preprocessing step, essentially independent samples from the posterior distribution are obtained at a computational cost which does not depend on the size of the data.

REFERENCES

- ANDERSON, D. F. (2007). A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J. Chem. Phys.* **127** 214107.
- ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. [MR2502648](#)
- BARDENET, R., DOUCET, A. and HOLMES, C. (2017). On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.* **18** 1515–1557. [MR3670492](#)
- BIERKENS, J. (2016). Non-reversible Metropolis–Hastings. *Stat. Comput.* **26** 1213–1228. [MR3538633](#)
- BIERKENS, J. (2017). Computer experiments accompanying J. Bierkens, P. Fearnhead and G. Roberts, the Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. Available at <https://github.com/jbierkens/zigzag-experiments>. Date accessed: 20-10-2017.
- BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2018). Supplement to “The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data.” DOI:[10.1214/18-AOS1715SUPP](https://doi.org/10.1214/18-AOS1715SUPP).

MSC2010 subject classifications. Primary 65C60; secondary 65C05, 62F15, 60J25.

Key words and phrases. MCMC, nonreversible Markov process, piecewise deterministic Markov process, stochastic gradient Langevin dynamics, sub-sampling, exact sampling.

- BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. [MR3655855](#)
- BIERKENS, J., ROBERTS, G. O. and ZITT, P.-A. (2017). Ergodicity of the zigzag process. Preprint. Available at [arXiv:1712.09875](#).
- BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2017). The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* To appear. Available at [arXiv:1510.02451](#).
- CHEN, T.-L. and HWANG, C.-R. (2013). Accelerating reversible Markov chains. *Statist. Probab. Lett.* **83** 1956–1962. [MR3079029](#)
- DELIAGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2017). Exponential ergodicity of the bouncy particle sampler. Preprint. Available at [arXiv:1705.04579](#).
- DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222.
- DUBEY, K. A., REDDI, S. J., WILLIAMSON, S. A., POZOS, B., SMOLA, A. J. and XING, E. P. (2016). Variance reduction in stochastic gradient Langevin dynamics. In *Advances in Neural Information Processing Systems* 1154–1162.
- DUNCAN, A. B., LELIÈVRE, T. and PAVLIOTIS, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163** 457–491. [MR3483241](#)
- FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. [MR3843382](#)
- FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** 977–994. [MR3052846](#)
- FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. [MR3542627](#)
- GIBSON, M. A. and BRUCK, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104** 1876–1889.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. [MR3363437](#)
- HWANG, C.-R., HWANG-MA, S.-Y. and SHEU, S. J. (1993). Accelerating Gaussian diffusions. *Ann. Appl. Probab.* **3** 897–913. [MR1233633](#)
- JACOB, P. E. and THIERY, A. H. (2015). On nonnegative unbiased estimators. *Ann. Statist.* **43** 769–784. [MR3319143](#)
- JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Stat.* **41** 851–864. [MR0263198](#)
- LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413. [MR0546120](#)
- LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. [MR3694589](#)
- MA, Y.-A., CHEN, T. and FOX, E. (2015). A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems* 2917–2925.
- MACLAURIN, D. and ADAMS, R. P. (2014). Firefly Monte Carlo: Exact MCMC with subsets of data. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence* AUAI Press, Arlington, VA.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087.
- MONMARCHÉ, P. (2014). Hypocoercive relaxation to equilibrium for some kinetic models via a third order differential inequality. Available at [arXiv:1306.4548](#).
- NEAL, R. M. (1998). Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In *Learning in Graphical Models* 205–228. Springer, Berlin.

- NEISWANGER, W., WANG, C. and XING, E. (2013). Asymptotically exact, embarrassingly parallel MCMC. Available at [arXiv:1311.4780](#).
- PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2016). Stochastic bouncy particle sampler. Preprint. Available at [arXiv:1609.00770](#).
- PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* (3) **85** 1–5.
- POLLOCK, M., FEARNHEAD, P., JOHANSEN, A. M. and ROBERTS, G. O. (2016). The scalable Langevin exact algorithm: Bayesian inference for big data. Available at [arXiv:1609.03436](#).
- QUIROZ, M., VILLANI, M. and KOHN, R. (2015). Speeding up MCMC by efficient data subsampling. Riksbank Research Paper Series 121.
- REY-BELLET, L. and SPILIOPOULOS, K. (2015). Irreversible Langevin samplers and variance reduction: A large deviations approach. *Nonlinearity* **28** 2081–2103. [MR3366637](#)
- ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. [MR1440273](#)
- SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOGH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *Int. J. Manag. Sci. Eng. Manag.* **11** 78–88.
- TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D* **240** 410–414.
- VOLLMER, S. J., ZYGALAKIS, K. C. and TEH, Y. W. (2016). Exploration of the (non-)asymptotic bias and variance of stochastic gradient Langevin dynamics. *J. Mach. Learn. Res.* **17** 1–48. [MR3555050](#)
- WANG, X. and DUNSON, D. B. (2013). Parallelizing MCMC via Weierstrass sampler. Available at [arXiv:1312.4605](#).
- WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* 681–688.

ESTIMATION OF LARGE COVARIANCE AND PRECISION MATRICES FROM TEMPORALLY DEPENDENT OBSERVATIONS

BY HAI SHU¹ AND BIN NAN¹

University of Michigan and University of California, Irvine

We consider the estimation of large covariance and precision matrices from high-dimensional sub-Gaussian or heavier-tailed observations with slowly decaying temporal dependence. The temporal dependence is allowed to be long-range so with longer memory than those considered in the current literature. We show that several commonly used methods for independent observations can be applied to the temporally dependent data. In particular, the rates of convergence are obtained for the generalized thresholding estimation of covariance and correlation matrices, and for the constrained ℓ_1 minimization and the ℓ_1 penalized likelihood estimation of precision matrix. Properties of sparsistency and sign-consistency are also established. A gap-block cross-validation method is proposed for the tuning parameter selection, which performs well in simulations. As a motivating example, we study the brain functional connectivity using resting-state fMRI time series data with long-range temporal dependence.

REFERENCES

- [1] ATHREYA, K. B. and LAHIRI, S. N. (2006). *Measure Theory and Probability Theory*. Springer, New York. [MR2247694](#)
- [2] BAI, J. and NG, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *J. Bus. Econom. Statist.* **23** 49–60. [MR2108691](#)
- [3] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. [MR2567175](#)
- [4] BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. [MR1235416](#)
- [5] BANERJEE, O., EL GHAOU, L. and D’ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. [MR2417243](#)
- [6] BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. [MR3357870](#)
- [7] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245](#)
- [8] BHATTACHARJEE, M. and BOSE, A. (2014). Consistency of large dimensional sample covariance matrix under weak dependence. *Stat. Methodol.* **20** 11–26. [MR3205718](#)
- [9] BHATTACHARJEE, M. and BOSE, A. (2014). Estimation of autocovariance matrices for infinite dimensional vector linear process. *J. Time Series Anal.* **35** 262–281. [MR3194965](#)

MSC2010 subject classifications. Primary 62H12; secondary 62H35.

Key words and phrases. Brain functional connectivity, correlation matrix, heavy tail, high-dimensional data, long memory, minimax optimal convergence rates, nonstationarity, sub-Gaussian tail, temporal dependence.

- [10] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. [MR2485008](#)
- [11] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. [MR2387969](#)
- [12] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. [MR1324786](#)
- [13] BOCHUD, T. and CHALLET, D. (2007). Optimal approximations of power laws with exponentials: Application to volatility models with long memory. *Quant. Finance* **7** 585–589. [MR2374605](#)
- [14] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. [MR2178042](#)
- [15] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. [MR1093459](#)
- [16] BUCKNER, R. L., SEPULCRE, J., TALUKDAR, T., KRIENEN, F. M., LIU, H., HEDDEN, T., ANDREWS-HANNA, J. R., SPERLING, R. A. and JOHNSON, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer’s disease. *J. Neurosci.* **29** 1860–1873.
- [17] CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. [MR2847949](#)
- [18] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- [19] CAI, T. T., LIU, W. and ZHOU, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488. [MR3476606](#)
- [20] CAI, T. T. and YUAN, M. (2012). Adaptive covariance matrix estimation through block thresholding. *Ann. Statist.* **40** 2014–2042. [MR3059075](#)
- [21] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. [MR2676885](#)
- [22] CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. [MR3097607](#)
- [23] CAO, G., BACHEGA, L. R. and BOUMAN, C. A. (2011). The sparse matrix transform for covariance estimation and analysis of high dimensional signals. *IEEE Trans. Image Process.* **20** 625–640. [MR2799176](#)
- [24] CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021. [MR3161455](#)
- [25] CHEN, X., XU, M. and WU, W. B. (2016). Regularized estimation of linear functionals of precision matrices for high-dimensional time series. *IEEE Trans. Signal Process.* **64** 6459–6470. [MR3566612](#)
- [26] CHLEBUS, E. (2009). An approximate formula for a partial sum of the divergent p -series. *Appl. Math. Lett.* **22** 732–737. [MR2514902](#)
- [27] CIUCIU, P., ABRY, P. and HE, B. J. (2014). Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks. *NeuroImage* **95** 248–263.
- [28] COLE, M. W., PATHAK, S. and SCHNEIDER, W. (2010). Identifying the brain’s most globally connected regions. *Neuroimage* **49** 3132–3148.
- [29] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Mathematical Series **9**. Princeton Univ. Press, Princeton, NJ. [MR0016588](#)
- [30] DEMKO, S., MOSS, W. F. and SMITH, P. W. (1984). Decay rates for inverses of band matrices. *Math. Comp.* **43** 491–499. [MR0758197](#)
- [31] EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. [MR2485011](#)
- [32] FANG, Y., WANG, B. and FENG, Y. (2016). Tuning-parameter selection in regularized estimations of large covariance matrices. *J. Stat. Comput. Simul.* **86** 494–509. [MR3421741](#)

- [33] FOMIN, V. (1999). *Optimal Filtering. Vol. II: Spatio-Temporal Fields. Mathematics and Its Applications* **481**. Kluwer Academic, Dordrecht. [MR1707320](#)
- [34] FOUCART, S. and RAUHUT, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Birkhäuser/Springer, New York. [MR3100033](#)
- [35] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- [36] GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Series Anal.* **4** 221–238. [MR0738585](#)
- [37] GOLUB, G. H. and VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd ed. Johns Hopkins Univ. Press, Baltimore, MD. [MR1417720](#)
- [38] GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1** 15–29. [MR0605572](#)
- [39] HARRISON, L., PENNY, W. D. and FRISTON, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage* **19** 1477–1491.
- [40] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. [MR2722294](#)
- [41] HE, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *J. Neurosci.* **31** 13786–13795.
- [42] HINICH, M. J. (1982). Testing for Gaussianity and linearity of a stationary time series. *J. Time Series Anal.* **3** 169–176. [MR0695228](#)
- [43] HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. [MR0614953](#)
- [44] HSIEH, C.-J., SUSTIK, M. A., DHILLON, I. S. and RAVIKUMAR, P. (2014). QUIC: Quadratic approximation for sparse inverse covariance estimation. *J. Mach. Learn. Res.* **15** 2911–2947. [MR3277149](#)
- [45] HU, T.-C., ROSALSKY, A. and VOLODIN, A. (2008). On convergence properties of sums of dependent random variables under second moment and covariance restrictions. *Statist. Probab. Lett.* **78** 1999–2005. [MR2458009](#)
- [46] HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. [MR2277742](#)
- [47] JIANG, T. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48. [MR2082906](#)
- [48] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. [MR2572459](#)
- [49] LI, X., ZHAO, T., YUAN, X. and LIU, H. (2015). The `flare` package for high dimensional linear regression and precision matrix estimation in R. *J. Mach. Learn. Res.* **16** 553–557. [MR3335796](#)
- [50] LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. [MR3319140](#)
- [51] LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10** 2295–2328. [MR2563983](#)
- [52] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. [MR0242239](#)
- [53] MANOLAKIS, D. G., INGLE, V. K. and KOGON, S. M. (2005). *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing* **46**. Artech House, Norwood, MA.
- [54] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- [55] PALMA, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley-Interscience, Hoboken, NJ. [MR2297359](#)

- [56] POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M., SCHLAGGAR, B. L. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.
- [57] PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. [MR0269062](#)
- [58] RACINE, J. (2000). Consistent cross-validatory model-selection for dependent data: Hv-block cross-validation. *J. Econometrics* **99** 39–61.
- [59] RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. [MR2836766](#)
- [60] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. [MR2417391](#)
- [61] ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. [MR2504372](#)
- [62] ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539–550. [MR2672482](#)
- [63] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. [MR3125258](#)
- [64] RYALI, S., CHEN, T., SUPEKAR, K. and MENON, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage* **59** 3852–3861.
- [65] SHU, H. and NAN, B. (2019). Supplement to “Estimation of large covariance and precision matrices from temporally dependent observations.” DOI:[10.1214/18-AOS1716SUPP](#).
- [66] SRIPADA, C., ANGSTADT, M., KESSLER, D., PHAN, K. L., LIBERZON, I., EVANS, G. W., WELSH, R. C., KIM, P. and SWAIN, J. E. (2014). Volitional regulation of emotions produces distributed alterations in connectivity between visual, attention control, and default networks. *NeuroImage* **89** 110–121.
- [67] SYED, M. N., PRINCIPE, J. C. and PARDALOS, P. M. (2012). Correntropy in data classification. In *Dynamics of Information Systems: Mathematical Foundations*. Springer Proc. Math. Stat. **20** 81–117. Springer, New York. [MR3067311](#)
- [68] TAGLIAZUCCHI, E., VON WEGNER, F., MORZELEWSKI, A., BRODBECK, V., JAHNKE, K. and LAUFS, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proc. Natl. Acad. Sci. USA* **110** 15419–15424.
- [69] TAQQU, M. S. (2003). Fractional Brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence* 5–38. Birkhäuser, Boston, MA. [MR1956042](#)
- [70] 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](#)
- [71] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [72] WU, W.-B. and WU, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. [MR3466186](#)
- [73] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#)
- [74] WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. [MR2589209](#)
- [75] YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. [MR2719856](#)
- [76] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. [MR2367824](#)

- [77] ZHOU, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.* **42** 532–562. [MR3210978](#)

BOOTSTRAP TUNING IN GAUSSIAN ORDERED MODEL SELECTION¹

BY VLADIMIR SPOKOINY^{*,†,‡,§,¶} AND NIKLAS WILLRICH^{*}

Weierstrass Institute Berlin^{}, Humboldt University Berlin[†], IITP RAS[‡], HSE[§] and Skoltech Moscow[¶]*

The paper focuses on the problem of model selection in linear Gaussian regression with unknown possibly inhomogeneous noise. For a given family of linear estimators $\{\tilde{\theta}_m, m \in \mathcal{M}\}$, ordered by their variance, we offer a new “smallest accepted” approach motivated by Lepski’s device and the multiple testing idea. The procedure selects the smallest model which satisfies the acceptance rule based on comparison with all larger models. The method is completely data-driven and does not use any prior information about the variance structure of the noise: its parameters are adjusted to the underlying possibly heterogeneous noise by the so-called “propagation condition” using a wild bootstrap method. The validity of the bootstrap calibration is proved for finite samples with an explicit error bound. We provide a comprehensive theoretical study of the method, describe in details the set of possible values of the selected model $\hat{m} \in \mathcal{M}$ and establish some oracle error bounds for the corresponding estimator $\hat{\theta} = \tilde{\theta}_{\hat{m}}$.

REFERENCES

- ARLOT, S. (2009). Model selection by resampling penalization. *Electron. J. Stat.* **3** 557–624. [MR2519533](#)
- BARAUD, Y., HUET, S. and LAURENT, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225–251. [MR1962505](#)
- BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. [MR1679028](#)
- BERAN, R. (1986). Discussion: Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1295–1298.
- BIRGÉ, L. (2001). An alternative point of view on Lepski’s method. In *State of the Art in Probability and Statistics (Leiden, 1999)*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. [MR1836557](#)
- BIRGÉ, L. and MASSART, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138** 33–73. [MR2288064](#)
- CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. [MR2283712](#)
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. [MR3262468](#)
- DALALYAN, A. S. and SALMON, J. (2012). Sharp oracle inequalities for aggregation of affine estimators. *Ann. Statist.* **40** 2327–2355. [MR3059085](#)

MSC2010 subject classifications. Primary 62G05; secondary 62G09, 62J15.

Key words and phrases. Smallest accepted, oracle, propagation condition.

- GACH, F., NICKL, R. and SPOKOINY, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 900–914. [MR3112439](#)
- GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. [MR2604707](#)
- GOEMAN, J. J. and SOLARI, A. (2010). The sequential rejection principle of familywise error control. *Ann. Statist.* **38** 3782–3810. [MR2766868](#)
- GOLDENSHLUGER, A. (2009). A universal procedure for aggregating estimators. *Ann. Statist.* **37** 542–568. [MR2488362](#)
- HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. [MR1245774](#)
- IBRAGIMOV, I. A. and HAS’MINSKIĬ, R. Z. (1981). *Statistical Estimation: Asymptotic Theory. Applications of Mathematics* **16**. Springer, New York. [MR0620321](#)
- KNEIP, A. (1994). Ordered linear smoothers. *Ann. Statist.* **22** 835–866. [MR1292543](#)
- LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. [MR1447734](#)
- LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. [MR1604408](#)
- LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. [MR1091202](#)
- LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. [MR1147167](#)
- LEPSKIĬ, O. V. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatn. Primen.* **37** 468–481. [MR1214353](#)
- MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. [MR1212176](#)
- MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. [MR0468056](#)
- MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. [MR2319879](#)
- PINSKER, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Probl. Inf. Transm.* **16** 52–68. [MR0624591](#)
- ROMANO, J. P. and WOLF, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica* **73** 1237–1282. [MR2149247](#)
- SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. [MR1425962](#)
- SPOKOINY, V. (2012). Parametric estimation. Finite sample theory. *Ann. Statist.* **40** 2877–2909. [MR3097963](#)
- SPOKOINY, V. and VIAL, C. (2009). Parameter tuning in pointwise adaptation using a propagation approach. *Ann. Statist.* **37** 2783–2807. [MR2541447](#)
- SPOKOINY, V., WANG, W. and HÄRDLE, W. K. (2013). Local quantile regression. *J. Statist. Plann. Inference* **143** 1109–1129. [MR3049611](#)
- SPOKOINY, V. and WILLRICH, N. (2019). Supplement to “Bootstrap tuning in Gaussian ordered model selection.” DOI:[10.1214/18-AOS1717SUPP](#).
- SPOKOINY, V. and ZHILOVA, M. (2015). Bootstrap confidence sets under model misspecification. *Ann. Statist.* **43** 2653–2675. [MR3405607](#)
- WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. [MR0868303](#)

SEQUENTIAL CHANGE-POINT DETECTION BASED ON NEAREST NEIGHBORS

BY HAO CHEN¹

University of California, Davis

We propose a new framework for the detection of change-points in on-line, sequential data analysis. The approach utilizes nearest neighbor information and can be applied to sequences of multivariate observations or non-Euclidean data objects, such as network data. Different stopping rules are explored, and one specific rule is recommended due to its desirable properties. An accurate analytic approximation of the average run length is derived for the recommended rule, making it an easy off-the-shelf approach for real multivariate/object sequential data monitoring applications. Simulations reveal that the new approach has better performance than likelihood-based approaches for high dimensional data. The new approach is illustrated through a real dataset in detecting global structural changes in social networks.

REFERENCES

- BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214. [MR0682809](#)
- CHAN, H. P. and WALTHER, G. (2015). Optimal detection of multi-sample aligned sparse signals. *Ann. Statist.* **43** 1865–1895. [MR3375870](#)
- CHEN, H. (2019). Supplement to “Sequential change-point detection based on nearest neighbors.” DOI:[10.1214/18-AOS1718SUPP](#).
- CHEN, H. and ZHANG, N. (2015). Graph-based change-point detection. *Ann. Statist.* **43** 139–176. [MR3285603](#)
- EAGLE, N., PENTLAND, A. S. and LAZER, D. (2009). Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **106** 15274–15278.
- HEARD, N. A., WESTON, D. J., PLATANIOTI, K. and HAND, D. J. (2010). Bayesian anomaly detection methods for social networks. *Ann. Appl. Stat.* **4** 645–662. [MR2758643](#)
- HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. [MR0947577](#)
- KAPPENMAN, J. (2012). A perfect storm of planetary proportions. *IEEE Spectrum* **49** 26–31.
- MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. [MR2650748](#)
- QU, M., SHIH, F. Y., JING, J. and WANG, H. (2005). Automatic solar filament detection using image processing techniques. *Sol. Phys.* **228** 119–135.
- SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. [MR0860514](#)
- SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. [MR0799155](#)

MSC2010 subject classifications. Primary 62G32; secondary 60K35.

Key words and phrases. Change-point, sequential detection, graph-based tests, nonparametrics, scan statistic, tail probability, high-dimensional data, network data, non-Euclidean data.

- SIEGMUND, D. and VENKATRAMAN, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23** 255–271. [MR1331667](#)
- SIEGMUND, D. and YAKIR, B. (2007). *The Statistics of Gene Mapping*. Springer, New York. [MR2301277](#)
- TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. [MR3241619](#)
- TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2008). Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Anal.* **27** 441–475. [MR2460208](#)
- WALD, A. (1973). *Sequential Analysis*. Dover, Mineola, NY.
- WANG, H., TANG, M., PARK, Y. and PRIEBE, C. E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Trans. Signal Process.* **62** 703–717. [MR3160307](#)
- XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. In 2013 *Information Theory and Applications Workshop (ITA)* 1–20. IEEE, Los Alamitos, CA.
- YANG, W., LIPSITCH, M. and SHAMAN, J. (2015). Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl. Acad. Sci. USA* **112** 2723–2728.
- YANG, S., SANTILLANA, M. and KOU, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proc. Natl. Acad. Sci. USA* **112** 14473–14478.

PREDICTION WHEN FITTING SIMPLE MODELS TO HIGH-DIMENSIONAL DATA¹

BY LUKAS STEINBERGER AND HANNES LEEB

Albert Ludwig University of Freiburg and University of Vienna

We study linear subset regression in the context of a high-dimensional linear model. Consider $y = \vartheta + \theta'z + \epsilon$ with univariate response y and a d -vector of random regressors z , and a submodel where y is regressed on a set of p explanatory variables that are given by $x = M'z$, for some $d \times p$ matrix M . Here, “high-dimensional” means that the number d of available explanatory variables in the overall model is much larger than the number p of variables in the submodel. In this paper, we present Pinsker-type results for prediction of y given x . In particular, we show that the mean squared prediction error of the best linear predictor of y given x is close to the mean squared prediction error of the corresponding Bayes predictor $\mathbb{E}[y|x]$, provided only that $p/\log d$ is small. We also show that the mean squared prediction error of the (feasible) least-squares predictor computed from n independent observations of (y, x) is close to that of the Bayes predictor, provided only that both $p/\log d$ and p/n are small. Our results hold uniformly in the regression parameters and over large collections of distributions for the design variables z .

REFERENCES

- ABADIE, A., IMBENS, G. W. and ZHENG, F. (2014). Inference for misspecified models with fixed regressors. *J. Amer. Statist. Assoc.* **109** 1601–1614. [MR3293613](#)
- BACHOC, F., LEEB, H. and PÖTSCHER, B. M. (2015). Valid confidence intervals for post-model-selection prediction. Arxiv preprint. Available at [arXiv:1412.4605](#).
- BERAN, R. and DÜMBGEN, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26** 1826–1856. [MR1673280](#)
- BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. [MR3099122](#)
- BRANNATH, W. and SCHARPENBERG, M. (2014). Interpretation of linear regression coefficients under mean model miss-specification. Arxiv preprint. Available at [arXiv:1409.8544](#).
- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer, Heidelberg. [MR2807761](#)
- BUJA, A. R., BROWN, L. D., GEORGE, E., PITKIN, E., TRASKIN, M., ZHAN, K. and ZHAO, L. (2014). A conspiracy of random predictors and model violations against classical inference in regression. Arxiv preprint. Available at [arXiv:1404.1578](#).
- DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815. [MR0751274](#)
- DÜMBGEN, L. and DEL CONTE-ZERIAL, P. (2013). On low-dimensional projections of high-dimensional distributions. In *From Probability to Statistics and Back: High-Dimensional Models and Processes*. Inst. Math. Stat. (IMS) Collect. **9** 91–104. IMS, Beachwood, OH. [MR3186751](#)

MSC2010 subject classifications. Primary 62H99; secondary 62F99, 62G99.

Key words and phrases. Pinsker theorem, best linear predictor, Bayes predictor, linear subset regression, non-Gaussian data, high-dimensional models, small sample size.

- EATON, M. L. (1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20** 272–276. [MR0866075](#)
- EL KAROUI, N. (2010). The spectrum of kernel random matrices. *Ann. Statist.* **38** 1–50. [MR2589315](#)
- GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988. [MR2108039](#)
- HALL, P. and LI, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. [MR1232523](#)
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics* 221–233. Univ. California Press, Berkeley, CA. [MR0216620](#)
- LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. [MR3485948](#)
- LEEB, H. (2008). Evaluation and selection of models for out-of-sample prediction when the sample size is small relative to the complexity of the data-generating process. *Bernoulli* **14** 661–690. [MR2537807](#)
- LEEB, H. (2009). Conditional predictive inference post model selection. *Ann. Statist.* **37** 2838–2876. [MR2541449](#)
- LEEB, H. (2013). On the conditional distributions of low-dimensional projections from high-dimensional data. *Ann. Statist.* **41** 464–483. [MR3099110](#)
- LEEB, H., PÖTSCHER, B. M. and EWALD, K. (2015). On various confidence intervals post-model-selection. *Statist. Sci.* **30** 216–227. [MR3353104](#)
- PINSKER, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Probl. Inf. Transm.* **16** 120–133. [MR0624591](#)
- ROSENTHAL, H. P. (1970). On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. [MR0271721](#)
- SRIVASTAVA, N. and VERSHYNIN, R. (2013). Covariance estimation for distributions with $2 + \varepsilon$ moments. *Ann. Probab.* **41** 3081–3111. [MR3127875](#)
- STEINBERGER, L. (2015). Statistical inference in high-dimensional linear regression based on simple working models. Ph.D. thesis, Univ. Vienna.
- STEINBERGER, L. and LEEB, H. (2018). On conditional moments of high-dimensional random vectors given lower-dimensional projections. *Bernoulli* **24** 565–591. [MR3706769](#)
- TAYLOR, J., LOCKHART, R., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). Exact post-selection inference for forward stepwise least angle regression. Arxiv preprint. Available at [arXiv:1401.3889](#).

TWO-SAMPLE AND ANOVA TESTS FOR HIGH DIMENSIONAL MEANS¹

BY SONG XI CHEN, JUN LI AND PING-SHOU ZHONG

Peking University, Kent State University and University of Illinois at Chicago

This paper considers testing the equality of two high dimensional means. Two approaches are utilized to formulate L_2 -type tests for better power performance when the two high dimensional mean vectors differ only in sparsely populated coordinates and the differences are faint. One is to conduct thresholding to remove the nonsignal bearing dimensions for variance reduction of the test statistics. The other is to transform the data via the precision matrix for signal enhancement. It is shown that the thresholding and data transformation lead to attractive detection boundaries for the tests. Furthermore, we demonstrate explicitly the effects of precision matrix estimation on the detection boundary for the test with thresholding and data transformation. Extension to multi-sample ANOVA tests is also investigated. Numerical studies are performed to confirm the theoretical findings and demonstrate the practical implementations.

REFERENCES

- BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. [MR1399305](#)
- BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. [MR2387969](#)
- BICKEL, P. J. and LEVINA, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. [MR2485008](#)
- CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. [MR3164870](#)
- CHEN, S. X., LI, J. and ZHONG, P.-S. (2019). Supplement to “Two-sample and ANOVA tests for high dimensional means.” DOI:[10.1214/18-AOS1720SUPP](#).
- CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#)
- DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student’s t -statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 283–301. [MR2815777](#)
- DONOHU, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#)
- DONOHU, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. [MR1311089](#)

MSC2010 subject classifications. Primary 62H15; secondary 62E20.

Key words and phrases. ANOVA, data transformation, large p small n , sparse signals, thresholding, two-sample tests for means.

- FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. [MR1395735](#)
- FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2015). Two-sample Behrens–Fisher problem for high-dimensional data. *Statist. Sinica* **25** 1297–1312. [MR3409068](#)
- GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). A two-sample test for equality of means in high dimension. *J. Amer. Statist. Assoc.* **110** 837–849. [MR3367268](#)
- GRÖCHENIG, K. and LEINERT, M. (2006). Symmetry and inverse-closedness of matrix algebras and functional calculus for infinite matrices. *Trans. Amer. Math. Soc.* **358** 2695–2711. [MR2204052](#)
- HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. [MR2662357](#)
- HE, J. and CHEN, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *J. Econometrics* **194** 283–297. [MR3536981](#)
- HOTELLING, H. (1931). The generalization of student's ratio. *Ann. Math. Stat.* **2** 360–378.
- INGSTER, YU. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#)
- JAFFARD, S. (1990). Propriétés des matrices “bien localisées” près de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **7** 461–476. [MR1138533](#)
- Ji, P. and JIN, J. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *Ann. Statist.* **40** 73–103. [MR3013180](#)
- PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford Studies in Probability **4**. The Clarendon Press, New York. [MR1353441](#)
- QIU, Y. and CHEN, S. X. (2015). Bandwidth selection for high-dimensional covariance matrix estimation. *J. Amer. Statist. Assoc.* **110** 1160–1174. [MR3420692](#)
- SRIVASTAVA, M. S., KATAYAMA, S. and KANO, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.* **114** 349–358. [MR2993891](#)
- SUN, Q. (2005). Wiener's lemma for infinite matrices with polynomial off-diagonal decay. *C. R. Math. Acad. Sci. Paris* **340** 567–570. [MR2138705](#)
- WAGAMAN, A. S. and LEVINA, E. (2009). Discovering sparse covariance structures with the isomap. *J. Comput. Graph. Statist.* **18** 551–572. [MR2751641](#)
- WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#)
- ZHONG, P.-S., CHEN, S. X. and XU, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.* **41** 2820–2851. [MR3161449](#)

VALID CONFIDENCE INTERVALS FOR POST-MODEL-SELECTION PREDICTORS

BY FRANÇOIS BACHOC*, HANNES LEEB^{†,‡,1} AND BENEDIKT M. PÖTSCHER[†]

*University Paul Sabatier**, *University of Vienna[†]* and *DataScience@Univie[‡]*

We consider inference post-model-selection in linear regression. In this setting, Berk et al. [*Ann. Statist.* **41** (2013a) 802–837] recently introduced a class of confidence sets, the so-called PoSI intervals, that cover a certain non-standard quantity of interest with a user-specified minimal coverage probability, irrespective of the model selection procedure that is being used. In this paper, we generalize the PoSI intervals to confidence intervals for post-model-selection predictors.

REFERENCES

- ANDREWS, D. W. K. and GUGGENBERGER, P. (2009). Hybrid and size-corrected subsampling methods. *Econometrica* **77** 721–762. [MR2531360](#)
- BACHOC, F., LEEB, H. and PÖTSCHER, B. M. (2019). Supplement to “Valid confidence intervals for post-model-selection predictors.” DOI:[10.1214/18-AOS1721SUPP](#).
- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2011). Inference for high-dimensional sparse econometric models. In *Advances in Economics and Econometrics. 10th World Congress of the Econometric Society, Vol. III* 245–295.
- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. [MR3207983](#)
- BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013a). Valid post-selection inference. *Ann. Statist.* **41** 802–837. [MR3099122](#)
- BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013b). Valid post-selection inference. Unpublished version. Available at <http://www-stat.wharton.upenn.edu/~lzhao/papers/MyPublication/24PoSI-submit.pdf>.
- CASTERA, L., CHAN, H. L. Y., ARRESE, M., AFDHAL, N., BEDOSSA, P., FRIEDRICH-RUST, M., HAN, K.-H. and PINZANI, M. (2015). EASL-ALEH clinical practice guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J. Hepatol.* **63** 237–264.
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FITHIAN, W., SUN, D. and TAYLOR, J. (2015). Optimal inference after model selection. Available at [arXiv:1410.2597](#).
- GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988. [MR2108039](#)
- JAUPI, L. (2014). Variable selection methods for multivariate process monitoring. In *Proceedings of the World Congress of Engineering 2014, Vol. II* (S. I. Ao, L. Gelman, D. Hukins, A. Hunter and A. M. Korsunsky, eds.) 1116–1120.

MSC2010 subject classifications. Primary 62F25; secondary 62J05.

Key words and phrases. Inference post-model-selection, confidence intervals, optimal post-model-selection predictors, nonstandard targets, linear regression.

- KABAILA, P. and LEEB, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *J. Amer. Statist. Assoc.* **101** 619–629. [MR2256178](#)
- LEE, J. D. and TAYLOR, J. (2014). Exact post model selection inference for marginal screening. In *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) 136–144. Curran Associates, Red Hook, NY.
- LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. [MR3485948](#)
- LEE, H. (2009). Conditional predictive inference post model selection. *Ann. Statist.* **37** 2838–2876. [MR2541449](#)
- LEE, H. and PÖTSCHER, B. M. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* **19** 100–142. [MR1965844](#)
- LEE, H. and PÖTSCHER, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21** 21–59. [MR2153856](#)
- LEE, H. and PÖTSCHER, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Statist.* **34** 2554–2591. [MR2291510](#)
- LEE, H. and PÖTSCHER, B. M. (2017). Testing in the presence of nuisance parameters: Some comments on tests post-model-selection and random critical values. In *Big and Complex Data Analysis* (S. E. Ahmed, ed.) 69–82. Springer, Cham. [MR3644121](#)
- LEE, H., PÖTSCHER, B. M. and EWALD, K. (2015). On various confidence intervals post-model-selection. *Statist. Sci.* **30** 216–227. [MR3353104](#)
- LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- PÖTSCHER, B. M. (2009). Confidence sets based on sparse estimators are necessarily large. *Sankhyā* **71** 1–18. [MR2579644](#)
- PÖTSCHER, B. M. and SCHNEIDER, U. (2010). Confidence sets based on penalized maximum likelihood estimators in Gaussian regression. *Electron. J. Stat.* **4** 334–360. [MR2645488](#)
- RAWLINGS, J. O., PANTULA, S. G. and DICKEY, D. A. (1998). *Applied Regression Analysis: A Research Tool*, 2nd ed. Springer, New York. [MR1631919](#)
- SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York. [MR0116429](#)
- SCHNEIDER, U. (2016). Confidence sets based on thresholding estimators in high-dimensional Gaussian regression models. *Econometric Rev.* **35** 1412–1455. [MR3511026](#)
- SOUDERS, T. M. and STENBAKKEN, G. N. (1991). Cutting the high cost of testing. *IEEE Spectrum* **28** 48–51.
- TIAN, X. and TAYLOR, J. (2015). Asymptotics of selective inference. Available at [arXiv:1501.03588](#).
- TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2015). Uniform asymptotic inference and the bootstrap after model selection. Available at [arXiv:1506.06266](#).
- TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. [MR3538689](#)
- 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](#)
- WASSERMAN, L. (2014). Discussion: “A significance test for the lasso” [[MR3210970](#)]. *Ann. Statist.* **42** 501–508. [MR3210975](#)
- WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. [MR2543689](#)
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHANG, K. (2013). Rank-extreme association of Gaussian vectors and low-rank detection. Available at [arXiv:1306.0623](#).

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](#)

A ROBUST AND EFFICIENT APPROACH TO CAUSAL INFERENCE BASED ON SPARSE SUFFICIENT DIMENSION REDUCTION

BY SHUIJIE MA^{*,1,4}, LIPING ZHU^{†,2,4}, ZHIWEI ZHANG^{*}, CHIH-LING TSAI[‡]
 AND RAYMOND J. CARROLL^{§,¶,3}

University of California, Riverside^{}, Renmin University of China[†],
 University of California, Davis[‡], Texas A&M University[§]
 and University of Technology[¶]*

A fundamental assumption used in causal inference with observational data is that treatment assignment is ignorable given measured confounding variables. This assumption of no missing confounders is plausible if a large number of baseline covariates are included in the analysis, as we often have no prior knowledge of which variables can be important confounders. Thus, estimation of treatment effects with a large number of covariates has received considerable attention in recent years. Most existing methods require specifying certain parametric models involving the outcome, treatment and confounding variables, and employ a variable selection procedure to identify confounders. However, selection of a proper set of confounders depends on correct specification of the working models. The bias due to model misspecification and incorrect selection of confounding variables can yield misleading results. We propose a robust and efficient approach for inference about the average treatment effect via a flexible modeling strategy incorporating penalized variable selection. Specifically, we consider an estimator constructed based on an efficient influence function that involves a propensity score and an outcome regression. We then propose a new sparse sufficient dimension reduction method to estimate these two functions without making restrictive parametric modeling assumptions. The proposed estimator of the average treatment effect is asymptotically normal and semiparametrically efficient without the need for variable selection consistency. The proposed methods are illustrated via simulation studies and a biomedical application.

REFERENCES

- ABADIE, A. and IMBENS, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74** 235–267. [MR2194325](#)
- BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. [MR2216189](#)
- BECK, A. and TBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. [MR2486527](#)

MSC2010 subject classifications. Primary 62G08; secondary 62G10, 62G20, 62J07.

Key words and phrases. Average treatment effect, dimension reduction, high-dimensional data, multiple-index model, outcome regression, semiparametric efficiency.

- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. [MR3207983](#)
- BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. [MR3099122](#)
- BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- BURA, E. and COOK, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. *J. Amer. Statist. Assoc.* **96** 996–1003. [MR1946367](#)
- CAO, W., TSIATIS, A. A. and DAVIDIAN, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96** 723–734. [MR2538768](#)
- CHAN, K. C. G. and YAM, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statist. Sci.* **29** 380–396. [MR3264551](#)
- CHARLTON, K., KOWAL, P., SORIANO, M. M., WILLIAMS, S., BANKS, E., VO, K. and BYLES, J. (2014). Fruit and vegetable intake and body mass index in a large sample of middle-aged Australian men and women. *Nutrients* **6** 2305–2319.
- CHEN, L. and HUANG, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J. Amer. Statist. Assoc.* **107** 1533–1545. [MR3036414](#)
- COOK, R. D. and LEE, H. (1999). Dimension reduction in binary response regression. *J. Amer. Statist. Assoc.* **94** 1187–1200. [MR1731482](#)
- COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. [MR1902895](#)
- DUAN, N. and LI, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530. [MR1105834](#)
- FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. [MR3397349](#)
- FENG, Z., WEN, X. M., YU, Z. and ZHU, L. (2013). On partial sufficient dimension reduction with applications to partially linear multi-index models. *J. Amer. Statist. Assoc.* **108** 237–246. [MR3174616](#)
- FREEDMAN, D. A. and BERK, R. A. (2008). Weighting regressions by propensity scores. *Eval. Rev.* **32** 392–409.
- GHOSH, D. (2011). Propensity score modelling in observational studies using dimension reduction methods. *Statist. Probab. Lett.* **81** 813–820. [MR2793748](#)
- GONG, P., ZHANG, C., LU, Z., HUANG, J. Z. and YE, J. (2013). A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *Proceedings of the 30th International Conference on Machine Learning (ICML)* **28** 37–45.
- HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. [MR1612242](#)
- HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *Statist. Sci.* **1** 297–318. [MR0858512](#)
- HECKMAN, J. J., ICHIMURA, H. and TODD, P. (1998). Matching as an econometric evaluation estimator. *Rev. Econ. Stud.* **65** 261–294. [MR1623713](#)
- HEO, M., KIM, R. S., WYLIE-ROSETT, J., ALLISON, D. B., HEYMSFIELD, S. B. and FAITH, M. S. (2011). Inverse association between fruit and vegetable intake and BMI even after controlling for demographic, socioeconomic and lifestyle factors. *Obesity Facts* **4** 449–455.
- HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. [MR1995826](#)
- IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. [MR3153941](#)

- KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. [MR2420458](#)
- KAUFMAN, A., AUGUSTON, E. M. and PATRICK, H. (2012). Unraveling the relationship between smoking and weight: The role of sedentary behavior. *J. Obesity*. DOI:[10.1155/2012/735465](#).
- LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. [MR1137117](#)
- LI, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. *J. Amer. Statist. Assoc.* **87** 1025–1039. [MR1209564](#)
- LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- LUO, W. and LI, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika* **103** 875–887. [MR3620445](#)
- LUO, W., ZHU, Y. and GHOSH, D. (2017). On estimating regression-based causal effects using sufficient dimension reduction. *Biometrika* **104** 51–65. [MR3626479](#)
- MA, Y. and ZHU, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. [MR2949349](#)
- MA, S., ZHU, L., ZHANG, Z., TSAI, C.-L. and CARROLL, R. J. (2018). Supplement to “A robust and efficient approach to causal inference based on sparse sufficient dimension reduction.” DOI:[10.1214/18-AOS1722SUPP](#).
- MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415. [MR0679685](#)
- RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. [MR2719855](#)
- 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. Model.* **7** 1393–1512. [MR0877758](#)
- ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.
- ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. [MR0742974](#)
- ROSENBAUM, P. R. and RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79** 516–524.
- ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33–38.
- ROTNITZKY, A., LEI, Q., SUED, M. and ROBINS, J. M. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika* **99** 439–456. [MR2931264](#)
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.
- SCHATZKIN, A., SUBAR, A. F., THOMPSON, F. E., HARLAN, L. C., TANGREA, J., HOLLENBECK, A. R., HURWITZ, P. E., COYLE, L., SCHUSSLER, N., MICHAUD, D. S., FREEDMAN, L. S., BROWN, C. C., MIDTHUNE, D. and KIPNIS, V. (2001). Design and serendipity in establishing a large cohort with wide dietary intake distributions: The national institutes of health-aarp diet and health study. *Am. J. Epidemiol.* **154** 1119–1125.
- SEKHON, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *J. Stat. Softw.* **42** 1–52.
- SNOWDEN, J. M., ROSE, S. and MORTIMER, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *Am. J. Epidemiol.* **173** 731–738.

- STEFFEN, L. M., JACOBS, D. R., MURTAUGH, M. A., MORAN, A., STEINBERGER, J., HONG, C. P. and SINAICO, A. R. (2003). Whole grain intake is associated with lower body mass and greater insulin sensitivity among adolescents. *Am. J. Epidemiol.* **158** 243–250.
- TAN, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Amer. Statist. Assoc.* **101** 1619–1637. [MR2279484](#)
- TAN, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97** 661–682. [MR2672490](#)
- VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. [MR3224285](#)
- VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. 25, 23. [MR2349918](#)
- VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality. Springer Series in Statistics.* Springer, New York. [MR1958123](#)
- VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics.* Springer, New York. [MR2867111](#)
- VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. [MR2306500](#)
- WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. [MR2543689](#)
- XIA, Y. (2008). A multiple-index model and dimension reduction. *J. Amer. Statist. Assoc.* **103** 1631–1640. [MR2504209](#)
- XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. [MR1924297](#)
- YIN, X., LI, B. and COOK, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *J. Multivariate Anal.* **99** 1733–1757. [MR2444817](#)
- YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. [MR2212574](#)
- ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- 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](#)
- ZHOU, S., VAN DE GEER, S. and BÜHLMANN, P. (2009). Adaptive Lasso for high dimensional regression and Gaussian graphical modeling. Available at [arxiv:0903.2515](#).

THE MAXIMUM LIKELIHOOD THRESHOLD OF A PATH DIAGRAM¹

BY MATHIAS DRTON^{*,†}, CHRISTOPHER FOX[‡], ANDREAS KÄUFL[§] AND
GUILLAUME POULIOT^{‡,¶}

University of Washington^{*}, *University of Copenhagen*[†], *University of Chicago*[‡],
University of Augsburg[§] and *École Polytechnique de Montréal*[¶]

Linear structural equation models postulate noisy linear relationships between variables of interest. Each model corresponds to a path diagram, which is a mixed graph with directed edges that encode the domains of the linear functions and bidirected edges that indicate possible correlations among noise terms. Using this graphical representation, we determine the maximum likelihood threshold, that is, the minimum sample size at which the likelihood function of a Gaussian structural equation model is almost surely bounded. Our result allows the model to have feedback loops and is based on decomposing the path diagram with respect to the connected components of its bidirected part. We also prove that if the sample size is below the threshold, then the likelihood function is almost surely unbounded. Our work clarifies, in particular, that standard likelihood inference is applicable to sparse high-dimensional models even if they feature feedback loops.

REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. [MR1439312](#)
- BARBER, R. F., DRTON, M. and WEIHS, L. (2015). SEMID: Identifiability of linear structural equation models. R package version 0.2.
- BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. [MR0489333](#)
- BUHL, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* **20** 263–270. [MR1241392](#)
- DAHL, J., VANDENBERGHE, L. and ROYCHOWDHURY, V. (2008). Covariance selection for non-chordal graphs via chordal embedding. *Optim. Methods Softw.* **23** 501–520. [MR2440363](#)
- DIESTEL, R. (2010). *Graph Theory*, 4th ed. *Graduate Texts in Mathematics* **173**. Springer, Heidelberg. [MR2744811](#)
- DRTON, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. [MR2502658](#)
- DRTON, M. (2016). Algebraic problems in structural equation modeling. Available at [arXiv:1612.05994](#).
- DRTON, M., FOX, C. and WANG, Y. S. (2018). Computation of maximum likelihood estimates in cyclic structural equation models. *Ann. Statist.* To appear. Available at [arXiv:1610.03434](#).

MSC2010 subject classifications. 62H12, 62J05.

Key words and phrases. Covariance matrix, graphical model, maximum likelihood, normal distribution, path diagram, structural equation model.

- DRTON, M. and RICHARDSON, T. S. (2008). Graphical methods for efficient likelihood inference in Gaussian covariance models. *J. Mach. Learn. Res.* **9** 893–914. [MR2417257](#)
- DRTON, M. and YU, J. (2010). On a parametrization of positive semidefinite matrices with zeros. *SIAM J. Matrix Anal. Appl.* **31** 2665–2680. [MR2740626](#)
- EVANS, R. J. and RICHARDSON, T. S. (2016). Smooth, identifiable supermodels of discrete DAG models with latent variables. Available at [arXiv:1511.06813](#).
- FOYGEL, R., DRAISMA, J. and DRTON, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.* **40** 1682–1713. [MR3015040](#)
- GRACE, J. B., ANDERSON, T. M., SEABLOOM, E. W., BORER, E. T., ADLER, P. B., HARPOLE, W. S., HAUTIER, Y., HILLEBRAND, H., LIND, E. M., PÄRTEL, M., BAKKER, J. D., BUCKLEY, Y. M., CRAWLEY, M. J., DAMSCHEN, E. I., DAVIES, K. F., FAY, P. A., FIRN, J., GRUNER, D. S., HECTOR, A., KNOPS, J. M. H., MACDOUGALL, A. S., MELBOURNE, B. A., MORGAN, J. W., ORROCK, J. L., PROBER, S. M. and SMITH, M. D. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529** 390–393.
- GRONE, R., JOHNSON, C. R., DE SÁ, E. M. and WOLKOWICZ, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.* **58** 109–124. [MR0739282](#)
- GROSS, E. and SULLIVANT, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli* **24** 386–407. [MR3706762](#)
- HORN, R. A. and JOHNSON, C. R. (1990). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. [MR1084815](#)
- KAUERMANN, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Stat.* **23** 105–116. [MR1380485](#)
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series **17**. Clarendon Press, New York. [MR1419991](#)
- MAATHUIS, M. H., COLOMBO, D., KALISCH, M. and BÜHLMANN, P. (2010). Predicting causal effects in large-scale systems from observational data. *Nat. Methods* **7** 247–248.
- OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** 763–765. [MR0331643](#)
- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2548166](#)
- SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. MIT Press, Cambridge, MA. [MR1815675](#)
- UHLER, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.* **40** 238–261. [MR3014306](#)
- WOODBURY, M. A. (1950). *Inverting Modified Matrices*. Statistical Research Group, Memo. Rep. **42**. Princeton Univ., Princeton, NJ. [MR0038136](#)
- WRIGHT, S. (1921). Correlation and causation. *J. Agricultural Research* **20** 557–585.
- WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.

CONVEX REGULARIZATION FOR HIGH-DIMENSIONAL MULTIRESPONSE TENSOR REGRESSION

BY GARVESH RASKUTTI*,¹, MING YUAN*,^{†,2} AND HAN CHEN*,²

University of Wisconsin-Madison and Columbia University[†]*

In this paper, we present a general convex optimization approach for solving high-dimensional multiple response tensor regression problems under low-dimensional structural assumptions. We consider using convex and weakly decomposable regularizers assuming that the underlying tensor lies in an unknown low-dimensional subspace. Within our framework, we derive general risk bounds of the resulting estimate under fairly general dependence structure among covariates. Our framework leads to upper bounds in terms of two very simple quantities, the *Gaussian width* of a convex set in tensor space and the *intrinsic dimension* of the low-dimensional tensor subspace. To the best of our knowledge, this is the first general framework that applies to multiple response problems. These general bounds provide useful upper bounds on rates of convergence for a number of fundamental statistical models of interest including multiresponse regression, vector autoregressive models, low-rank tensor models and pairwise interaction models. Moreover, in many of these settings we prove that the resulting estimates are minimax optimal. We also provide a numerical study that both validates our theoretical guarantees and demonstrates the breadth of our framework.

REFERENCES

- [1] ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York. [MR0771294](#)
- [2] BASU, S. and MICHAELIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. [MR3357870](#)
- [3] BHATIA, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer, New York. [MR1477662](#)
- [4] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- [5] CHEN, S., LYU, M. R., KING, I. and XU, Z. (2013). Exact and stable recovery of pairwise interaction tensors. In *Advances in Neural Information Processing Systems*.
- [6] COHEN, S. B. and COLLINS, M. (2012). Tensor decomposition for fast parsing with latent-variable PCFGs. In *Advances in Neural Information Processing Systems*.
- [7] GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low- n -rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010, 19. [MR2765628](#)
- [8] GORDON, Y. (1988). On Milman’s inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. [MR0950977](#)

MSC2010 subject classifications. 60K35.

Key words and phrases. Tensor regression, convex regularization, Gaussian width, intrinsic dimension, low-rank, sparsity.

- [9] HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations. Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. [MR3616141](#)
- [10] HOFF, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat.* **9** 1169–1193. [MR3418719](#)
- [11] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. [MR2535056](#)
- [12] LI, N. and LI, B. (2010). Tensor completion for on-board compression of hyperspectral images. In *17th IEEE International Conference on Image Processing (ICIP)* 517–520.
- [13] LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. [MR2172368](#)
- [14] MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. [MR3565471](#)
- [15] MESGARANI, N., SLANEY, M. and SHAMMA, S. (2006). Content-based audio classification based on multiscale spectro-temporal features. *IEEE Trans. Speech Audio Process.* **14** 920–930.
- [16] MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *International Conference on Machine Learning*.
- [17] NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. [MR2930649](#)
- [18] 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](#)
- [19] NION, D. and SIDIROPOULOS, N. D. (2010). Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar. *IEEE Trans. Signal Process.* **58** 5693–5705. [MR2789612](#)
- [20] PISIER, G. (1989). *The Volume of Convex Bodies and Banach Space Geometry. Cambridge Tracts in Mathematics* **94**. Cambridge Univ. Press, Cambridge. [MR1036275](#)
- [21] QIN, Z., SCHEINBERG, K. and GOLDFARB, D. (2013). Efficient block-coordinate descent algorithms for the group Lasso. *Math. Program. Comput.* **5** 143–169. [MR3069877](#)
- [22] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. [MR2719855](#)
- [23] RASKUTTI, G., YUAN, M. and CHEN, H. (2019). Supplement to “Convex regularization for high-dimensional multiresponse tensor regression.” DOI:[10.1214/18-AOS1725SUPP](#).
- [24] RENDLE, S., MARINHO, L. B., NANOPOULOS, A. and SCHMIDT-THIEME, L. (2009). Learning optimal ranking with tensor factorization for tag recommendation. In *SIGKDD*.
- [25] RENDLE, S. and SCHMIDT-THIEME, L. (2010). Pairwise interaction tensor factorization for personalized tag recommendation. In *ICDM*.
- [26] ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. [MR0274683](#)
- [27] SEMERCI, O., HAO, N., KILMER, M. E. and MILLER, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23** 1678–1693. [MR3191324](#)
- [28] SIMON, N., FRIEDMAN, J. and HASTIE, T. (2013). A blockwise coordinate descent algorithm for penalized multiresponse and grouped multinomial regression. Technical report, Georgia Tech.
- [29] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)

- [30] TURLACH, B. A., VENABLES, W. N. and WRIGHT, S. J. (2005). Simultaneous variable selection. *Technometrics* **47** 349–363. [MR2164706](#)
- [31] VAN DE GEER, S. (2014). Weakly decomposable regularization penalties and structured sparsity. *Scand. J. Stat.* **41** 72–86. [MR3181133](#)
- [32] YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. [MR2212574](#)
- [33] YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. [MR3529132](#)
- [34] ZHOU, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. Technical report, ETH, Zurich. Available at [arXiv:0912.4045](#).

LARGE SAMPLE THEORY FOR MERGED DATA FROM MULTIPLE SOURCES

BY TAKUMI SAEGUSA

University of Maryland

We develop large sample theory for merged data from multiple sources. Main statistical issues treated in this paper are (1) the same unit potentially appears in multiple datasets from overlapping data sources, (2) duplicated items are not identified and (3) a sample from the same data source is dependent due to sampling without replacement. We propose and study a new weighted empirical process and extend empirical process theory to a dependent and biased sample with duplication. Specifically, we establish the uniform law of large numbers and uniform central limit theorem over a class of functions along with several empirical process results under conditions identical to those in the i.i.d. setting. As applications, we study infinite-dimensional M -estimation and develop its consistency, rates of convergence and asymptotic normality. Our theoretical results are illustrated with simulation studies and a real data example.

REFERENCES

- [1] ALEXANDER, K. S. (1985). Rates of growth for weighted empirical processes. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*. Wadsworth Statist./Probab. Ser. 475–493. Wadsworth, Belmont, CA. [MR0822047](#)
- [2] BAE, J. and LEVENTAL, S. (1995). Uniform CLT for Markov chains and its invariance principle: A martingale approach. *J. Theoret. Probab.* **8** 549–570. [MR1340827](#)
- [3] BERKES, I. and PHILIPP, W. (1977/78). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrsch. Verw. Gebiete* **41** 115–137. [MR0464344](#)
- [4] BERTAIL, P., CHAUTRU, E. and CLÉMENÇON, S. (2017). Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.* **44** 97–111. [MR3619696](#)
- [5] BOISTARD, H., LOPUHAÄ, H. P. and RUIZ-GAZEN, A. (2017). Functional central limit theorems for single-stage sampling designs. *Ann. Statist.* **45** 1728–1758. [MR3670194](#)
- [6] BRESLOW, N. E. and CHATTERJEE, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **48** 457–468.
- [7] BRESLOW, N. E. and WELLNER, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Stat.* **34** 86–102. [MR2325244](#)
- [8] BRESLOW, N. E. and WELLNER, J. A. (2008). A Z-theorem with estimated nuisance parameters and correction note for: “Weighted likelihood for semiparametric models and

MSC2010 subject classifications. Primary 62E20; secondary 62G20, 62D99, 62N01.

Key words and phrases. Calibration, data integration, empirical process, nonregular, sampling without replacement, semiparametric model.

- two-phase stratified samples, with application to Cox regression" [Scand. J. Statist. **34** (2007) 86–102; MR2325244]. *Scand. J. Stat.* **35** 186–192. [MR2391566](#)
- [9] BRICK, J. M., DIPKO, S., PRESSER, S., TUCKER, C. and YUAN, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opin. Q.* **70** 780–793.
- [10] CANTELLI, F. P. (1933). Sulla determinazione empirica delle leggi di probabilit . *G. Ist. Ital. Attuari* **4** 421–424.
- [11] CERVANTES, I. F., JONES, M. E., ROJAS, L. A., BRICK, J. M., KURATA, J. and GRANT, D. (2006). A review of the sample design for the California health interview survey. In *Proceedings of the Social Statistics Section* 3023–3030. Amer. Statist. Assoc., Alexandria, VA.
- [12] CHATTERJEE, N., CHEN, Y.-H., MAAS, P. and CARROLL, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J. Amer. Statist. Assoc.* **111** 107–117. [MR3494641](#)
- [13] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. [MR0341758](#)
- [14] D'ANGIO, G. J., BRESLOW, N., BECKWITH, J. B., EVANS, A., BAUM, H., DELORIMIER, A., FERNBACH, D., HRABOVSKY, E., JONES, B. and KELALIS, P. (1989). Treatment of Wilms' tumor. Results of the Third National Wilms' Tumor Study. *Cancer* **64** 349–360.
- [15] DE LEEUW, E. D. (2005). To mix or not to mix data collection modes in surveys. *J. Off. Stat.* **21** 233–255.
- [16] DEVILLE, J.-C. and S RNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. [MR1173804](#)
- [17] DILLMAN, D. A., SMYTH, J. D., CHRISTIAN and MELANI, L. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 4th ed. Wiley, New York.
- [18] DING, Y. and NAN, B. (2011). A sieve M -theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Statist.* **39** 3032–3061. [MR3012400](#)
- [19] DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov–Smirnov theorems. *Ann. Math. Stat.* **23** 277–281. [MR0047288](#)
- [20] DUDLEY, R. M. (1981). Donsker classes of functions. In *Statistics and Related Topics (Ottawa, Ont., 1980)* 341–352. North-Holland, Amsterdam. [MR0665285](#)
- [21] FELLEGI, I. P. and SUNTER, A. B. (1969). A theory for record linkage. *J. Amer. Statist. Assoc.* **64** 1183–1210.
- [22] GIN , E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–998. [MR0757767](#)
- [23] GLIVENKO, V. (1933). Sulla determinazione empirica della legge di probabilit . *G. Ist. Ital. Attuari* **4** 92–99.
- [24] H JEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magy. Tud. Akad. Mat. Kut. Int z. K zl.* **5** 361–374. [MR0125612](#)
- [25] HARTLEY, H. O. (1962). Multiple frame surveys. In *Proceedings of the Social Statistics Section* 203–206. Amer. Statist. Assoc., Alexandria, VA.
- [26] HARTLEY, H. O. (1974). Multiple frame methodology and selected applications. *Sankhy , Ser. C* **36** 99–118.
- [27] HARTLEY, H. O. and SIELKEN, R. L. JR. (1975). A "super-population viewpoint" for finite population sampling. *Biometrics* **31** 411–422. [MR0386084](#)
- [28] HERZOG, T. N., SCHEUREN, F. J. and WINKLER, W. E. (2007). *Data Quality and Record Linkage Techniques*. 1st ed. Springer, Berlin.
- [29] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. [MR0053460](#)

- [30] HU, S. S., BALLUZ, L., BATTAGLIA, M. P. and FRANKEL, M. R. (2011). Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. *Am. J. Epidemiol.* **173** 703–711.
- [31] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. [MR1394975](#)
- [32] HUANG, J. and WELLNER, J. A. (1997). Interval censored survival data: A review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* 123–169. Springer, Berlin.
- [33] IACHAN, R. and DENNIS, M. L. (1993). A multiple frame approach to sampling the homeless and transient population. *J. Off. Stat.* **9** 747–764.
- [34] ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. [MR0648029](#)
- [35] KALTON, G. and ANDERSON, D. W. (1986). Sampling rare populations. *J. R. Stat. Soc., A* **149** 65–82.
- [36] KEIDING, N. and LOUIS, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. Roy. Statist. Soc. Ser. A* **179** 319–376. [MR3461587](#)
- [37] KIM, G. and CHAMBERS, R. (2012). Regression analysis under incomplete linkage. *Comput. Statist. Data Anal.* **56** 2756–2770. [MR2915160](#)
- [38] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics.* Springer, New York. [MR2724368](#)
- [39] LAHIRI, P. and LARSEN, M. D. (2005). Regression analysis with linked data. *J. Amer. Statist. Assoc.* **100** 222–230. [MR2156832](#)
- [40] LEVENTAL, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *J. Theoret. Probab.* **2** 271–287. [MR0996990](#)
- [41] LOHR, S. and RAO, J. N. K. (2006). Estimation in multiple-frame surveys. *J. Amer. Statist. Assoc.* **101** 1019–1030. [MR2324141](#)
- [42] LU, Y. (2012). Regression coefficient estimation in dual frame surveys. In *Proceedings of the Section on Survey Research Methods* 4687–4695. Amer. Statist. Assoc., Alexandria, VA.
- [43] LU, Y. and LOHR, S. L. (2010). Gross flow estimation in dual frame surveys. *Surv. Methodol.* **36** 13–22.
- [44] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96** 190–217. [MR2202406](#)
- [45] METCALF, P. and SCOTT, A. (2009). Using multiple frames in health surveys. *Stat. Med.* **28** 1512–1523. [MR2649709](#)
- [46] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. [MR1245301](#)
- [47] RANALLI, M. G., ARCOS, A., DEL MAR RUEDA, M. and TEODORO, A. (2016). Calibration estimation in dual-frame surveys. *Stat. Methods Appl.* **25** 321–349. [MR3539496](#)
- [48] RAO, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Off. Stat.* **10** 153–165.
- [49] RAO, J. N. K. and WU, C. (2010). Pseudo-empirical likelihood inference for multiple frame surveys. *J. Amer. Statist. Assoc.* **105** 1494–1503. [MR2796566](#)
- [50] RUBIN-BLEUER, S. and SCHIOPU KRATINA, I. (2005). On the two-phase framework for joint model and design-based inference. *Ann. Statist.* **33** 2789–2810. [MR2253102](#)
- [51] SAEGUSA, T. (2019). Supplement to “Large sample theory for merged data from multiple sources.” DOI:[10.1214/18-AOS1727SUPP](#).
- [52] SAEGUSA, T. and WELLNER, J. A. (2013). Weighted likelihood estimation under two-phase sampling. *Ann. Statist.* **41** 269–295. [MR3059418](#)
- [53] SKINNER, C. J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *J. Amer. Statist. Assoc.* **86** 779–784. [MR1147105](#)

- [54] SKINNER, C. J. and RAO, J. N. K. (1996). Estimation in dual frame surveys with complex designs. *J. Amer. Statist. Assoc.* **91** 349–356. [MR1394091](#)
- [55] VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1999)*. *Lecture Notes in Math.* **1781** 331–457. Springer, Berlin. [MR1915446](#)
- [56] VAN DER VAART, A. W. (1995). Efficiency of infinite-dimensional M -estimators. *Stat. Neerl.* **49** 9–30. [MR1333176](#)
- [57] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- [58] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. [MR1385671](#)
- [59] WINKLER, W. E. (1995). *Matching and Record Linkage* 353–384. Wiley, New York.
- [60] ZIEGLER, K. (1997). Functional central limit theorems for triangular arrays of function-indexed processes under uniformly integrable entropy conditions. *J. Multivariate Anal.* **62** 233–272. [MR1473875](#)
- [61] ZIEGLER, K. (2001). Uniform laws of large numbers for triangular arrays of function-indexed processes under random entropy conditions. *Results Math.* **39** 374–389. [MR1834583](#)

KHINCHINE'S THEOREM AND EDGEWORTH APPROXIMATIONS FOR WEIGHTED SUMS

BY SERGEY G. BOBKOV¹

University of Minnesota and National Research University Higher School of Economics

Let F_n denote the distribution function of the normalized sum of n i.i.d. random variables. In this paper, polynomial rates of approximation of F_n by the corrected normal laws are considered in the model where the underlying distribution has a convolution structure. As a basic tool, the convergence part of Khinchine's theorem in metric theory of Diophantine approximations is extended to the class of product characteristic functions.

REFERENCES

- [1] ANGST, J. and POLY, G. (2017). A weak Cramér condition and application to Edgeworth expansions. *Electron. J. Probab.* **22** Paper No. 59, 24. [MR3683368](#)
- [2] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. [MR0436272](#)
- [3] BOBKOV, S. G. (2016). Closeness of probability distributions in terms of Fourier–Stieltjes transforms. *Uspekhi Mat. Nauk* **71** 37–98. [MR3588939](#)
- [4] BOBKOV, S. G. (2017). Asymptotic expansions for products of characteristic functions under moment assumptions of non-integer orders. In *Convexity and Concentration* (E. Carlen, M. Madiman and E. Werner, eds.). *IMA Vol. Math. Appl.* **161** 297–357. [MR3837276](#)
- [5] BOBKOV, S. G. (2017). Central limit theorem and Diophantine approximations. *J. Theoret. Probab.* **31** 2390–2411. [MR3866618](#)
- [6] CASSELS, J. W. S. (1957). *An Introduction to Diophantine Approximation. Cambridge Tracts in Mathematics and Mathematical Physics* **45**. Cambridge Univ. Press, New York. [MR0087708](#)
- [7] ESSEEN, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. [MR0014626](#)
- [8] PETROV, V. V. (1975). *Sums of Independent Random Variables. Ergebnisse der Mathematik und ihrer Grenzgebiete* **82**. Springer, New York. Translated from the Russian by A. A. Brown. [MR0388499](#)
- [9] PETROV, V. V. (1987). *Limit Theorems for Sums of Independent Random Variables* (Russian). Nauka, Moscow. [MR0896036](#)
- [10] SCHMIDT, W. M. (1980). *Diophantine Approximation. Lecture Notes in Math.* **785**. Springer, Berlin. [MR0568710](#)
- [11] SCHMIDT, W. M. (1991). *Diophantine Approximations and Diophantine Equations. Lecture Notes in Math.* **1467**. Springer, Berlin. [MR1176315](#)

MSC2010 subject classifications. 60F05.

Key words and phrases. Central limit theorem, Edgeworth approximations.

DISTRIBUTED INFERENCE FOR QUANTILE REGRESSION PROCESSES

BY STANISLAV VOLGUSHEV^{*,1}, SHIH-KANG CHAO^{†,2} AND
 GUANG CHENG^{†,2,3}

University of Toronto^{} and Purdue University[†]*

The increased availability of massive data sets provides a unique opportunity to discover subtle patterns in their distributions, but also imposes overwhelming computational challenges. To fully utilize the information contained in big data, we propose a two-step procedure: (i) estimate conditional quantile functions at different levels in a parallel computing environment; (ii) construct a conditional quantile regression process through projection based on these estimated quantile curves. Our general quantile regression framework covers both linear models with fixed or growing dimension and series approximation models. We prove that the proposed procedure does not sacrifice any statistical inferential accuracy provided that the number of distributed computing units and quantile levels are chosen properly. In particular, a sharp upper bound for the former and a sharp lower bound for the latter are derived to capture the minimal computational cost from a statistical perspective. As an important application, the statistical inference on conditional distribution functions is considered. Moreover, we propose computationally efficient approaches to conducting inference in the distributed estimation setting described above. Those approaches directly utilize the availability of estimators from subsamples and can be carried out at almost no additional computational cost. Simulations confirm our statistical inferential theory.

REFERENCES

- BANERJEE, M., DUROT, C. and SEN, B. (2019). Divide and conquer in non-standard problems and the super-efficiency phenomenon. *Ann. Statist.* To appear.
- BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and FERNÁNDEZ-VAL, I. (2017). Conditional quantile processes based on series or many regressors. Preprint. Available at [arXiv:1105.6154](https://arxiv.org/abs/1105.6154).
- CHAO, S.-K., VOLGUSHEV, S. and CHENG, G. (2017). Quantile processes for semi and nonparametric regression. *Electron. J. Stat.* **11** 3272–3331. [MR3708539](#)
- CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125. [MR2667913](#)
- HAGEMANN, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *J. Amer. Statist. Assoc.* **112** 446–456. [MR3646584](#)
- JORDAN, M. I. (2013). On statistics, computation and scalability. *Bernoulli* **19** 1378–1390. [MR3102908](#)

MSC2010 subject classifications. 62F12, 62G15, 62G20.

Key words and phrases. B-spline estimation, conditional distribution function, distributed computing, divide-and-conquer, quantile regression process.

- KATO, K. (2012). Asymptotic normality of Powell's kernel estimator. *Ann. Inst. Statist. Math.* **64** 255–273. [MR2878905](#)
- KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. [MR3248677](#)
- KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. [MR2268657](#)
- LI, R., LIN, D. K. J. and LI, B. (2013). Statistical inference in massive data sets. *Appl. Stoch. Models Bus. Ind.* **29** 399–409. [MR3117826](#)
- POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. [MR0853049](#)
- SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory*. Wiley, New York. [MR0606200](#)
- SENGUPTA, S., VOLGUSHEV, S. and SHAO, X. (2016). A subsampled double bootstrap for massive data. *J. Amer. Statist. Assoc.* **111** 1222–1232. [MR3561944](#)
- SHANG, Z. and CHENG, G. (2017). Computational limits of a distributed algorithm for smoothing spline. *J. Mach. Learn. Res.* **18** Paper No. 108. [MR3725447](#)
- SHI, C., LU, W. and SONG, R. (2017). A massive data framework for M-estimators with cubic-rate. *J. Amer. Statist. Assoc.* To appear. DOI:[10.1080/01621459.2017.1360779](#).
- VOLGUSHEV, S. (2013). Smoothed quantile regression processes for binary response models. Preprint. Available at [arXiv:1302.5644](#).
- VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Supplement to “Distributed inference for quantile regression processes.” DOI:[10.1214/18-AOS1730SUPP](#).
- WHITE, T. (2012). *Hadoop: The Definitive Guide*. O'Reilly Media/Yahoo Press.
- XU, G., SHANG, Z. and CHENG, G. (2016). Optimal tuning for divide-and-conquer kernel ridge regression with massive data. Preprint. Available at [arXiv:1612.05907](#).
- ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Communication-efficient algorithms for statistical optimization. *J. Mach. Learn. Res.* **14** 3321–3363. [MR3144464](#)
- ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. [MR3450540](#)
- ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.* **44** 1400–1437. [MR3519928](#)
- ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. [MR1673277](#)

GAUSSIAN APPROXIMATION OF MAXIMA OF WIENER FUNCTIONALS AND ITS APPLICATION TO HIGH-FREQUENCY DATA¹

BY YUTA KOIKE

University of Tokyo and JST CREST

This paper establishes an upper bound for the Kolmogorov distance between the maximum of a high-dimensional vector of smooth Wiener functionals and the maximum of a Gaussian random vector. As a special case, we show that the maximum of multiple Wiener–Itô integrals with common orders is well approximated by its Gaussian analog in terms of the Kolmogorov distance if their covariance matrices are close to each other and the maximum of the fourth cumulants of the multiple Wiener–Itô integrals is close to zero. This may be viewed as a new kind of fourth moment phenomenon, which has attracted considerable attention in the recent studies of probability. This type of Gaussian approximation result has many potential applications to statistics. To illustrate this point, we present two statistical applications in high-frequency financial econometrics: One is the hypothesis testing problem for the absence of lead-lag effects and the other is the construction of uniform confidence bands for spot volatility.

REFERENCES

- [1] AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton, NJ.
- [2] ALVAREZ, A., PANLOUP, F., PONTIER, M. and SAVY, N. (2012). Estimation of the instantaneous volatility. *Stat. Inference Stoch. Process.* **15** 27–59. [MR2892587](#)
- [3] ANDERSEN, T. G., BOLLERSLEV, T. and DOBREV, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *J. Econometrics* **138** 125–180. [MR2380695](#)
- [4] BENNEDSEN, M., HOUNYO, U., LUNDE, A. and PAKKANEN, M. S. (2016). The local fractional bootstrap. Working paper. Available at <https://arxiv.org/abs/1605.00868>.
- [5] BIBINGER, M., JIRAK, M. and VETTER, M. (2017). Nonparametric change-point analysis of volatility. *Ann. Statist.* **45** 1542–1578. [MR3670188](#)
- [6] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- [7] CHANG, J., QIU, Y., YAO, Q. and ZOU, T. (2017). On the statistical inference for large precision matrices with dependent data. Working paper. Available at <https://arxiv.org/abs/1603.06663>.
- [8] CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. [MR3782380](#)

MSC2010 subject classifications. 60G15, 60H07, 62E17, 62G20.

Key words and phrases. Bootstrap, fourth moment phenomenon, Malliavin calculus, maximum, Stein’s method, uniform confidence bands.

- [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] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. [MR3262468](#)
- [11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. [MR3262461](#)
- [12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Testing many moment inequalities. Working paper. Available at <https://arxiv.org/abs/1312.7614v4>.
- [13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. [MR3350040](#)
- [14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. [MR3565470](#)
- [15] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. [MR3693963](#)
- [16] CHRISTENSEN, K., OOMEN, R. and RENÒ, R. (2016). The drift burst hypothesis. Available at <https://ssrn.com/abstract=2842535>.
- [17] DALALYAN, A. and YOSHIDA, N. (2011). Second-order asymptotic expansion for a non-synchronous covariation estimator. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 748–789. [MR2841074](#)
- [18] DAVIDSON, R. and FLACHAIRE, E. (2008). The wild bootstrap, tamed at last. *J. Econometrics* **146** 162–169. [MR2459651](#)
- [19] DAVIDSON, R. and MACINNON, J. G. (1999). The size distortion of bootstrap tests. *Econometric Theory* **15** 361–376. [MR1704228](#)
- [20] FAN, J. and WANG, Y. (2008). Spot volatility estimation for high-frequency data. *Stat. Interface* **1** 279–288. [MR2476744](#)
- [21] GÖTZE, F. and TIKHOMIROV, A. N. (1999). Asymptotic distribution of quadratic forms. *Ann. Probab.* **27** 1072–1098. [MR1699003](#)
- [22] HAYASHI, T. and KOIKE, Y. (2016). Wavelet-based methods for high-frequency lead-lag analysis. Working paper. Available at <https://arxiv.org/abs/1612.01232>.
- [23] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379. [MR2132731](#)
- [24] HOFFMANN, M., ROSENBAUM, M. and YOSHIDA, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19** 426–461. [MR3037160](#)
- [25] JANSON, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge Univ. Press, Cambridge. [MR1474726](#)
- [26] KANAYA, S. (2017). Uniform convergence rates of kernel-based nonparametric estimators for continuous time diffusion processes: A damping function approach. *Econometric Theory* **33** 874–914. [MR3667040](#)
- [27] KANAYA, S. and KRISTENSEN, D. (2016). Estimation of stochastic volatility models by non-parametric filtering. *Econometric Theory* **32** 861–916. [MR3530455](#)
- [28] KATO, K. and KURISU, D. (2017). Bootstrap confidence bands for spectral estimation of Lévy densities under high-frequency observations. Working paper. Available at <https://arxiv.org/abs/1705.00586>.
- [29] KATO, K. and SASAKI, Y. (2016). Uniform confidence bands in deconvolution with unknown error distribution. Working paper. Available at <https://arxiv.org/abs/1608.02251>.
- [30] KOIKE, Y. (2019). Supplement to “Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data.” DOI:10.1214/18-AOS1731SUPP.

- [31] KRISTENSEN, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* **26** 60–93. [MR2587103](#)
- [32] KUSUOKA, S. and YOSHIDA, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probab. Theory Related Fields* **116** 457–484. [MR1757596](#)
- [33] LEE, S. S. and MYKLAND, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financ. Stud.* **21** 2535–2563.
- [34] MANCINO, M. E. and RECCHIONI, M. C. (2015). Fourier spot volatility estimator: Asymptotic normality and efficiency with liquid and illiquid high-frequency data. *PLoS ONE* **10** 1–33.
- [35] MOSSEL, E., O'DONNELL, R. and OLESZKIEWICZ, K. (2010). Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.* (2) **171** 295–341. [MR2630040](#)
- [36] MYKLAND, P. A. and ZHANG, L. (2008). Inference for volatility-type objects and implications for hedging. *Stat. Interface* **1** 255–278. [MR2476743](#)
- [37] NOURDIN, I. (2013). Lectures on Gaussian approximations with Malliavin calculus. In *Séminaire de Probabilités XLV* (C. Donati-Martin, A. Lejay and A. Rouault, eds.). *Lecture Notes in Math.* **2078** 3–89. Springer, Cham. [MR3185909](#)
- [38] NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. [MR2520122](#)
- [39] NOURDIN, I. and PECCATI, G. (2010). Stein's method meets Malliavin calculus: A short survey with new estimates. In *Recent Development in Stochastic Dynamics and Stochastic Analysis* (J. Duan, S. Luo and C. Wang, eds.). *Interdiscip. Math. Sci.* **8** 207–236. World Sci. Publ., Hackensack, NJ. [MR2807823](#)
- [40] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. [MR2962301](#)
- [41] NOURDIN, I., PECCATI, G. and REINERT, G. (2009). Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.* **257** 593–609. [MR2527030](#)
- [42] NOURDIN, I., PECCATI, G. and REINERT, G. (2010). Invariance principles for homogeneous sums: Universality of Gaussian Wiener chaos. *Ann. Probab.* **38** 1947–1985. [MR2722791](#)
- [43] NOURDIN, I., PECCATI, G. and RÉVEILLAC, A. (2010). Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 45–58. [MR2641769](#)
- [44] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Springer, Berlin. [MR2200233](#)
- [45] NUALART, D. and PECCATI, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. [MR2118863](#)
- [46] PALMES, C. and WOERNER, J. H. C. (2016). A mathematical analysis of the Gumbel test for jumps in stochastic volatility models. *Stoch. Anal. Appl.* **34** 852–881. [MR3529199](#)
- [47] ROBERT, C. Y. and ROSENBAUM, M. (2010). On the limiting spectral distribution of the covariance matrices of time-lagged processes. *J. Multivariate Anal.* **101** 2434–2451. [MR2719873](#)
- [48] SABEL, T. (2014). Simultaneous confidence statements about the diffusion coefficient of an Itô-process with application to spot volatility estimation. Ph.D. thesis, Georg-August Univ. School of Science.
- [49] SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462. [MR3581829](#)
- [50] ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. [MR3718156](#)

- [51] ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. [MR3779697](#)

CAUSAL DANTZIG: FAST INFERENCE IN LINEAR STRUCTURAL EQUATION MODELS WITH HIDDEN VARIABLES UNDER ADDITIVE INTERVENTIONS

BY DOMINIK ROTHENHÄUSLER, PETER BÜHLMANN AND
 NICOLAI MEINSHAUSEN

ETH Zürich

Causal inference is known to be very challenging when only observational data are available. Randomized experiments are often costly and impractical and in instrumental variable regression the number of instruments has to exceed the number of causal predictors. It was recently shown in Peters, Bühlmann and Meinshausen (2016) (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012) that causal inference for the full model is possible when data from distinct observational environments are available, exploiting that the conditional distribution of a response variable is invariant under the correct causal model. Two shortcomings of such an approach are the high computational effort for large-scale data and the assumed absence of hidden confounders. Here, we show that these two shortcomings can be addressed if one is willing to make a more restrictive assumption on the type of interventions that generate different environments. Thereby, we look at a different notion of invariance, namely inner-product invariance. By avoiding a computationally cumbersome reverse-engineering approach such as in Peters, Bühlmann and Meinshausen (2016), it allows for large-scale causal inference in linear structural equation models. We discuss identifiability conditions for the causal parameter and derive asymptotic confidence intervals in the low-dimensional setting. In the case of nonidentifiability, we show that the solution set of causal Dantzig has predictive guarantees under certain interventions. We derive finite-sample bounds in the high-dimensional setting and investigate its performance on simulated datasets.

REFERENCES

- ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141. [MR0331612](#)
- ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. [MR1439312](#)
- ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.
- BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York. [MR0996025](#)
- BOWDEN, R. J. and TURKINGTON, D. A. (1990). *Instrumental Variables*. Cambridge Univ. Press, Cambridge. Reprint of the 1984 original. [MR1113481](#)

MSC2010 subject classifications. Primary 62J99, 62H99; secondary 68T99.

Key words and phrases. Causal inference, structural equation models, high-dimensional consistency.

- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- CHICKERING, D. M. (2003). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554. [MR1991085](#)
- DIDELEZ, V., MENG, S. and SHEEHAN, N. A. (2010). Assumptions of IV methods for observational epidemiology. *Statist. Sci.* **25** 22–40. [MR2741813](#)
- HAUSER, A. and BÜHLMANN, P. (2012). Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.* **13** 2409–2464. [MR2973606](#)
- HAUSER, A. and BÜHLMANN, P. (2015). Jointly interventional and observational data: Estimation of interventional Markov equivalence classes of directed acyclic graphs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 291–318. [MR3299409](#)
- HOYER, P. O., JANZING, D., MOOIJ, J. M., PETERS, J. and SCHÖLKOPF, B. (2009). Nonlinear causal discovery with additive noise models. *Adv. Neural Inf. Process. Syst.* **21** 689–696.
- KEMMEREN, P., SAMEITH, K., VAN DE PASCH, L. A., BENSCHOP, J. J., LENSTRA, T. L., MARGARITIS, T., O’DUIBHIR, E., APWEILER, E., VAN WAGENINGEN, S. et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157** 740–752.
- LEWBEL, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econom. Statist.* **30** 67–80. [MR2899185](#)
- MAATHUIS, M. H., KALISCH, M. and BÜHLMANN, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Statist.* **37** 3133–3164. [MR2549555](#)
- MEINSHAUSEN, N., HAUSER, A., MOOIJ, J. M., PETERS, J., VERSTEEG, P. and BÜHLMANN, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci. USA* **113** 7361–7368.
- NEYMAN, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. [Translated and edited by D. Dabrowska and T. Speed] *Statist. Sci.* **5** (1990) 465–480.
- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2548166](#)
- PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012. With comments and a rejoinder. [MR3557186](#)
- R CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RICHARDSON, T. and ROBINS, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Working Paper 128, Center for the Statistics and the Social Sciences, Univ. Washington Series.
- ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.
- ROTHENHÄUSLER, D., BÜHLMANN, P. and MEINSHAUSEN, N. (2019). Supplement to “Causal Dantzig: Fast inference in linear structural equation models with hidden variables under additive interventions.” DOI:[10.1214/18-AOS1732SUPP](#).
- ROTHENHÄUSLER, D., HEINZE, C., PETERS, J. and MEINSHAUSEN, N. (2015). Backshift: Learning causal cyclic graphs from unknown shift interventions. *Adv. Neural Inf. Process. Syst.* **29** 1513–1521.
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

- SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. [MR2274431](#)
- TIAN, J. and PEARL, J. (2001). Causal discovery from changes. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI)* 512–522.
- VANDERWEELE, T. J. and ROBINS, J. M. (2010). Signed directed acyclic graphs for causal inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 111–127. [MR2751246](#)
- 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](#)
- VERMA, T. and PEARL, J. (1991). Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence (UAI)* 255–270.
- WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* **11** 285–300. [MR0002739](#)
- WANG, L. and TCHETGEN TCHETGEN, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 531–550. [MR3798877](#)
- WRIGHT, P. G. (1928). *The Tariff on Animal and Vegetable Oils*. The Macmillan Company, New York.
- YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. [MR2756192](#)

NONPENALIZED VARIABLE SELECTION IN HIGH-DIMENSIONAL LINEAR MODEL SETTINGS VIA GENERALIZED FIDUCIAL INFERENCE

BY JONATHAN P. WILLIAMS AND JAN HANNIG

University of North Carolina at Chapel Hill

Standard penalized methods of variable selection and parameter estimation rely on the magnitude of coefficient estimates to decide which variables to include in the final model. However, coefficient estimates are unreliable when the design matrix is collinear. To overcome this challenge, an entirely new perspective on variable selection is presented within a generalized fiducial inference framework. This new procedure is able to effectively account for linear dependencies among subsets of covariates in a high-dimensional setting where p can grow almost exponentially in n , as well as in the classical setting where $p \leq n$. It is shown that the procedure very naturally assigns small probabilities to subsets of covariates which include redundancies by way of explicit L_0 minimization. Furthermore, with a typical sparsity assumption, it is shown that the proposed method is consistent in the sense that the probability of the true sparse subset of covariates converges in probability to 1 as $n \rightarrow \infty$, or as $n \rightarrow \infty$ and $p \rightarrow \infty$. Very reasonable conditions are needed, and little restriction is placed on the class of possible subsets of covariates to achieve this consistency result.

REFERENCES

- [1] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. [MR2502648](#)
- [2] BEAUMONT, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160.
- [3] BERGER, J. O. and PERICCHI, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. In *Model Selection. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **38** 135–207. IMS, Beachwood, OH. [MR2000753](#)
- [4] BERK, R. A. (2008). *Statistical Learning from a Regression Perspective*. Springer, New York. [MR2722293](#)
- [5] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. [MR3476618](#)
- [6] BONDELL, H. D. and REICH, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Amer. Statist. Assoc.* **107** 1610–1624. [MR3036420](#)
- [7] 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](#)

MSC2010 subject classifications. 62J05, 62F12, 62A01.

Key words and phrases. Best subset selection, high-dimensional regression, L_0 minimization, feature selection.

- [8] CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- [9] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- [10] GHOSH, J. and GHATTAS, A. E. (2015). Bayesian variable selection under collinearity. *Amer. Statist.* **69** 165–173. [MR3391636](#)
- [11] HANNIG, J., IYER, H., LAI, R. C. S. and LEE, T. C. M. (2016). Generalized fiducial inference: A review and new results. *J. Amer. Statist. Assoc.* **111** 1346–1361. [MR3561954](#)
- [12] JAMESON, G. J. O. (2013). Inequalities for gamma function ratios. *Amer. Math. Monthly* **120** 936–940. [MR3139572](#)
- [13] JOHNSON, V. E. and ROSSELL, D. (2012). Bayesian model selection in high-dimensional settings. *J. Amer. Statist. Assoc.* **107** 649–660. [MR2980074](#)
- [14] LAI, R. C. S., HANNIG, J. and LEE, T. C. M. (2015). Generalized fiducial inference for ultrahigh-dimensional regression. *J. Amer. Statist. Assoc.* **110** 760–772. [MR3367262](#)
- [15] LUO, S. and CHEN, Z. (2013). Extended BIC for linear regression models with diverging number of relevant features and high or ultra-high feature spaces. *J. Statist. Plann. Inference* **143** 494–504. [MR2995110](#)
- [16] NARISSETTY, N. N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. [MR3210987](#)
- [17] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. [MR2854348](#)
- [18] ROČKOVÁ, V. and GEORGE, E. I. (2018). The Spike-and-Slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. [MR3803476](#)
- [19] ROSSELL, D. and TELESKA, D. (2017). Nonlocal priors for high-dimensional estimation. *J. Amer. Statist. Assoc.* **112** 254–265. [MR3646569](#)
- [20] SHIN, M., BHATTACHARYA, A. and JOHNSON, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statist. Sinica* **28** 1053–1078. [MR3791100](#)
- [21] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [22] 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](#)

SUPER-RESOLUTION ESTIMATION OF CYCLIC ARRIVAL RATES

BY NINGYUAN CHEN^{*,1}, DONALD K. K. LEE[†]

AND SAHAND N. NEGAHBAN^{†,2}

HKUST^{} and Yale University[†]*

Exploiting the fact that most arrival processes exhibit cyclic behaviour, we propose a simple procedure for estimating the intensity of a nonhomogeneous Poisson process. The estimator is the super-resolution analogue to Shao (2010) and Shao and Lii [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 99–122], which is a sum of p sinusoids where p and the amplitude and phase of each wave are not known and need to be estimated. This results in an interpretable yet flexible specification that is suitable for use in modelling as well as in high resolution simulations.

Our estimation procedure sits in between classic periodogram methods and atomic/total variation norm thresholding. Through a novel use of window functions in the point process domain, our approach attains super-resolution without semidefinite programming. Under suitable conditions, finite sample guarantees can be derived for our procedure. These resolve some open questions and expand existing results in spectral estimation literature.

REFERENCES

- BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264–296. [MR0171334](#)
- BEBBINGTON, M. and ZITIKIS, R. (2004). A robust heuristic estimator for the period of a Poisson intensity function. *Methodol. Comput. Appl. Probab.* **6** 441–462. [MR2108562](#)
- BELITSER, E., SERRA, P. and VAN ZANTEN, H. (2013). Estimating the period of a cyclic non-homogeneous Poisson process. *Scand. J. Stat.* **40** 204–218. [MR3066411](#)
- BHASKAR, B. N., TANG, G. and RECHT, B. (2013). Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Process.* **61** 5987–5999. [MR3138795](#)
- BROWN, L., CAI, T., ZHANG, R., ZHAO, L. and ZHOU, H. (2010). The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probab. Theory Related Fields* **146** 401–433. [MR2574733](#)
- CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2013). Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19** 1229–1254. [MR3132912](#)
- CHEN, N., LEE, D. K. K. and NEGAHBAN, S. N. (2019). Supplement to “Super-resolution estimation of cyclic arrival rates.” DOI:10.1214/18-AOS1736SUPP.
- CHEN, N., LEE, D. K. K. and SHEN, H. (2018). Can customer arrival rates be modelled by sine waves? Working paper.
- DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. [MR1311089](#)

MSC2010 subject classifications. Primary 62M15, 90B22; secondary 60G55.

Key words and phrases. Spectral estimation, periodogram, window function, thresholding, non-homogeneous Poisson process, queueing theory.

- DUTT, A. and ROKHLIN, V. (1993). Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* **14** 1368–1393. [MR1241591](#)
- FERNANDEZ-GRANDA, C. (2013). Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications* 145–148.
- HELMERS, R. and MANGKU, I. W. (2003). On estimating the period of a cyclic Poisson process. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **42** 345–356. IMS, Beachwood, OH. [MR2138302](#)
- LEWIS, P. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *J. Sound Vibration* **12** 353–375.
- LI, T.-H. (2014). *Time Series with Mixed Spectra*. CRC Press, Boca Raton, FL. [MR3098581](#)
- MOITRA, A. (2015). Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing* 821–830. ACM, New York. [MR3388262](#)
- OSIPOV, A., ROKHLIN, V. and XIAO, H. (2013). *Prolate Spheroidal Wave Functions of Order Zero: Mathematical Tools for Bandlimited Approximation. Applied Mathematical Sciences* **187**. Springer, New York. [MR3136430](#)
- PRABHU, K. (2013). *Window Functions and Their Applications in Signal Processing*. CRC Press, New York.
- RICE, J. A. and ROSENBLATT, M. (1988). On frequency estimation. *Biometrika* **75** 477–484. [MR0967586](#)
- SHAO, N. (2010). Modeling almost periodicity in point processes. Ph.D. thesis, Univ. California, Riverside.
- SHAO, N. and LII, K.-S. (2011). Modelling non-homogeneous Poisson processes with almost periodic intensity functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 99–122. [MR2797738](#)
- TANG, G., BHASKAR, B. N. and RECHT, B. (2015). Near minimax line spectral estimation. *IEEE Trans. Inform. Theory* **61** 499–512. [MR3299978](#)
- VERE-JONES, D. (1982). On the estimation of frequency in point-process data. *J. Appl. Probab.* **19A** 383–394. [MR0633207](#)

SEQUENTIAL MULTIPLE TESTING WITH GENERALIZED ERROR CONTROL: AN ASYMPTOTIC OPTIMALITY THEORY¹

BY YANGLEI SONG AND GEORGIOS FELLOURIS

University of Illinois, Urbana–Champaign

The sequential multiple testing problem is considered under two generalized error metrics. Under the first one, the probability of at least k mistakes, of any kind, is controlled. Under the second, the probabilities of at least k_1 false positives and at least k_2 false negatives are simultaneously controlled. For each formulation, the optimal expected sample size is characterized, to a first-order asymptotic approximation as the error probabilities go to 0, and a novel multiple testing procedure is proposed and shown to be asymptotically efficient under every signal configuration. These results are established when the data streams for the various hypotheses are independent and each local log-likelihood ratio statistic satisfies a certain strong law of large numbers. In the special case of i.i.d. observations in each stream, the gains of the proposed sequential procedures over fixed-sample size schemes are quantified.

REFERENCES

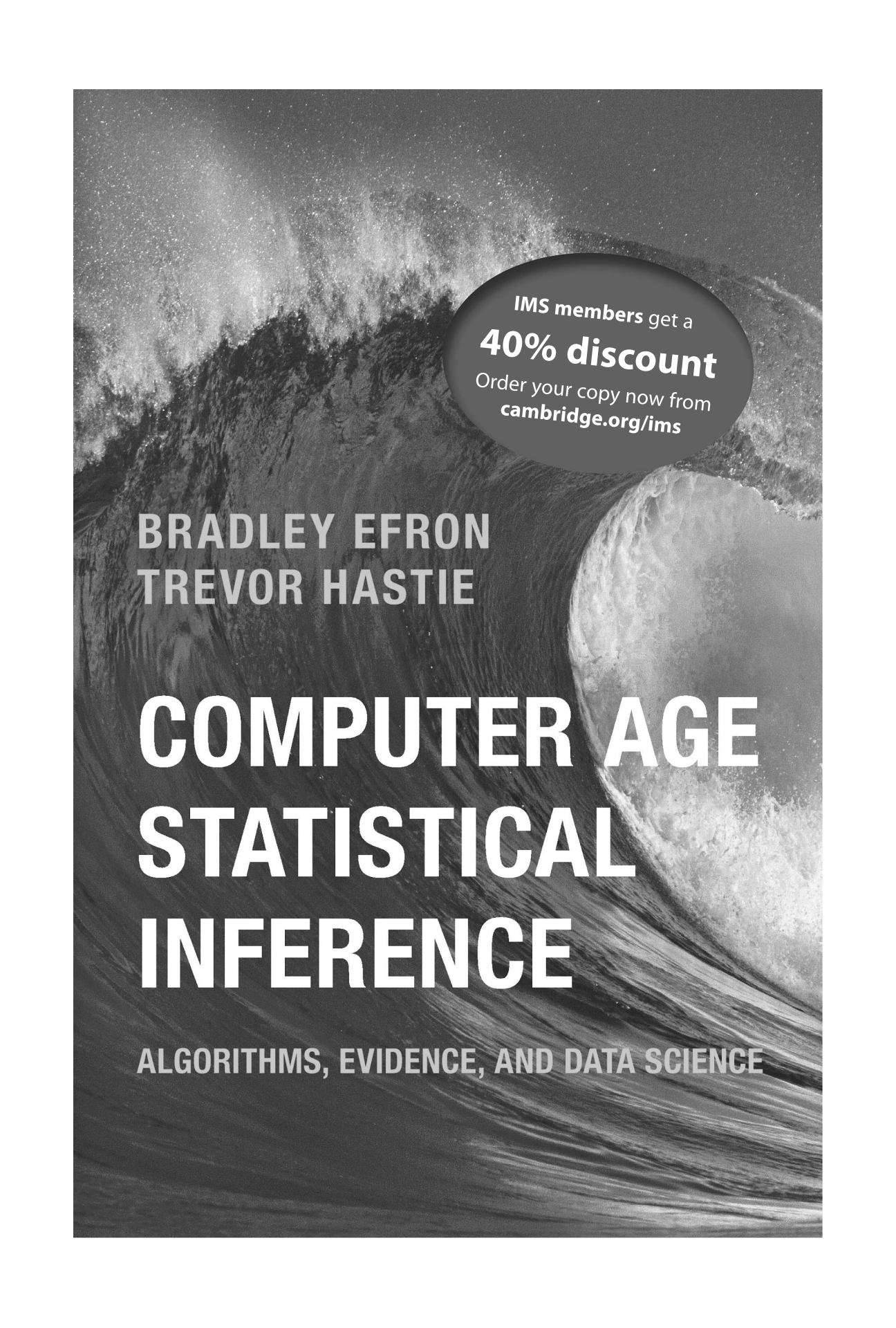
- [1] BARTROFF, J. (2018). Multiple hypothesis tests controlling generalized error rates for sequential data. *Statist. Sinica* **28** 363–398. [MR3752265](#)
- [2] BARTROFF, J. and LAI, T. L. (2010). Multistage tests of multiple hypotheses. *Comm. Statist. Theory Methods* **39** 1597–1607.
- [3] BARTROFF, J. and SONG, J. (2014). Sequential tests of multiple hypotheses controlling type I and II familywise error rates. *J. Statist. Plann. Inference* **153** 100–114. [MR3229025](#)
- [4] 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](#)
- [5] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245](#)
- [6] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. [MR2850212](#)
- [7] DE, S. K. and BARON, M. (2012). Sequential Bonferroni methods for multiple hypothesis testing with strong control of family-wise error rates I and II. *Sequential Anal.* **31** 238–262. [MR2911288](#)
- [8] DE, S. K. and BARON, M. (2012). Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *J. Statist. Plann. Inference* **142** 2059–2070. [MR2903412](#)
- [9] DE, S. K. and BARON, M. (2015). Sequential tests controlling generalized familywise error rates. *Stat. Methodol.* **23** 88–102. [MR3278804](#)

MSC2010 subject classifications. 62L10.

Key words and phrases. Multiple testing, sequential analysis, asymptotic optimality, generalized familywise error rates, misclassification rate.

- [10] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. Springer, New York. [MR1619036](#)
- [11] FORESTI, G. L., REGAZZONI, C. S. and VARSHNEY, P. K. (2003). *Multisensor Surveillance Systems: The Fusion Perspective*. Springer, New York.
- [12] GUO, W., HE, L. and SARKAR, S. K. (2014). Further results on controlling the false discovery proportion. *Ann. Statist.* **42** 1070–1101. [MR3210996](#)
- [13] HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. [MR0538597](#)
- [14] HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75** 383–386.
- [15] HOMMEL, G. and HOFFMANN, T. (1988). Controlled uncertainty. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing* 154–161. Springer, Berlin.
- [16] HSU, P. L. and ROBBINS, H. (1947). Complete convergence and the law of large numbers. *Proc. Natl. Acad. Sci. USA* **33** 25–31. [MR0019852](#)
- [17] KITTUR, A., CHI, E. H. and SUH, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 453–456. ACM, New York.
- [18] LEHMANN, E. L. and ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. [MR2195631](#)
- [19] LEHMANN, E. L., ROMANO, J. P. and SHAFFER, J. P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33** 1084–1108. [MR2195629](#)
- [20] LI, Y., NITINAWARAT, S. and VEERAVALLI, V. V. (2014). Universal outlier hypothesis testing. *IEEE Trans. Inform. Theory* **60** 4066–4082. [MR3225950](#)
- [21] MALLOY, M. L. and NOWAK, R. D. (2014). Sequential testing for sparse recovery. *IEEE Trans. Inform. Theory* **60** 7862–7873. [MR3285750](#)
- [22] MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. [MR0468056](#)
- [23] PEÑA, E. A., HABIGER, J. D. and WU, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* **39** 556–583. [MR2797856](#)
- [24] RAPPAPORT, T. S. (1996). *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, Upper Saddle River, NJ.
- [25] ROMANO, J. P. and SHAIKH, A. M. (2006). Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.* **34** 1850–1873. [MR2283720](#)
- [26] ROMANO, J. P. and WOLF, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. [MR2351090](#)
- [27] SONG, Y. and FELLOURIS, G. (2016). Logarithmically efficient simulation for misclassification probabilities in sequential multiple testing. In *2016 Winter Simulation Conference (WSC)* 314–325.
- [28] SONG, Y. and FELLOURIS, G. (2017). Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electron. J. Stat.* **11** 338–363. [MR3606774](#)
- [29] SONG, Y. and FELLOURIS, G. (2019). Supplement to “Sequential multiple testing with generalized error control: An asymptotic optimality theory.” DOI:[10.1214/18-AOS1737SUPP](#).
- [30] STOREY, J. D. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 347–368. [MR2323757](#)
- [31] SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. [MR2649603](#)
- [32] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. [MR3241619](#)

- [33] TARTAKOVSKY, A. G. (1998). Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case. *Stat. Inference Stoch. Process.* **1** 265–295. [MR2797137](#)
- [34] TARTAKOVSKY, A. G., LI, X. R. and YARALOV, G. (2003). Sequential detection of targets in multichannel systems. *IEEE Trans. Inform. Theory* **49** 425–445. [MR1966790](#)
- [35] WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. [MR0013275](#)



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

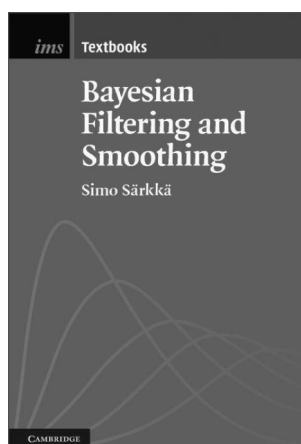
COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE



The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Bayesian Filtering and Smoothing

Simo Särkkä

Filtering and smoothing methods are used to produce an accurate estimate of the state of a time-varying system based on multiple observational inputs (data). Interest in these methods has exploded in recent years, with numerous applications emerging in fields such as navigation, aerospace engineering, telecommunications and medicine. This compact, informal introduction for graduate students and advanced undergraduates presents the current state-of-the-art filtering and smoothing methods in a unified Bayesian framework. Readers learn what non-linear Kalman filters and particle filters are, how they are related, and their relative advantages and disadvantages. They also discover how state-of-the-art Bayesian parameter estimation methods can be combined with state-of-the-art filtering and smoothing algorithms. The book's practical and algorithmic approach assumes only modest mathematical prerequisites. Examples include MATLAB computations, and the numerous end-of-chapter exercises include computational assignments. MATLAB/GNU Octave source code is available for download at www.cambridge.org/sarkka, promoting hands-on work with the methods.

**IMS member? Claim
your 40% discount:
www.cambridge.org/ims**

**Hardback US\$59.40
Paperback US\$22.19**

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.