

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

- Exact recovery in the Ising blockmodel
QUENTIN BERTHET, PHILIPPE RIGOLLET AND PIYUSH SRIVASTAVA 1805
- Maximum likelihood estimation in Gaussian models under total positivity
STEFFEN LAURITZEN, CAROLINE UHLER AND PIOTR ZWIERNIK 1835
- Maximum likelihood estimation in transformed linear regression with nonnormal errors
XINGWEI TONG, FUQING GAO, KANI CHEN, DINGJIAO CAI AND JIANGUO SUN 1864
- Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates SIVARAMAN BALAKRISHNAN AND LARRY WASSERMAN 1893
- The BLUE in continuous-time regression models with correlated errors
HOLGER DETTE, ANDREY PEPELYSHEV AND ANATOLY ZHIGLJAVSKY 1928
- Adaptive-to-model checking for regressions with diverging number of predictors
FALONG TAN AND LIXING ZHU 1960
- Nonparametric screening under conditional strictly convex loss for ultrahigh dimensional sparse data XU HAN 1995
- Local stationarity and time-inhomogeneous Markov chains LIONEL TRUQUET 2023
- High-dimensional change-point detection under sparse alternatives
FARIDA ENIKEEVA AND ZAID HARCHAOUTI 2051
- Perturbation bootstrap in adaptive Lasso
DEBRAJ DAS, KARL GREGORY AND S. N. LAHIRI 2080
- Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions PIERRE ALQUIER, VINCENT COTTET AND GUILLAUME LECUÉ 2117
- Cross validation for locally stationary processes
STEFAN RICHTER AND RAINER DAHLHAUS 2145
- Generalized cluster trees and singular measures YEN-CHI CHEN 2174
- Spectral method and regularized MLE are both optimal for top- K ranking
YUXIN CHEN, JIANQING FAN, CONG MA AND KAIZHENG WANG 2204
- Negative association, ordering and convergence of resampling methods
MATHIEU GERBER, NICOLAS CHOPIN AND NICK WHITELEY 2236
- On deep learning as a remedy for the curse of dimensionality in nonparametric regression BENEDIKT BAUER AND MICHAEL KOHLER 2261
- Convergence rates of least squares regression estimators with heavy-tailed errors
QIYANG HAN AND JON A. WELLNER 2286
- Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression QIAN QIN AND JAMES P. HOBERT 2320
- On testing conditional qualitative treatment effects
CHENGCHUN SHI, RUI SONG AND WENBIN LU 2348
- Dynamic network models and graphon estimation MARIANNA PENSKEY 2378

THE ANNALS OF STATISTICS

Vol. 47, No. 4, pp. 1805–2403 August 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, New Jersey 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 4, August 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.

EXACT RECOVERY IN THE ISING BLOCKMODEL

BY QUENTIN BERTHET¹, PHILIPPE RIGOLLET² AND PIYUSH SRIVASTAVA³

*University of Cambridge, Massachusetts Institute of Technology and
 Tata Institute of Fundamental Research*

We consider the problem associated to recovering the block structure of an Ising model given independent observations on the binary hypercube. This new model, called the Ising blockmodel, is a perturbation of the mean field approximation of the Ising model known as the Curie–Weiss model: the sites are partitioned into two blocks of equal size and the interaction between those of the same block is stronger than across blocks, to account for more order within each block. We study probabilistic, statistical and computational aspects of this model in the high-dimensional case when the number of sites may be much larger than the sample size.

REFERENCES

- ABBE, E. (2017). Community detection and stochastic block models: Recent developments. Preprint. Available at [arXiv:1703.10146](https://arxiv.org/abs/1703.10146).
- ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. [MR3447993](https://arxiv.org/abs/1601.02658)
- ABBE, E. and SANDON, C. (2015). Detection in the stochastic block model with multiple clusters: Proof of the achievability conjectures, acyclic BP, and the information-computation gap. Preprint. Available at [arXiv:1512.09080](https://arxiv.org/abs/1512.09080).
- ABBE, E. and SANDON, C. (2016a). Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 1334–1342. Curran Associates, Inc., New York.
- ABBE, E. and SANDON, C. (2016b). Crossing the KS threshold in the stochastic block model with information theory. In *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)* 840–844. IEEE, New York.
- ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. In *Proceedings of the 1998 ACM–SIAM Symposium on Discrete Algorithms* 594–598. SIAM, Philadelphia, PA.
- 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.
- BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. Preprint. Available at [arXiv:1601.02658](https://arxiv.org/abs/1601.02658).
- BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2019). Supplement to “Exact recovery in the Ising blockmodel.” DOI:[10.1214/17-AOS1620SUPP](https://doi.org/10.1214/17-AOS1620SUPP).
- BESAG, J. (1986). On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

MSC2010 subject classifications. Primary 62H30; secondary 82B20.

Key words and phrases. Ising blockmodel, Curie–Weiss, stochastic blockmodel, planted partition, spectral partitioning.

- BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575](#)
- BRESLER, G. (2015). Efficiently learning Ising models on arbitrary graphs [extended abstract]. In *Proceedings of the 2015 ACM Symposium on Theory of Computing* 771–782. ACM, New York.
- BRESLER, G., GAMARNIK, D. and SHAH, D. (2014). Learning graphical models from the glauher dynamics. Preprint. Available at [arXiv:1410.7659](#).
- BRESLER, G., MOSSEL, E. and SLY, A. (2008). Reconstruction of Markov random fields from samples: Some observations and algorithms. In *Approximation, Randomization and Combinatorial Optimization. Lecture Notes in Computer Science* **5171** 343–356. Springer, Berlin.
- BUNEA, F., GIRAUD, C. and LUO, X. (2015). Minimax optimal variable clustering in G -models via Cord. Preprint. Available at [arXiv:1508.01939](#).
- BUNEA, F., GIRAUD, C., ROYER, M. and VERZELEN, N. (2016). PECOK: A convex optimization approach to variable clustering. Preprint. Available at [arXiv:1606.05100](#).
- DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* (3) **84** 066106.
- DIACONIS, P., GOEL, S. and HOLMES, S. (2008). Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat.* **2** 777–807.
- DYER, M. E. and FRIEZE, A. M. (1989). The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489.
- FEDELE, M. and UNGUENDOLI, F. (2012). Rigorous results on the bipartite mean-field model. *J. Phys. A* **45** 385001. [MR2970551](#)
- FEIGE, U. and KRAUTHGAMER, R. (2002). A polylogarithmic approximation of the minimum bi-section. *SIAM J. Comput.* **31** 1090–1118 (electronic).
- GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2015). Achieving optimal misclassification proportion in stochastic block model. Preprint. Available at [arXiv:1505.03772](#).
- GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2016). Community detection in degree-corrected block models. Preprint. Available at [arXiv:1607.06993](#).
- GAREY, M. R., JOHNSON, D. S. and STOCKMEYER, L. (1976). Some simplified NP-complete graph problems. *Theoret. Comput. Sci.* **1** 237–267.
- GOEMANS, M. X. and WILLIAMSON, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.* **42** 1115–1145.
- HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inform. Theory* **62** 2788–2797.
- HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137.
- ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31** 253–258.
- LAURENT, M. and POLJAK, S. (1996). On the facial structure of the set of correlation matrices. *SIAM J. Matrix Anal. Appl.* **17** 530–547.
- LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. [MR1419991](#)
- LAURITZEN, S. L. and SHEEHAN, N. A. (2003). Graphical models for genetic analyses. *Statist. Sci.* **18** 489–514. [MR2059327](#)
- LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.* **2017** 073403. [MR3683819](#)
- LI, L., LU, P. and YIN, Y. (2012). Correlation decay up to uniqueness in spin systems. In *Proceedings of the Twenty-Fourth Annual ACM–SIAM Symposium on Discrete Algorithms* 67–84. SIAM, Philadelphia, PA. [MR3185380](#)

- MANNING, C. D. and SCHÜTZE, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.
- MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, New York.
- MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 529–537. IEEE Computer Soc., Los Alamitos, CA.
- MOITRA, A., PERRY, W. and WEIN, A. S. (2016). How robust are reconstruction thresholds for community detection? In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* 828–841.
- MONTANARI, A. and SABERI, A. (2010). The spread of innovations in social networks. *Proc. Natl. Acad. Sci. USA* **107** 20196–20201.
- MOSSEL, E., NEEMAN, J. and SLY, A. (2013). A proof of the block model threshold conjecture. Preprint. Available at [arXiv:1311.4115](https://arxiv.org/abs/1311.4115).
- MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461.
- MOSSEL, E., NEEMAN, J. and SLY, A. (2016). Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.* **26** 2211–2256. [MR3543895](https://arxiv.org/abs/1508.07908)
- MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2016). Global testing against sparse alternatives under Ising models. Preprint. Available at [arXiv:1611.08293](https://arxiv.org/abs/1611.08293).
- 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.
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](https://arxiv.org/abs/1008.3583)
- SCHNEIDMAN, E., BERRY, M. J., SEGEV, R. and BIALEK, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440** 1007–1012.
- SEBASTIANI, P., RAMONI, M. F., NOLAN, V., BALDWIN, C. T. and STEINBERG, M. H. (2005). Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat. Genet.* **37** 435–440.
- SINCLAIR, A., SRIVASTAVA, P. and THURLEY, M. (2014). Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.* **155** 666–686.
- SLY, A. and SUN, N. (2014). Counting in two-spin models on d -regular graphs. *Ann. Probab.* **42** 2383–2416. [MR3265170](https://arxiv.org/abs/1306.2808)
- TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.
- TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. [MR2724359](https://arxiv.org/abs/0904.2795)
- WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. [MR3546438](https://arxiv.org/abs/1506.02531)
- WEITZ, D. (2006). Counting independent sets up to the tree threshold. In *Proceedings of the 2006 ACM Symposium on the Theory of Computing* 140–149. ACM, New York.

MAXIMUM LIKELIHOOD ESTIMATION IN GAUSSIAN MODELS UNDER TOTAL POSITIVITY

BY STEFFEN LAURITZEN, CAROLINE UHLER¹ AND PIOTR ZWIERNIK²

*Massachusetts Institute of Technology, University of Copenhagen and
 Universitat Pompeu Fabra*

We analyze the problem of maximum likelihood estimation for Gaussian distributions that are multivariate totally positive of order two (MTP₂). By exploiting connections to phylogenetics and single-linkage clustering, we give a simple proof that the maximum likelihood estimator (MLE) for such distributions exists based on $n \geq 2$ observations, irrespective of the underlying dimension. Slawski and Hein [*Linear Algebra Appl.* **473** (2015) 145–179], who first proved this result, also provided empirical evidence showing that the MTP₂ constraint serves as an implicit regularizer and leads to sparsity in the estimated inverse covariance matrix, determining what we name the ML graph. We show that we can find an upper bound for the ML graph by adding edges corresponding to correlations in excess of those explained by the maximum weight spanning forest of the correlation matrix. Moreover, we provide globally convergent coordinate descent algorithms for calculating the MLE under the MTP₂ constraint which are structurally similar to iterative proportional scaling. We conclude the paper with a discussion of signed MTP₂ distributions.

REFERENCES

- [1] ANANDKUMAR, A., TAN, V. Y. F., HUANG, F. and WILLSKY, A. S. (2012). High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion. *J. Mach. Learn. Res.* **13** 2293–2337. [MR2973603](#)
- [2] BARTOLUCCI, F. and BESAG, J. (2002). A recursive algorithm for Markov random fields. *Biometrika* **89** 724–730. [MR1929176](#)
- [3] BARTOLUCCI, F. and FORCINA, A. (2000). A likelihood ratio test for MTP₂ within binary variables. *Ann. Statist.* **28** 1206–1218. [MR1811325](#)
- [4] BHATTACHARYA, B. (2012). Covariance selection and multivariate dependence. *J. Multivariate Anal.* **106** 212–228.
- [5] BØLVIKEN, E. (1982). Probability inequalities for the multivariate normal with nonnegative partial correlations. *Scand. J. Stat.* **9** 49–58. [MR0651862](#)
- [6] BUHL, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* **20** 263–270.
- [7] CHOI, M. J., TAN, V. Y. F., ANANDKUMAR, A. and WILLSKY, A. S. (2011). Learning latent tree graphical models. *J. Mach. Learn. Res.* **12** 1771–1812.
- [8] COLANGELO, A., SCARSINI, M. and SHAKED, M. (2005). Some notions of multivariate positive dependence. *Insurance Math. Econom.* **37** 13–26.

MSC2010 subject classifications. Primary 60E15, 62H99; secondary 15B48.

Key words and phrases. MTP₂ distribution, attractive Gaussian Markov random field (GMRF), nonfrustrated GRMF, Gaussian graphical model, inverse M-matrix, ultrametric.

- [9] DELLACHERIE, C., MARTINEZ, S. and SAN MARTIN, J. (2014). *Inverse M-Matrices and Ultrametric Matrices* **2118**. Springer, Berlin.
- [10] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.
- [11] DJOLONGA, J. and KRAUSE, A. (2015). Scalable variational inference in log-supermodular models. In *International Conference on Machine Learning (ICML)*.
- [12] EGILMEZ, H. E., PAVEZ, E. and ORTEGA, A. (2016). Graph learning from data under structural and Laplacian constraints. Available at [arXiv:1611.0518](https://arxiv.org/abs/1611.0518).
- [13] FALLAT, S., LAURITZEN, S. L., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Ann. Statist.* **45** 1152–1184.
- [14] FELSENSTEIN, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25** 471–492.
- [15] FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103.
- [16] GOWER, J. C. and ROSS, G. J. S. (1969). Minimum spanning trees and single linkage cluster analysis. *Appl. Statist.* **18** 54–61. [MR0242315](#)
- [17] GRANT, M. and BOYD, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. Available at <http://cvxr.com/cvx>.
- [18] GROSS, E. and SULLIVANT, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli* **24** 386–407.
- [19] HØJSGAARD, S., EDWARDS, D. and LAURITZEN, S. (2012). *Graphical Models with R*. Springer, New York.
- [20] JOHNSON, C. R. and SMITH, R. L. (1996). The completion problem for M -matrices and inverse M -matrices. *Linear Algebra Appl.* **241–243** 655–667.
- [21] JOHNSON, C. R. and SMITH, R. L. (1999). Path product matrices. *Linear and Multilinear Algebra* **46** 177–191.
- [22] JOHNSON, C. R. and SMITH, R. L. (2011). Inverse M -matrices, II. *Linear Algebra Appl.* **435** 953–983. [MR2807211](#)
- [23] KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498.
- [24] KARLIN, S. and RINOTT, Y. (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Ann. Statist.* **9** 1035–1049. [MR0628759](#)
- [25] KARLIN, S. and RINOTT, Y. (1983). M -matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.* **52** 419–438.
- [26] LAURITZEN, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford.
- [27] LEDERMANN, W. (1940). On a problem concerning matrices with variable diagonal elements. *Proc. Roy. Soc. Edinburgh Sect. A* **60** 1–17. [MR0001206](#)
- [28] LUO, Z. Q. and TSENG, P. (1992). On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.* **72** 7–35.
- [29] MALIOUTOV, D. M., JOHNSON, J. K. and WILLSKY, A. S. (2006). Walk-sums and belief propagation in Gaussian graphical models. *J. Mach. Learn. Res.* **7** 2031–2064.
- [30] MALLE, B. F. and HOROWITZ, L. M. (1995). The puzzle of negative self-views: An exploration using the schema concept. *J. Pers. Soc. Psychol.* **68** 470.
- [31] NEWMAN, C. M. (1983). A general central limit theorem for FKG systems. *Comm. Math. Phys.* **91** 75–80.
- [32] OSTROWSKI, A. (1937). Über die Determinanten mit überwiegender Hauptdiagonale. *Comment. Math. Helv.* **10** 69–96. [MR1509568](#)
- [33] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)* **9** 223–252. [MR1611693](#)

- [34] SEMPLE, C. and STEEL, M. A. (2003). *Phylogenetics* **24**. Oxford Univ. Press, London.
- [35] SHAPIRO, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Stat. Rev.* **56** 49–62.
- [36] SHIERS, N., ZWIERNIK, P., ASTON, J. and SMITH, J. Q. (2016). The correlation space of Gaussian latent tree models and model selection without fitting. *Biometrika* **103** 531–545.
- [37] SLAWSKI, M. and HEIN, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra Appl.* **473** 145–179.
- [38] SPEARMAN, C. (1928). The abilities of man. *Science* **68** 38.
- [39] SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150. [MR0829559](#)
- [40] UHLER, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.* **40** 238–261. [MR3014306](#)
- [41] WERMUTH, N. and SCHEIDT, E. (1977). Algorithm AS 105: Fitting a covariance selection model to a matrix. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **26** 88–92.
- [42] ZWIERNIK, P. (2016). *Semialgebraic Statistics and Latent Tree Models. Monographs on Statistics and Applied Probability* **146**. Chapman & Hall/CRC, Boca Raton, FL. [MR3379921](#)

MAXIMUM LIKELIHOOD ESTIMATION IN TRANSFORMED LINEAR REGRESSION WITH NONNORMAL ERRORS

BY XINGWEI TONG^{*,1}, FUQING GAO^{†,2}, KANI CHEN[‡], DINGJIAO CAI[§] AND
 JIANGUO SUN[¶]

Beijing Normal University^{}, Wuhan University[†], Hong Kong University of
 Science and Technology[‡], Henan University of Economics and Law[§] and
 University of Missouri[¶]*

This paper discusses the transformed linear regression with non-normal error distributions, a problem that often occurs in many areas such as economics and social sciences as well as medical studies. The linear transformation model is an important tool in survival analysis partly due to its flexibility. In particular, it includes the Cox model and the proportional odds model as special cases when the error follows the extreme value distribution and the logistic distribution, respectively. Despite the popularity and generality of linear transformation models, however, there is no general theory on the maximum likelihood estimation of the regression parameter and the transformation function. One main difficulty for this is that the transformation function near the tails diverges to infinity and can be quite unstable. It affects the accuracy of the estimation of the transformation function and regression parameters. In this paper, we develop the maximum likelihood estimation approach and provide the near optimal conditions on the error distribution under which the consistency and asymptotic normality of the resulting estimators can be established. Extensive numerical studies suggest that the methodology works well, and an application to the data on a typhoon forecast is provided.

REFERENCES

- [1] BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252. [MR0192611](#)
- [2] CHEN, K., JIN, Z. and YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89** 659–668. [MR1929170](#)
- [3] CHEN, K. and TONG, X. (2010). Varying coefficient transformation models with censored data. *Biometrika* **97** 969–976. [MR2746165](#)
- [4] CHEN, S. (2002). Rank estimation of transformation models. *Econometrica* **70** 1683–1697. [MR1929984](#)
- [5] CHENG, S. C., WEI, L. J. and YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82** 835–845. [MR1380818](#)
- [6] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. [MR0341758](#)
- [7] DIAO, G., ZENG, D. and YANG, S. (2013). Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics* **69** 840–849. [MR3146780](#)

MSC2010 subject classifications. 62F12.

Key words and phrases. Linear transformation model, maximum likelihood estimation.

- [8] DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15** 325–345. [MR0885740](#)
- [9] FINE, J. P., YING, Z. and WEI, L. J. (1998). On the linear transformation model with censored observations. *Biometrika* **85** 980–986.
- [10] HAN, A. K. (1987). A nonparametric analysis of transformations. *J. Econometrics* **35** 191–209. [MR0903182](#)
- [11] HOROWITZ, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* **64** 103–137. [MR1366143](#)
- [12] HUANG, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563. [MR1742499](#)
- [13] KHAN, S. and TAMER, E. (2007). Partial rank estimation of duration models with general forms of censoring. *J. Econometrics* **136** 251–280. [MR2328593](#)
- [14] MA, S. and KOSOROK, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Ann. Statist.* **33** 2256–2290. [MR2211086](#)
- [15] MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731. [MR1292537](#)
- [16] MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23** 182–198. [MR1331663](#)
- [17] MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. [MR1803168](#)
- [18] SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. [MR1201705](#)
- [19] TONG, X., GAO, F., CHEN, K., CAI, D. and SUN, J. (2019). Supplement to “Maximum likelihood estimation in transformed linear regression with nonnormal errors.” DOI:10.1214/18-AOS1726SUPP.
- [20] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- [21] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.
- [22] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics. Springer Series in Statistics*. Springer, New York. [MR1385671](#)
- [23] ZENG, D. and LIN, D. Y. (2006). Maximum likelihood estimation in semiparametric transformation models for counting processes. *Biometrika* **93** 627–640.
- [24] ZENG, D. and LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 507–564. [MR2370068](#)
- [25] ZENG, D. L., MICHAEL, R. and KOSOROK, L. D. Y. (2014). Nonparametric maximum likelihood estimation in linear regression with unknown transformation. Unpublished manuscript.

HYPOTHESIS TESTING FOR DENSITIES AND HIGH-DIMENSIONAL MULTINOMIALS: SHARP LOCAL MINIMAX RATES¹

BY SIVARAMAN BALAKRISHNAN AND LARRY WASSERMAN

Carnegie Mellon University

We consider the goodness-of-fit testing problem of distinguishing whether the data are drawn from a specified distribution, versus a composite alternative separated from the null in the total variation metric. In the discrete case, we consider goodness-of-fit testing when the null distribution has a possibly growing or unbounded number of categories. In the continuous case, we consider testing a Hölder density with exponent $0 < s \leq 1$, with possibly unbounded support, in the low-smoothness regime where the Hölder parameter is not assumed to be constant. In contrast to existing results, we show that the minimax rate and critical testing radius in these settings depend strongly, and in a precise way, on the null distribution being tested and this motivates the study of the (local) minimax rate as a function of the null distribution. For multinomials, the local minimax rate has been established in recent work. We revisit and extend these results and develop two modifications to the χ^2 -test whose performance we characterize. For testing Hölder densities, we show that the usual binning tests are inadequate in the low-smoothness regime and we design a spatially adaptive partitioning scheme that forms the basis for our locally minimax optimal tests. Furthermore, we provide the first local minimax lower bounds for this problem which yield a sharp characterization of the dependence of the critical radius on the null hypothesis being tested. In the low-smoothness regime, we also provide adaptive tests that adapt to the unknown smoothness parameter. We illustrate our results with a variety of simulations that demonstrate the practical utility of our proposed tests.

REFERENCES

- [1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. [MR2722464](#)
- [2] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. [MR2797847](#)
- [3] ARIAS-CASTRO, E., PELLETIER, B. and SALIGRAMA, V. (2018). Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *J. Nonparametr. Stat.* **30** 448–471. [MR3794401](#)
- [4] BALAKRISHNAN, S. and WASSERMAN, L. (2019). Supplement to “Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates.” DOI:[10.1214/18-AOS1729SUPP](#).
- [5] BALAKRISHNAN, S. and WASSERMAN, L. (2018). Hypothesis testing for high-dimensional multinomials: A selective review. *Ann. Appl. Stat.* To appear.

MSC2010 subject classifications. 60K35.

Key words and phrases. Local-minimax, nonparametric goodness-of-fit testing.

- [6] BARRON, A. R. (1989). Uniformly powerful goodness of fit tests. *Ann. Statist.* **17** 107–124. [MR0981439](#)
- [7] BATU, T., FISCHER, E., FORTNOW, L., KUMAR, R., RUBINFELD, R. and WHITE, P. (2001). Testing random variables for independence and identity. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 442–451. IEEE Computer Soc., Los Alamitos, CA. [MR1948733](#)
- [8] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. [MR3127849](#)
- [9] CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. [MR3379081](#)
- [10] CARPENTIER, A. (2015). Testing the regularity of a smooth signal. *Bernoulli* **21** 465–488. [MR3322327](#)
- [11] CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*. Duxbury, Pacific Grove, CA.
- [12] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. [MR3269982](#)
- [13] CRAMÉR, H. (1928). On the composition of elementary errors. *Scand. Actuar. J.* **1928** 13–74.
- [14] DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York. [MR0780746](#)
- [15] DIACONIS, P. and MOSTELLER, F. (2006). Methods for studying coincidences. In *Selected Papers of Frederick Mosteller* (S. E. Fienberg and D. C. Hoaglin, eds.) 605–622. Springer, New York.
- [16] DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* 685–694. IEEE Computer Soc., Los Alamitos, CA. [MR3631031](#)
- [17] FIENBERG, S. E. (1979). The use of chi-squared statistics for categorical data problems. *J. Roy. Statist. Soc. Ser. B* **41** 54–64. [MR0535545](#)
- [18] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics **40**. Cambridge Univ. Press, New York. [MR3588285](#)
- [19] GOLDBREICH, O. and RON, D. (2011). On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Lecture Notes in Computer Science* **6650** 68–75. Springer, Heidelberg. [MR2844253](#)
- [20] INGSTER, Y. I. (1990). Minimax detection of a signal in ℓ_p -metrics. *J. Math. Sci.* **68** 503–515.
- [21] INGSTER, Y. I. (1997). Adaptive chi-square tests. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **244** 150–166, 333. [MR1700386](#)
- [22] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Lecture Notes in Statistics **169**. Springer, New York. [MR1991446](#)
- [23] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. [MR2747131](#)
- [24] LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. [MR0334381](#)
- [25] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. [MR1639875](#)
- [26] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. [MR2135927](#)
- [27] MARRIOTT, P., SABOLOVA, R., VAN BEVER, G. and CRITCHLEY, F. (2015). Geometry of goodness-of-fit testing in high dimensional low sample size modelling. In *Geometric Science of Information. Lecture Notes in Computer Science* **9389** 569–576. Springer, Cham. [MR3442239](#)
- [28] MORRIS, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188. [MR0370871](#)

- [29] NEYMAN, J. and PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. Lond. Ser. A* **231** 289–337.
- [30] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. [MR3161450](#)
- [31] PANINSKI, L. (2008). A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inform. Theory* **54** 4750–4755. [MR2591136](#)
- [32] PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Ser. 5* **50** 157–175.
- [33] READ, T. R. C. and CRESSIE, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York. [MR0955054](#)
- [34] RON, D. (2008). Property testing: A learning theory perspective. *Found. Trends Mach. Learn.* **1** 307–402.
- [35] SMIRNOFF, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow Univ. Math. Bull.* **2** 3–14. [MR0002062](#)
- [36] SNEDECOR, G. W. and COCHRAN, W. G. (1980). *Statistical Methods*, 7th ed. Iowa State Univ. Press, Ames, IA. [MR0614143](#)
- [37] VALIANT, G. and VALIANT, P. (2014). An automatic inequality prover and instance optimal identity testing. In *55th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2014* 51–60. IEEE Computer Soc., Los Alamitos, CA. [MR3344854](#)
- [38] VON MISES, R. (1951). *Wahrscheinlichkeit, Statistik und Wahrheit*. Springer, Vienna. [MR0041364](#)
- [39] WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.

THE BLUE IN CONTINUOUS-TIME REGRESSION MODELS WITH CORRELATED ERRORS¹

BY HOLGER DETTE*, ANDREY PEPELYSHEV^{†,2} AND
 ANATOLY ZHIGLJAVSKY[†]

Ruhr-Universität Bochum and Cardiff University[†]*

In this paper, the problem of best linear unbiased estimation is investigated for continuous-time regression models. We prove several general statements concerning the explicit form of the best linear unbiased estimator (BLUE), in particular when the error process is a smooth process with one or several derivatives of the response process available for construction of the estimators. We derive the explicit form of the BLUE for many specific models including the cases of continuous autoregressive errors of order two and integrated error processes (such as integrated Brownian motion). The results are illustrated on many examples.

REFERENCES

- ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. [MR2604700](#)
- ANDERSON, T. W. (1970). Efficient estimation of regression coefficients in time series. Technical report, DTIC document.
- BERLINET, A. and THOMAS-AGNAN, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, Boston. [MR2239907](#)
- BROCKWELL, P. J. (2001). Continuous-time ARMA processes. In *Stochastic Processes: Theory and Methods. Handbook of Statist.* **19** 249–276. North-Holland, Amsterdam. [MR1861726](#)
- BROCKWELL, P., DAVIS, R. and YANG, Y. (2007). Continuous-time Gaussian autoregression. *Statist. Sinica* **17** 63.
- DETTE, H., KONSTANTINOU, M. and ZHIGLJAVSKY, A. (2017). A new approach to optimal designs for correlated observations. *Ann. Statist.* **45** 1579–1608. [MR3670189](#)
- DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2013). Optimal design for linear models with correlated observations. *Ann. Statist.* **41** 143–176. [MR3059413](#)
- DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2016). Optimal designs in regression with correlated errors. *Ann. Statist.* **44** 113–152. [MR3449764](#)
- DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2019). Supplement to “The BLUE in continuous-time regression models with correlated errors.” DOI:[10.1214/18-AOS1734SUPP](#).
- FEDOROV, V. V. and MÜLLER, W. G. (2007). Optimum design for correlated fields via covariance kernel expansions. In *MODa 8—Advances in Model-Oriented Design and Analysis*. 57–66. Physica-Verlag/Springer, Heidelberg. [MR2409030](#)
- FREEDEN, W. (1999). *Multiscale Modelling of Spaceborne Geodata*. B. G. Teubner, Stuttgart. [MR1715230](#)

MSC2010 subject classifications. Primary 62M10; secondary 62M09.

Key words and phrases. Linear regression, correlated observations, signed measures, optimal design, BLUE, AR processes, continuous autoregressive model.

- GANTMACHER, F. R. (1959). *The Theory of Matrices*. Vols. 1, 2. Chelsea Publishing Co., New York. Translated by K. A. Hirsch. [MR0107649](#)
- GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277. [MR0039202](#)
- GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Stat.* **25** 252–272. [MR0062402](#)
- HÁJEK, J. (1956). Linear estimation of the mean value of a stationary random process with convex correlation function. *Czechoslovak Math. J.* **6** 94–117. [MR0080399](#)
- HANNAN, E. J. (1975). Linear regression in continuous time. *J. Aust. Math. Soc.* **19** 146–159. [MR0375699](#)
- HE, S. W. and WANG, J. G. (1989). On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model. *J. Time Series Anal.* **10** 315–323. [MR1038465](#)
- KANWAL, R. P. (1997). *Linear Integral Equations*, 2nd ed. Birkhäuser, Inc., Boston, MA. [MR1427946](#)
- KHOLEVO, A. S. (1969). On estimates of regression coefficients. *Theory Probab. Appl.* **14** 79–104. [MR0254985](#)
- LOH, W.-L. and LAM, T.-K. (2000). Estimating structured correlation matrices in smooth Gaussian random field models. *Ann. Statist.* **28** 880–904. [MR1792792](#)
- MORRIS, M. D., MITCHELL, T. J. and YLVISAKER, D. (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics* **35** 243–255. [MR1234641](#)
- MÜLLER, W. G. and PÁZMAN, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika* **90** 423–434. [MR1986657](#)
- NÄTHER, W. (1985). *Effective Observation of Random Fields*. Teubner Verlagsgesellschaft, Leipzig.
- OSBORNE, M. A., GARNETT, R. and ROBERTS, S. J. (2009). Gaussian processes for global optimization. In *3rd International Conference on Learning and Intelligent Optimization (LION3)* 1–15. Citeseer.
- PARZEN, E. (1961). An approach to time series analysis. *Ann. Math. Stat.* **32** 951–989. [MR0143315](#)
- PRONZATO, L. and MÜLLER, W. G. (2012). Design of computer experiments: Space filling and beyond. *Stat. Comput.* **22** 681–701. [MR2909615](#)
- RAMM, A. G. (1980). *Theory and Applications of Some New Classes of Integral Equations*. Springer, New York. [MR0601947](#)
- RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA. [MR2514435](#)
- RITTER, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Springer, Berlin. [MR1763973](#)
- RITTER, K., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks–Ylvisaker conditions. *Ann. Appl. Probab.* **5** 518–540. [MR1336881](#)
- ROSENBLATT, M. (1956). Some regression problems in time series analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 165–186. Univ. California Press, Berkeley. [MR0084920](#)
- SACKS, J. and YLVISAKER, N. D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Stat.* **37** 66–89. [MR0192601](#)
- SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Stat.* **39** 49–69. [MR0220424](#)
- SACKS, J. and YLVISAKER, D. (1970). Designs for regression problems with correlated errors. III. *Ann. Math. Stat.* **41** 2057–2074. [MR0270530](#)
- SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. [MR3127866](#)

- STEIN, M. L. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.
- SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J. Statist. Plann. Inference* **143** 1249–1272. [MR3055745](#)
- YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Vol. I: Basic Results*. Springer, New York. [MR0893393](#)

ADAPTIVE-TO-MODEL CHECKING FOR REGRESSIONS WITH DIVERGING NUMBER OF PREDICTORS¹

BY FALONG TAN* AND LIXING ZHU^{†,‡}

*Hunan University**, *Beijing Normal University[†]* and *Hong Kong Baptist University[‡]*

In this paper, we construct an adaptive-to-model residual-marked empirical process as the base of constructing a goodness-of-fit test for parametric single-index models with diverging number of predictors. To study the relevant asymptotic properties, we first investigate, under the null and alternative hypothesis, the estimation consistency and asymptotically linear representation of the nonlinear least squares estimator for the parameters of interest and then the convergence of the empirical process to a Gaussian process. We prove that under the null hypothesis the convergence of the process holds when the number of predictors diverges to infinity at a certain rate that can be of order, in some cases, $o(n^{1/3}/\log n)$ where n is the sample size. The convergence is also studied under the local and global alternative hypothesis. These results are readily applied to other model checking problems. Further, by modifying the approach in the literature to suit the diverging dimension settings, we construct a martingale transformation and then the asymptotic properties of the test statistic are investigated. Numerical studies are conducted to examine the performance of the test.

REFERENCES

- BIERENS, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20** 105–134. [MR0685673](#)
- BIERENS, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58** 1443–1458. [MR1080813](#)
- COOK, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York. [MR1645673](#)
- COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. [MR1902895](#)
- COOK, R. D. and WEISBERG, S. (1991). Discussion of Sliced inverse regression for dimension reduction, by K. C. Li. *J. Amer. Statist. Assoc.* **86** 316–342.
- DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27** 1012–1040. [MR1724039](#)
- ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. [MR2328527](#)
- FAN, J. and HUANG, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96** 640–652. [MR1946431](#)

MSC2010 subject classifications. Primary 62G10; secondary 62M07.

Key words and phrases. Adaptive-to-model test, diverging number of predictors, empirical process, martingale transformation, parametric single-index models, sufficient dimension reduction.

- FAN, Y. and LI, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890. [MR1399221](#)
- FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. [MR2065194](#)
- GONZÁLEZ MANTEIGA, W. and CRUJEIRAS, R. (2013). An updated review of Goodness-of-Fit tests for regression models. *TEST* **22** 361–411.
- GUO, X., WANG, T. and ZHU, L. (2016). Model checking for parametric single-index models: A dimension reduction model-adaptive approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1013–1035. [MR3557187](#)
- HALL, P. and LI, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. [MR1232523](#)
- HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. [MR1245774](#)
- HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. [MR0356373](#)
- KHMALADZE, E. V. (1982). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.
- KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034. [MR2065196](#)
- KOUL, H. L. and NI, P. (2004). Minimum distance regression model checking. *J. Statist. Plann. Inference* **119** 109–141. [MR2018453](#)
- LAVERGNE, P. and PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143** 103–122. [MR2384435](#)
- LAVERGNE, P. and PATILEA, V. (2012). One for all and all for one: Regression checks with many regressors. *J. Bus. Econom. Statist.* **30** 41–52. [MR2899183](#)
- 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](#)
- LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. [MR2354409](#)
- PORTNOY, S. (1984). Asymptotic behavior of M -estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. [MR0760690](#)
- PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. [MR0811499](#)
- STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. [MR1439316](#)
- STUTE, W., GONZÁLEZ MANTEIGA, W. and PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149. [MR1614600](#)
- STUTE, W., THIES, S. and ZHU, L.-X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26** 1916–1934. [MR1673284](#)
- STUTE, W., XU, W. L. and ZHU, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467. [MR2521592](#)
- STUTE, W. and ZHU, L.-X. (2002). Model checks for generalized linear models. *Scand. J. Stat.* **29** 535–545. [MR1925573](#)
- TAN, F. and ZHU, L. (2019). Supplement to “Adaptive-to-model checking for regressions with diverging number of predictors.” DOI:10.1214/18-AOS1735SUPP.
- TAN, F., ZHU, X. and ZHU, L. (2018). A projection-based adaptive-to-model test for regressions. *Statist. Sinica* **28** 157–188.

- VAN KEILEGOM, I., GONZÁLEZ MANTEIGA, W. and SÁNCHEZ SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17** 401–415. [MR2434335](#)
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. [MR1385671](#)
- WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. [MR0624344](#)
- XIA, Q., XU, W. and ZHU, L. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statist. Sinica* **25** 1025–1044. [MR3409736](#)
- ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. [MR1413644](#)
- ZHU, L. X. and AN, H. Z. (1992). A method for testing nonlinearity in regression models. *J. Math. (Wuhan)* **12** 391–397. [MR1254621](#)
- ZHU, L.-X. and CHENG, P. (1994). The optimal lower bound for the tail probability of the Kolmogorov distance in the worst direction. *Sankhyā Ser. A* **56** 265–293. [MR1664915](#)
- ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. [MR2281245](#)
- ZHU, L.-P., ZHU, L.-X. and FENG, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105** 1455–1466. [MR2796563](#)
- ZHU, L., WANG, T., ZHU, L. and FERRÉ, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **97** 295–304. [MR2650739](#)
- ZOU, H. and ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37** 1733–1751. [MR2533470](#)

NONPARAMETRIC SCREENING UNDER CONDITIONAL STRICTLY CONVEX LOSS FOR ULTRAHIGH DIMENSIONAL SPARSE DATA

BY XU HAN

Temple University

Sure screening technique has been considered as a powerful tool to handle the ultrahigh dimensional variable selection problems, where the dimensionality p and the sample size n can satisfy the NP dimensionality $\log p = O(n^a)$ for some $a > 0$ [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** (2008) 849–911]. The current paper aims to simultaneously tackle the “universality” and “effectiveness” of sure screening procedures. For the “universality,” we develop a general and unified framework for nonparametric screening methods from a loss function perspective. Consider a loss function to measure the divergence of the response variable and the underlying nonparametric function of covariates. We newly propose a class of loss functions called conditional strictly convex loss, which contains, but is not limited to, negative log likelihood loss from one-parameter exponential families, exponential loss for binary classification and quantile regression loss. The sure screening property and model selection size control will be established within this class of loss functions. For the “effectiveness,” we focus on a goodness-of-fit nonparametric screening (Goffins) method under conditional strictly convex loss. Interestingly, we can achieve a better convergence probability of containing the true model compared with related literature. The superior performance of our proposed method has been further demonstrated by extensive simulation studies and some real scientific data example.

REFERENCES

- ANDERSON, M. J. and ROBINSON, J. (2001). Permutation tests for linear models. *Aust. N. Z. J. Stat.* **43** 75–88. [MR1837497](#)
- BARUT, E., FAN, J. and VERHASSELT, A. (2016). Conditional sure independence screening. *J. Amer. Statist. Assoc.* **111** 1266–1277. [MR3561948](#)
- BRËGMAN, L. M. (1967). A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Ž. Vyčisl. Mat. Mat. Fiz.* **7** 620–631. [MR0215617](#)
- mr BULDYGIN, V. and KOZACHENKO, Y. (2000). Metric characterization of random variables and random processes.. *Translations of Mathematical Monographs* **188**.
- CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2404.
- CHANG, J., TANG, C. Y. and WU, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* **41** 2123–2148. [MR3127860](#)

MSC2010 subject classifications. 62G99.

Key words and phrases. Ultrahigh dimensional variable selection, sure screening property, goodness-of-fit nonparametric screening, conditional strictly convex loss.

- DE BOOR, C. (1978). *A Practical Guide to Splines. Applied Mathematical Sciences* **27**. Springer, New York. [MR0507062](#)
- FAN, J. and FAN, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* **36** 2605–2637. [MR2485009](#)
- FAN, J., FENG, Y. and SONG, R. (2011). Nonparametric independence screening in sparse ultrahigh-dimensional additive models. *J. Amer. Statist. Assoc.* **106** 544–557. [MR2847969](#)
- FAN, J., FENG, Y. and TONG, X. (2012). A road to classification in high dimensional space: The regularized optimal affine discriminant. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 745–771. [MR2965958](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. [MR2530322](#)
- FAN, J., MA, Y. and DAI, W. (2014). Nonparametric independence screening in sparse ultrahigh-dimensional varying coefficient models. *J. Amer. Statist. Assoc.* **109** 1270–1284. [MR3265696](#)
- FAN, J., SAMWORTH, R. and WU, Y. (2009). Ultrahigh dimensional feature selection: Beyond the linear model. *J. Mach. Learn. Res.* **10** 2013–2038. [MR2550099](#)
- FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38** 3567–3604. [MR2766861](#)
- FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. [MR1473055](#)
- GAO, Q., WU, Y., ZHU, C. and WANG, Z. (2008). Asymptotic normality of maximum quasi-likelihood estimators in generalized linear models with fixed design. *J. Syst. Sci. Complex.* **21** 463–473. [MR2425677](#)
- GORDON, G. et al. (2002). Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma. *Cancer Research* **62** 4963–4967.
- HAN, X. (2019). Supplement to “Nonparametric screening under conditional strictly convex loss for ultrahigh dimensional sparse data.” DOI:[10.1214/18-AOS1738SUPP](#).
- HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. [MR3059421](#)
- HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer, New York. [MR1461808](#)
- KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. [MR2268657](#)
- LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. [MR1805785](#)
- LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. [MR3010900](#)
- LI, G., PENG, H., ZHANG, J. and ZHU, L. (2012). Robust rank correlation based screening. *Ann. Statist.* **40** 1846–1877. [MR3015046](#)
- MAI, Q. and ZOU, H. (2015). The fused Kolmogorov filter: A nonparametric model-free screening method. *Ann. Statist.* **43** 1471–1497. [MR3357868](#)
- MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. [MR2572443](#)
- SONG, R., LU, W., MA, S. and JENG, X. J. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika* **101** 799–814. [MR3286918](#)
- STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606. [MR0840516](#)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)

- WENG, H., FENG, Y. and QIAO, X. (2017). Regularization after retention in ultrahigh dimensional linear regression models. *Statist. Sinica*. In press.
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHANG, C., JIANG, Y. and SHANG, Z. (2009). New aspects of Bregman divergence in regression and classification with parametric and nonparametric estimation. *Canad. J. Statist.* **37** 119–139. [MR2509465](#)
- ZHAO, S. D. and LI, Y. (2012). Principled sure independence screening for Cox models with ultrahigh-dimensional covariates. *J. Multivariate Anal.* **105** 397–411. [MR2877525](#)
- ZHU, L.-P., LI, L., LI, R. and ZHU, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **106** 1464–1475. [MR2896849](#)
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)
- ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. [MR2137327](#)
- ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. [MR2435443](#)

LOCAL STATIONARITY AND TIME-INHOMOGENEOUS MARKOV CHAINS

BY LIONEL TRUQUET

CREST-ENSAI, CNRS UMR 9194

A primary motivation of this contribution is to define new locally stationary Markov models for categorical or integer-valued data. For this initial purpose, we propose a new general approach for dealing with time-inhomogeneity that extends the local stationarity notion developed in the time series literature. We also introduce a probabilistic framework which is very flexible and allows us to consider a much larger class of Markov chain models on arbitrary state spaces, including most of the locally stationary autoregressive processes studied in the literature. We consider triangular arrays of time-inhomogeneous Markov chains, defined by some families of contracting and slowly-varying Markov kernels. The finite-dimensional distribution of such Markov chains can be approximated locally with the distribution of ergodic Markov chains and some mixing properties are also available for these triangular arrays. As a consequence of our results, some classical geometrically ergodic homogeneous Markov chain models have a locally stationary version, which lays the theoretical foundations for new statistical modeling. Statistical inference of finite-state Markov chains can be based on kernel smoothing and we provide a complete and fast implementation of such models, directly usable by the practitioners. We also illustrate the theory on a real data set. A central limit theorem for Markov chains on more general state spaces is also provided and illustrated with the statistical inference in INAR models, Poisson ARCH models and binary time series models. Additional examples such as locally stationary regime-switching or SETAR models are also discussed.

REFERENCES

- [1] AL-OSH, M. A. and ALZAID, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Series Anal.* **8** 261–275. [MR0903755](#)
- [2] ATHREYA, K. B. and FUH, C. D. (1992). Bootstrapping Markov chains: Countable case. *J. Statist. Plann. Inference* **33** 311–331. [MR1200651](#)
- [3] AVERY, P. J. and HENDERSON, D. A. (1999). Detecting a changed segment in DNA sequences. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **48** 489–503. [MR1721441](#)
- [4] BARTOLI, N. and DEL MORAL, P. (2001). *Simulation et algorithmes stochastiques: une introduction avec applications*. Editions Cépaduès.
- [5] BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes. Statistical Research Monographs. Vol. II*. Univ. Chicago Press, Chicago, IL. [MR0123419](#)
- [6] BIRR, S., VOLGUSHEV, S., KLEY, T., DETTE, H. and HALLIN, M. (2017). Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1619–1643. [MR3731679](#)

MSC2010 subject classifications. Primary 62M05, 62G08; secondary 62M10.

Key words and phrases. Markov chains, local stationarity.

- [7] BOLCH, G., GREINER, S., DE MEER, H. and TRIVEDI, K. S. (2006). *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, 2nd ed. Wiley, Hoboken, NJ. [MR2216267](#)
- [8] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. [MR2178042](#)
- [9] BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions. Vol. 2*. Kendrick Press, Heber City, UT. [MR2325295](#)
- [10] BRILLINGER, D. R., MORETTIN, P. A., IRIZARRY, R. A. and CHIANN, C. (2000). Some wavelet-based analyses of Markov chain data. *Signal Process.* **80** 1607–1627.
- [11] CAO, X.-R. (1998). The Maclaurin series for performance functions of Markov chains. *Adv. in Appl. Probab.* **30** 676–692. [MR1663549](#)
- [12] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. [MR1429916](#)
- [13] DAHLHAUS, R., RICHTER, S. and WU, W. B. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. [MR1835040](#)
- [14] DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. [MR2278352](#)
- [15] DAVIS, R. A., HOLAN, S. H., LUND, R. and RAVISHANKER, N., eds. (2016) *Handbook of Discrete-Valued Time Series*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL. [MR3642975](#)
- [16] DEDECKER, J. and PRIEUR, C. (2004). Coupling for τ -dependent sequences and applications. *J. Theoret. Probab.* **17** 861–885. [MR2105738](#)
- [17] DOBRUSHIN, R. L. (1956). Central limit theorem for nonstationary Markov chains. II. *Teor. Veroyatn. Primen.* **1** 365–425. [MR0097112](#)
- [18] DOBRUSHIN, R. L. (1970). Definition of a system of random variables by means of conditional distributions. *Teor. Veroyatn. Primen.* **15** 469–497. [MR0298716](#)
- [19] DOUC, R., MOULINES, E. and ROSENTHAL, J. S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. [MR2099647](#)
- [20] DOUC, R., MOULINES, E. and STOFFER, D. S. (2014). *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*. Chapman & Hall/CRC Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL. [MR3289095](#)
- [21] DOUKHAN, P. (1994). *Mixing. Properties and Examples*. Lecture Notes in Statistics **85**. Springer, New York. [MR1312160](#)
- [22] FERRÉ, D., HERVÉ, Y. and LEDOUX, J. (2013). Regular perturbation of V -geometrically ergodic markov chains. *J. Appl. Probab.* **50** 184–194. [MR3076780](#)
- [23] DU, J. G. and LI, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Time Series Anal.* **12** 129–142. [MR1108796](#)
- [24] FOKIANOS, K. and MOYSIADIS, T. (2017). Binary time series models driven by a latent process. *Econ. Stat.* **2** 117–130. [MR3671156](#)
- [25] FOKIANOS, K., RAHBK, A. and TJØSTHEIM, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.* **104** 1430–1439. [MR2596998](#)
- [26] FRYZLEWICZ, P., SAPATINAS, T. and SUBBA RAO, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36** 742–786. [MR2396814](#)
- [27] FRYZLEWICZ, P. and SUBBA RAO, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* **17** 320–346. [MR2797994](#)
- [28] GRANGER, C. and STĂRICĂ, C. (2005). Nonstationarities in stock returns. *Rev. Econ. Stat.* **87** 503–522.
- [29] HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris’ ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser/Springer, Basel. [MR2857021](#)

- [30] HALL, P. and BURA, E. (2004). Nonparametric methods of inference for finite-state, inhomogeneous Markov processes. *Bernoulli* **10** 919–938. [MR2093617](#)
- [31] HAMILTON, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57** 357–384. [MR0996941](#)
- [32] HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications. Springer Series in Statistics*. Springer, New York. [MR2920735](#)
- [33] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen. [MR0322926](#)
- [34] KULPERGER, R. J. and PRAKASA RAO, B. L. S. (1989). Bootstrapping a finite state Markov chain. *Sankhyā Ser. A* **51** 178–191. [MR1065568](#)
- [35] LI, G., GUAN, B., LI, W. K. and YU, P. L. H. (2015). Hysteretic autoregressive time series models. *Biometrika* **102** 717–723. [MR3394287](#)
- [36] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2509253](#)
- [37] MOYSIADIS, T. and FOKIANOS, K. (2014). On binary and categorical time series models with feedback. *J. Multivariate Anal.* **131** 209–228. [MR3252645](#)
- [38] RAJAGOPALAN, B., LALL, U. and TARBOTON, D. G. (1996). Nonhomogeneous Markov model for daily precipitation. *J. Hydrol. Eng.* **1** 33–40.
- [39] RICHTER, S. and DAHLHAUS, R. (2017). Cross validation for locally stationary processes. Available at [arXiv:1705.10046](#).
- [40] SALOFF-COSTE, L. and ZÚÑIGA, J. (2007). Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Process. Appl.* **117** 961–979. [MR2340874](#)
- [41] SALOFF-COSTE, L. and ZÚÑIGA, J. (2011). Merging for inhomogeneous finite Markov chains, Part II: Nash and log-Sobolev inequalities. *Ann. Probab.* **39** 1161–1203. [MR2789587](#)
- [42] SARUKKAI, R. R. (2000). Link prediction and path analysis using Markov chains. *Comput. Netw.* **33** 377–386.
- [43] SENETA, E. (2006). *Non-negative Matrices and Markov Chains. Springer Series in Statistics*. Springer, New York. [MR2209438](#)
- [44] STOCKIS, J.-P., FRANKE, J. and KAMGAING, J. T. (2010). On geometric ergodicity of CHARME models. *J. Time Series Anal.* **31** 141–152. [MR2676147](#)
- [45] SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. [MR2285698](#)
- [46] TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach. Oxford Statistical Science Series 6*. The Clarendon Press, New York. [MR1079320](#)
- [47] TRUQUET, L. (2017). A perturbation analysis of some markov chains models with time-varying parameters. Available at [arXiv:1610.09272](#).
- [48] TRUQUET, L. (2019). Supplement to “Local stationarity and time-inhomogeneous Markov chains.” DOI:[10.1214/18-AOS1739SUPP](#).
- [49] VERGNE, N. (2008). Drifting Markov models with polynomial drift and applications to DNA sequences. *Stat. Appl. Genet. Mol. Biol.* **7** Art. 6, 45. [MR2386323](#)
- [50] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] 338*. Springer, Berlin. [MR2459454](#)
- [51] VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. [MR3097614](#)
- [52] WINKLER, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction. Applications of Mathematics (New York) 27*. Springer, Berlin. [MR1316400](#)
- [53] ZHANG, T. and WU, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. [MR3319142](#)

HIGH-DIMENSIONAL CHANGE-POINT DETECTION UNDER SPARSE ALTERNATIVES¹

BY FARIDA ENIKEEVA^{*,†} AND ZAID HARCHAOU[‡]

Université de Poitiers^{}, IITP[†] and University of Washington[‡]*

We consider the problem of detecting a change in mean in a sequence of high-dimensional Gaussian vectors. The change in mean may be occurring simultaneously in an unknown subset components. We propose a hypothesis test to detect the presence of a change-point and establish the detection boundary in different regimes under the assumption that the dimension tends to infinity and the length of the sequence grows with the dimension. A remarkable feature of the proposed test is that it does not require any knowledge of the subset of components in which the change in mean is occurring and yet automatically adapts to yield optimal rates of convergence over a wide range of statistical regimes.

REFERENCES

- [1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. [MR2722464](#)
- [2] BACON, R., BRINCHMANN, J., RICHARD, J., CONTINI, T., DRAKE, A., FRANX, M., TACCHELLA, S., VERNET, J., WISOTZKI, L., et al. (2015). The MUSE 3D view of the hubble deep field South. *Astron. Astrophys.* **575** A75.
- [3] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. [MR1935648](#)
- [4] BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs, NJ. [MR1210954](#)
- [5] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. [MR3185193](#)
- [6] BOURGUIGNON, S., MARY, D. and SLEZAK, E. (2011). Restoration of astrophysical spectra with sparsity constraints: Models and algorithms. *IEEE J. Sel. Top. Signal Process.* **5** 1002–1013.
- [7] BRODSKY, B. E. and DARKHOVSKY, B. S. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications* **243**. Kluwer Academic, Dordrecht. [MR1228205](#)
- [8] BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. [MR3160567](#)
- [9] 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](#)
- [10] CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. [MR1477909](#)

MSC2010 subject classifications. Primary 62G10, 62H15; secondary 60G35.

Key words and phrases. Change-point problem, high-dimensional data, minimax optimality, sparsity.

- [11] CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester. [MR2743035](#)
- [12] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#)
- [13] ENIKEEVA, F. and HARCHAOUI, Z. (2019). Supplement to “High-dimensional change-point detection under sparse alternatives.” DOI:[10.1214/18-AOS1740SUPP](#).
- [14] HARCHAOUI, Z., VALLET, F., LUNG-YUT-FONG, A. and CAPPÉ, O. (2009). A regularized kernel-based approach to unsupervised audio segmentation. In *ICASSP* 1665–1668.
- [15] HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Monographs on Statistics and Applied Probability **143**. CRC Press, Boca Raton, FL. [MR3616141](#)
- [16] INGSTER, Y. I., POUET, C. and TSYBAKOV, A. B. (2009). Classification of sparse high-dimensional vectors. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4427–4448. [MR2546395](#)
- [17] INGSTER, Y. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#)
- [18] INGSTER, Y. and SUSLINA, I. A. (2002). On the detection of a signal with a known shape in a multichannel system. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **294** 88–112, 261. [MR1976749](#)
- [19] INGSTER, Y. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. [MR1991446](#)
- [20] KOROSTELEV, A. and KOROSTELEVA, O. (2011). *Mathematical Statistics: Asymptotic Minimax Theory. Graduate Studies in Mathematics* **119**. Amer. Math. Soc., Providence, RI. [MR2767163](#)
- [21] KOROSTELEV, A. and LEPSKI, O. (2008). On a multi-channel change-point problem. *Math. Methods Statist.* **17** 187–197. [MR2448945](#)
- [22] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. [MR2135927](#)
- [23] LÉVY-LEDUC, C. and ROUEFF, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Ann. Appl. Stat.* **3** 637–662. [MR2750676](#)
- [24] MEILLIER, C., CHATELAIN, F., MICHEL, O., BACON, R., PIQUERAS, L., BACHER, R. and AYASSO, H. (2016). SELFI: An object-based, Bayesian method for faint emission line source detection in MUSE deep field data cubes. *Astron. Astrophys.* **588** A140.
- [25] OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.
- [26] PARIS, S., MARY, D. and FERRARI, A. (2013). Detection tests using sparse models, with application to hyperspectral data. *IEEE Trans. Signal Process.* **61** 1481–1494. [MR3043363](#)
- [27] SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York–Heidelberg. [MR0468067](#)
- [28] SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. [MR0799155](#)
- [29] STRANSKY, N., VALLOT, C., REYAL, F., BERNARD-PIERROT, I., DE MEDINA, S. G., SEGRAVES, R., DE RYCKE, Y., ELVIN, P., CASSIDY, A., et al. (2006). Regional copy number-independent deregulation of transcription in cancer. *Nat. Genet.* **38** 1386–1396.
- [30] VERT, J. P. and BLEAKLEY, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. In *Advances in Neural Information Processing Systems* 23 (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 2343–2351. Curran Associates, Red Hook, NY.
- [31] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. [MR3099117](#)

- [32] ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. [MR2672488](#)

PERTURBATION BOOTSTRAP IN ADAPTIVE LASSO

BY DEBRAJ DAS¹, KARL GREGORY² AND S. N. LAHIRI¹

*University of Wisconsin—Madison, University of South Carolina and
 North Carolina State University*

The Adaptive Lasso (Alasso) was proposed by Zou [*J. Amer. Statist. Assoc.* **101** (2006) 1418–1429] as a modification of the Lasso for the purpose of simultaneous variable selection and estimation of the parameters in a linear regression model. Zou [*J. Amer. Statist. Assoc.* **101** (2006) 1418–1429] established that the Alasso estimator is variable-selection consistent as well as asymptotically Normal in the indices corresponding to the nonzero regression coefficients in certain fixed-dimensional settings. In an influential paper, Minnier, Tian and Cai [*J. Amer. Statist. Assoc.* **106** (2011) 1371–1382] proposed a perturbation bootstrap method and established its distributional consistency for the Alasso estimator in the fixed-dimensional setting. In this paper, however, we show that this (naive) perturbation bootstrap fails to achieve second-order correctness in approximating the distribution of the Alasso estimator. We propose a modification to the perturbation bootstrap objective function and show that a suitably Studentized version of our modified perturbation bootstrap Alasso estimator achieves second-order correctness even when the dimension of the model is allowed to grow to infinity with the sample size. As a consequence, inferences based on the modified perturbation bootstrap will be more accurate than the inferences based on the oracle Normal approximation. We give simulation studies demonstrating good finite-sample properties of our modified perturbation bootstrap method as well as an illustration of our method on a real data set.

REFERENCES

- BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. [MR0471142](#)
- BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansions*. John Wiley & Sons, New York.
- BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575](#)
- BÜHLMANN, P., KALISCH, M. and MEIER, L. (2014). High-dimensional statistics with a view towards applications in biology. *Annu. Rev. Statist. Appl.* **1** 255–278.
- CAMPONOV, L. (2015). On the validity of the pairs bootstrap for Lasso estimators. *Biometrika* **102** 981–987. [MR3431568](#)
- CANER, M. and FAN, Q. (2010). The adaptive Lasso method for instrumental variable selection. Working paper, North Carolina State Univ., Raleigh, NC.

MSC2010 subject classifications. Primary 62J07; secondary 62G09, 62E20.

Key words and phrases. Alasso, naive perturbation bootstrap, modified perturbation bootstrap, second-order correctness, oracle.

- CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. [MR2157808](#)
- CHATTERJEE, A. and LAHIRI, S. N. (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.* **138** 4497–4509. [MR2680074](#)
- CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping Lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. [MR2847974](#)
- CHATTERJEE, A. and LAHIRI, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41** 1232–1259. [MR3113809](#)
- DAS, D., GREGORY, K. and LAHIRI, S. N. (2019). Supplement to “Perturbation bootstrap in adaptive Lasso.” DOI:[10.1214/18-AOS1741SUPP](#).
- DAS, D. and LAHIRI, S. N. (2019). Second order correctness of perturbation bootstrap M-estimator of multiple linear regression parameter. *Bernoulli* **25** 654–682. [MR3892332](#)
- DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. Available at [arXiv:1606.03940](#). [MR3713586](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.
- FUK, D. H. and NAGAEV, S. V. (1971). Probabilistic inequalities for sums of independent random variables. *Teor. Veroyatn. Primen.* **16** 660–675. [MR0293695](#)
- HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York. [MR1145237](#)
- HALL, P., LEE, Y. K., PARK, B. U. and PAUL, D. (2009). Tie-respecting bootstrap methods for estimating distributions of sets and functions of eigenvalues. *Bernoulli* **15** 380–401. [MR2543867](#)
- JIN, Z., YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. [MR1844838](#)
- KNIGHT, K. and FU, W. (2000). Asymptotics for Lasso-type estimators. *Ann. Statist.* **28** 1356–1378. [MR1805787](#)
- MINNIER, J., TIAN, L. and CAI, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Amer. Statist. Assoc.* **106** 1371–1382. [MR2896842](#)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- TURNBULL, H. W. (1930). A matrix form of Taylor’s theorem. *Proc. Edinb. Math. Soc.* (2) **33** 33–54.
- 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](#)
- WANG, X. and SONG, L. (2011). Adaptive Lasso variable selection for the accelerated failure models. *Comm. Statist. Theory Methods* **40** 4372–4386. [MR2864162](#)
- ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. Roy. Statist. Soc. Ser. B* **76** 217–242. [MR3153940](#)
- ZHOU, Q. M., SONG, P. X.-K. and THOMPSON, M. E. (2012). Information ratio test for model misspecification in quasi-likelihood inference. *J. Amer. Statist. Assoc.* **107** 205–213. [MR2949352](#)
- ZOU, H. (2006). The adaptive Lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)

ESTIMATION BOUNDS AND SHARP ORACLE INEQUALITIES OF REGULARIZED PROCEDURES WITH LIPSCHITZ LOSS FUNCTIONS¹

BY PIERRE ALQUIER², VINCENT COTTET AND GUILLAUME LECUÉ

CREST, CNRS, ENSAE, Université Paris Saclay

We obtain estimation error rates and sharp oracle inequalities for regularization procedures of the form

$$\hat{f} \in \operatorname{argmin}_{f \in F} \left(\frac{1}{N} \sum_{i=1}^N \ell_f(X_i, Y_i) + \lambda \|f\| \right)$$

when $\|\cdot\|$ is any norm, F is a convex class of functions and ℓ is a Lipschitz loss function satisfying a Bernstein condition over F . We explore both the bounded and sub-Gaussian stochastic frameworks for the distribution of the $f(X_i)$'s, with no assumption on the distribution of the Y_i 's. The general results rely on two main objects: a complexity function and a sparsity equation, that depend on the specific setting in hand (loss ℓ and norm $\|\cdot\|$).

As a proof of concept, we obtain minimax rates of convergence in the following problems: (1) matrix completion with any Lipschitz loss function, including the hinge and logistic loss for the so-called 1-bit matrix completion instance of the problem, and quantile losses for the general case, which enables to estimate any quantile on the entries of the matrix; (2) logistic LASSO and variants such as the logistic SLOPE, and also shape constrained logistic regression; (3) kernel methods, where the loss is the hinge loss, and the regularization function is the RKHS norm.

REFERENCES

- [1] ALQUIER, P. (2013). Bayesian methods for low-rank matrix estimation: Short survey and theoretical study. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **8139** 309–323. Springer, Heidelberg. [MR3133074](#)
- [2] ALQUIER, P., RIDGWAY, J. and CHOPIN, N. (2016). On the properties of variational approximations of Gibbs posteriors. *J. Mach. Learn. Res.* **17** 239. [MR3595173](#)
- [3] ALQUIER, P., COTTET, V. and LECUÉ, G. (2019). Supplement to “Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions.” DOI:10.1214/18-AOS1742SUPP.
- [4] AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. [MR2336861](#)
- [5] BARTHE, F., GUÉDON, O., MENDELSON, S. and NAOR, A. (2005). A probabilistic approach to the geometry of the l_p^n -ball. *Ann. Probab.* **33** 480–513. [MR2123199](#)
- [6] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. [MR2166554](#)

MSC2010 subject classifications. Primary 60K35, 62G08; secondary 62C20, 62G05, 62G20.

Key words and phrases. Empirical processes, high-dimensional statistics.

- [7] BARTLETT, P. L. and MENDELSON, S. (2006). Empirical minimization. *Probab. Theory Related Fields* **135** 311–334. [MR2240689](#)
- [8] BELLEC, P., LECUÉ, G. and TSYBAKOV, A. (2018). Slope meets Lasso: Improved oracle bounds and optimality *Ann. Statist.* **46** 3603–3642. [MR3852663](#)
- [9] BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. [MR2797841](#)
- [10] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. [MR3418717](#)
- [11] CAI, T. and ZHOU, W.-X. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *J. Mach. Learn. Res.* **14** 3619–3647. [MR3159403](#)
- [12] CANDÈS, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.
- [13] CATONI, O. (2004). *Statistical Learning Theory and Stochastic Optimization: Ecole d'Été de Probabilités de Saint-Flour, XXXI-2001*. **31**. Springer, Berlin. [MR2163920](#)
- [14] CATONI, O. (2007). *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **56**. IMS, Beachwood, OH. [MR2483528](#)
- [15] CHAFAÏ, D., GUÉDON, O., LECUÉ, G. and PAJOR, A. (2012). *Interactions Between Compressed Sensing Random Matrices and High Dimensional Geometry. Panoramas et Synthèses [Panoramas and Syntheses]* **37**. Société Mathématique de France, Paris. [MR3113826](#)
- [16] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. [MR2989474](#)
- [17] COTTET, V. and ALQUIER, P. (2016). 1-bit Matrix Completion: PAC-Bayesian Analysis of a Variational Approximation. Machine Learning. To appear. Preprint [arXiv:1604.04191](#).
- [18] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. Revised reprint of the 1989 original. [MR1932358](#)
- [19] GARCIA-MAGARIÑOS, M., ANTONIADIS, A., CAO, R. and GONZÁLEZ-MANTEIGA, W. (2010). Lasso logistic regression, GSoft and the cyclic coordinate descent algorithm: Application to gene expression data. *Stat. Appl. Genet. Mol. Biol.* **9** 30. [MR2721710](#)
- [20] GORDON, Y., LITVAK, A. E., MENDELSON, S. and PAJOR, A. (2007). Gaussian averages of interpolated bodies and applications to approximate reconstruction. *J. Approx. Theory* **149** 59–73. [MR2371614](#)
- [21] KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. [MR3160583](#)
- [22] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. [MR2329442](#)
- [23] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems. Lecture Notes in Math.* **2033**. Springer, Heidelberg. [MR2829871](#)
- [24] KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. [MR2906869](#)
- [25] KOLTCHINSKII, V. and PANCHENKO, D. (2002). Empirical margin distributions and bounding the generalization error of combined classifiers. *Ann. Statist.* **30** 1–50. [MR1892654](#)
- [26] LAFOND, J., KLOPP, O., MOULINES, E. and SALMON, J. (2014). Probabilistic low-rank matrix completion on finite alphabets. In *Advances in Neural Information Processing Systems* 1727–1735.
- [27] LECUÉ, G. (2011). *Interplay Between Concentration, Complexity and Geometry in Learning Theory with Applications to High Dimensional Data Analysis*. Habilitation à Diriger des Recherches Université, Paris-Est Marne-la-vallée.

- [28] LECUÉ, G. and MENDELSON, S. (2012). General nonexact oracle inequalities for classes with a subexponential envelope. *Ann. Statist.* **40** 832–860. [MR2933668](#)
- [29] LECUÉ, G. and MENDELSON, S. (2013). Learning subgaussian classes: Upper and minimax bounds. Technical Report CNRS, Ecole polytechnique and Technion—to appear in *Topics in Learning Theory*—Société Mathématique de France (S. Boucheron and N. Vayatis eds.).
- [30] LECUÉ, G. and MENDELSON, S. (2017). Regularization and the small-ball method II: Complexity dependent error rates. *J. Mach. Learn. Res.* **18** 146. [MR3763780](#)
- [31] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann. Statist.* **46** 611–641. [MR3782379](#)
- [32] MAI, T. T. and ALQUIER, P. (2015). A Bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. *Electron. J. Stat.* **9** 823–841. [MR3331862](#)
- [33] MAK, C. (1999). *Polychotomous Logistic Regression Via the Lasso*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Univ. Toronto. [MR2699823](#)
- [34] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. [MR1765618](#)
- [35] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2008). The group Lasso for logistic regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 53–71. [MR2412631](#)
- [36] MENDELSON, S. (2004). On the performance of kernel classes. *J. Mach. Learn. Res.* **4** 759–771. [MR2075996](#)
- [37] RAO, M. M. and REN, Z. D. (1991). *Theory of Orlicz Spaces. Monographs and Textbooks in Pure and Applied Mathematics* **146**. Dekker, New York. [MR1113700](#)
- [38] ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. [MR2816342](#)
- [39] SABBE, N., THAS, O. and OTTOY, J.-P. (2013). EMLasso: Logistic lasso with missing data. *Stat. Med.* **32** 3143–3157. [MR3073790](#)
- [40] SREBRO, N., RENNIE, J. and JAAKKOLA, T. S. (2004). Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems* 1329–1336.
- [41] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. [MR3485953](#)
- [42] TIAN, G.-L., TANG, M.-L., FANG, H.-B. and TAN, M. (2008). Efficient methods for estimating constrained parameters with applications to regularized (lasso) logistic regression. *Comput. Statist. Data Anal.* **52** 3528–3542. [MR2427362](#)
- [43] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. [MR2051002](#)
- [44] VAN DE GEER, S. (2016). *Estimation and Testing Under Sparsity. Lecture Notes in Math.* **2159**. Springer, Cham. [MR3526202](#)
- [45] VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. [MR2396809](#)
- [46] ZHANG, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Statist.* **32** 56–85. [MR2051001](#)

CROSS VALIDATION FOR LOCALLY STATIONARY PROCESSES¹

BY STEFAN RICHTER AND RAINER DAHLHAUS

Heidelberg University

We propose an adaptive bandwidth selector via cross validation for local M-estimators in locally stationary processes. We prove asymptotic optimality of the procedure under mild conditions on the underlying parameter curves. The results are applicable to a wide range of locally stationary processes such linear and nonlinear processes. A simulation study shows that the method works fairly well also in misspecified situations.

REFERENCES

- [1] ARKOUN, O. (2011). Sequential adaptive estimators in nonparametric autoregressive models. *Sequential Anal.* **30** 229–247. [MR2801140](#)
- [2] ARKOUN, O. and PERGAMENCHTCHIKOV, S. (2016). Sequential robust estimation for non-parametric autoregressive models. *Sequential Anal.* **35** 489–515. [MR3574333](#)
- [3] CHIU, S.-T. (1991). Bandwidth selection for kernel density estimation. *Ann. Statist.* **19** 1883–1905. [MR1135154](#)
- [4] DAHLHAUS, R. (2012). 13—Locally stationary processes. In *Time Series Analysis: Methods and Applications* (T. S. Rao, S. S. Rao and C. R. Rao, eds.). *Handbook of Statistics* **30** 351–413. Elsevier, Amsterdam. DOI:[10.1016/B978-0-444-53858-1.00013-2](#).
- [5] DAHLHAUS, R. and POLONIK, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. [MR2546797](#)
- [6] DAHLHAUS, R., RICHTER, S. and WU, W. B. (2018). Towards a general theory for non-linear locally stationary processes. *Bernoulli*. To appear.
- [7] DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. [MR2278352](#)
- [8] FRYZLEWICZ, P., SAPATINAS, T. and SUBBA RAO, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36** 742–786. [MR2396814](#)
- [9] GIRAUD, C., ROUEFF, F. and SANCHEZ-PEREZ, A. (2015). Aggregation of predictors for nonstationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes. *Ann. Statist.* **43** 2412–2450. [MR3405599](#)
- [10] HALL, P., MARRON, J. S. and PARK, B. U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* **92** 1–20. [MR1156447](#)
- [11] HÄRDLE, W., HALL, P. and MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83** 86–101. [MR0941001](#)
- [12] HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13** 1465–1481. [MR0811503](#)
- [13] KARMAKAR, S., RICHTER, S. and WU, W. B. (2018). Bahadur representation and simultaneous inference for time-varying models. Technical report.

MSC2010 subject classifications. Primary 62M10; secondary 62G20.

Key words and phrases. Locally stationary processes, cross validation, adaptive bandwidth selection, asymptotic optimality.

- [14] MALLAT, S., PAPANICOLAOU, G. and ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26** 1–47. [MR1611808](#)
- [15] RICHTER, S. and DAHLHAUS, R. (2019). Supplement to “Cross validation for locally stationary processes.” DOI:[10.1214/18-AOS1743SUPP](#).
- [16] SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. [MR2285698](#)
- [17] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#)
- [18] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. [MR2541444](#)

GENERALIZED CLUSTER TREES AND SINGULAR MEASURES

BY YEN-CHI CHEN

University of Washington

In this paper we study the α -cluster tree (α -tree) under both singular and nonsingular measures. The α -tree uses probability contents within a set created by the ordering of points to construct a cluster tree so that it is well defined even for singular measures. We first derive the convergence rate for a density level set around critical points, which leads to the convergence rate for estimating an α -tree under nonsingular measures. For singular measures, we study how the kernel density estimator (KDE) behaves and prove that the KDE is not uniformly consistent but pointwise consistent after rescaling. We further prove that the estimated α -tree fails to converge in the L_∞ metric but is still consistent under the integrated distance. We also observe a new type of critical points—the dimensional critical points (DCPs)—of a singular measure. DCPs are points that contribute to cluster tree topology but cannot be defined using density gradient. Building on the analysis of the KDE and DCPs, we prove the topological consistency of an estimated α -tree.

REFERENCES

- BALAKRISHNAN, S., NARAYANAN, S., RINALDO, A., SINGH, A. and WASSERMAN, L. (2012). Cluster trees on manifolds. In *Advances in Neural Information Processing Systems* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26** 2679–2687. Curran Associates, Red Hook, NY.
- BARYSHNIKOV, Y., BUBENIK, P. and KAHLE, M. (2014). Min-type Morse theory for configuration spaces of hard spheres. *Int. Math. Res. Not. IMRN* **9** 2577–2592. [MR3207377](#)
- BOBROWSKI, O., MUKHERJEE, S. and TAYLOR, J. E. (2017). Topological consistency via kernel estimation. *Bernoulli* **23** 288–328. [MR3556774](#)
- BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16** 77–102. [MR3317230](#)
- CADRE, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023. [MR2256570](#)
- CADRE, B., PELLETIER, B. and PUDLO, P. (2009). Clustering by estimation of density level sets at a fixed probability. Available at <https://hal.archives-ouvertes.fr/file/index/docid/397437/filename/tlevel.pdf>.
- CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)* **46** 255–308. [MR2476414](#)
- CHAUDHURI, K. and DASGUPTA, S. (2010). Rates of convergence for the cluster tree. In *Advances in Neural Information Processing Systems* 343–351.
- CHAUDHURI, K., DASGUPTA, S., KPOTUFE, S. and VON LUXBURG, U. (2014). Consistent procedures for cluster tree estimation and pruning. *IEEE Trans. Inform. Theory* **60** 7900–7912. [MR3285753](#)

MSC2010 subject classifications. Primary 62G20; secondary 62G05, 62G07.

Key words and phrases. Cluster tree, kernel density estimator, level set, singular measure, critical points, topological data analysis.

- CHAZAL, F., FASY, B., LECCI, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2017). Robust topological inference: Distance to a measure and kernel distance. *J. Mach. Learn. Res.* **18** Paper No. 159, 40. [MR3813808](#)
- CHEN, Y.-C. (2017). A tutorial on kernel density estimation and recent advances. ArXiv preprint. Available at [arXiv:1704.03924](#).
- CHEN, Y.-C. (2019). Supplement to “Generalized cluster trees and singular measures”. DOI:[10.1214/18-AOS1744SUPP](#).
- CHEN, Y.-C. and DOBRA, A. (2017). Measuring human activity spaces with density ranking based on GPS data. ArXiv preprint. Available at [arXiv:1708.05017](#).
- CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *Ann. Statist.* **43** 1896–1928. [MR3375871](#)
- CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2016). A comprehensive approach to mode clustering. *Electron. J. Stat.* **10** 210–241. [MR3466181](#)
- CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. [MR3750891](#)
- CHEN, Y.-C., KIM, J., BALAKRISHNAN, S., RINALDO, A. and WASSERMAN, L. (2016). Statistical inference for cluster trees. ArXiv preprint. Available at [arXiv:1605.06416](#).
- COHEN-STEINER, D., EDELSBRUNNER, H. and HARER, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* **37** 103–120. [MR2279866](#)
- EDELSBRUNNER, H. and HARER, J. (2008). Persistent homology—A survey. In *Surveys on Discrete and Computational Geometry. Contemp. Math.* **453** 257–282. Amer. Math. Soc., Providence, RI. [MR2405684](#)
- EDELSBRUNNER, H. and MOROZOV, D. (2013). Persistent homology: Theory and practice. In *European Congress of Mathematics* 31–50. Eur. Math. Soc., Zürich. [MR3469114](#)
- EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403. [MR2195639](#)
- ELDRIDGE, J., BELKIN, M. and WANG, Y. (2015). Beyond hartigan consistency: Merge distortion metric for hierarchical clustering. In *Proceedings of the 28th Conference on Learning Theory* 588–606.
- FASY, B. T., LECCI, F., RINALDO, A., WASSERMAN, L., BALAKRISHNAN, S. and SINGH, A. (2014). Confidence sets for persistence diagrams. *Ann. Statist.* **42** 2301–2339. [MR3269981](#)
- FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. [MR0110078](#)
- FRIEDMAN, G. (2014). Singular intersection homology. Preprint.
- GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2009). On the path density of a gradient field. *Ann. Statist.* **37** 3236–3271. [MR2549559](#)
- GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2014). Non-parametric ridge estimation. *Ann. Statist.* **42** 1511–1545. [MR3262459](#)
- GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré B, Probab. Stat.* **38** 907–921. [MR1955344](#)
- GORESKEY, M. and MACPHERSON, R. (1980). Intersection homology theory. *Topology* **19** 135–162. [MR0572580](#)
- GORESKEY, M. and MACPHERSON, R. (1988). *Stratified Morse Theory. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **14**. Springer, Berlin. [MR0932724](#)
- HARTIGAN, J. A. (1981). Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.* **76** 388–394. [MR0624340](#)
- KENT, B. P. (2013). Level set trees for applied statistics. Ph.D. thesis, Carnegie Mellon Univ., Pittsburgh, PA.
- KLEMELÄ, J. (2004). Visualization of multivariate density estimates with level set trees. *J. Comput. Graph. Statist.* **13** 599–620. [MR2087717](#)

- KLEMELÁ, J. (2006). Visualization of multivariate density estimates with shape trees. *J. Comput. Graph. Statist.* **15** 372–397. [MR2256150](#)
- KLEMELÁ, J. (2009). *Smoothing of Multivariate Data: Density Estimation and Visualization*. Wiley, Hoboken, NJ. [MR2640738](#)
- KPOTUFE, S. and LUXBURG, U. V. (2011). Pruning nearest neighbor cluster trees. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (L. Getoor and T. Scheffer, eds.) 225–232. International Machine Learning Society, Madison, WI.
- LALOE, T. and SERVIEN, R. (2013). Nonparametric estimation of regression level sets. *J. Korean Statist. Soc.*
- LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. [MR2954043](#)
- MAMMEN, E. and POLONIK, W. (2013). Confidence regions for level sets. *J. Multivariate Anal.* **122** 202–214. [MR3189318](#)
- MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. [MR2537201](#)
- MATTILA, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. *Cambridge Studies in Advanced Mathematics* **44**. Cambridge Univ. Press, Cambridge. [MR1333890](#)
- MILNOR, J. (1963). *Morse Theory. Based on Lecture Notes by M. Spivak and R. Wells*. *Annals of Mathematics Studies*, No. 51. Princeton Univ. Press, Princeton, NJ. [MR0163331](#)
- MOLCHANOV, I. S. (1990). Empirical estimation of quantiles of distributions of random closed sets. *Teor. Veroyatn. Primen.* **35** 586–592. [MR1091220](#)
- MORSE, M. (1925). Relations between the critical points of a real function of n independent variables. *Trans. Amer. Math. Soc.* **27** 345–396. [MR1501318](#)
- MORSE, M. (1930). The foundations of a theory of the calculus of variations in the large in m -space. II. *Trans. Amer. Math. Soc.* **32** 599–631. [MR1501555](#)
- POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—An excess mass approach. *Ann. Statist.* **23** 855–881. [MR1345204](#)
- PREISS, D. (1987). Geometry of measures in $\mathbf{R}^n$: Distribution, rectifiability, and densities. *Ann. of Math.* (2) **125** 537–643. [MR0890162](#)
- RINALDO, A. and WASSERMAN, L. (2010). Generalized density clustering. *Ann. Statist.* **38** 2678–2722. [MR2722453](#)
- RINALDO, A., SINGH, A., NUGENT, R. and WASSERMAN, L. (2012). Stability of density-based clustering. *J. Mach. Learn. Res.* **13** 905–948. [MR2930628](#)
- SCOTT, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed. Wiley, Hoboken, NJ. [MR3329609](#)
- SINGH, A., SCOTT, C. and NOWAK, R. (2009). Adaptive Hausdorff estimation of density level sets. *Ann. Statist.* **37** 2760–2782. [MR2541446](#)
- STEINWART, I. (2011). Adaptive density level set clustering. In *COLT* 703–738.
- STUETZLE, W. (2003). Estimating the cluster type of a density by analyzing the minimal spanning tree of a sample. *J. Classification* **20** 25–47. [MR1983120](#)
- TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. [MR1447735](#)
- TU, L. W. (2008). *An Introduction to Manifolds*. Springer, New York. [MR2356311](#)
- WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299. [MR1604445](#)
- WASSERMAN, L. (2006). *All of Nonparametric Statistics*. *Springer Texts in Statistics*. Springer, New York. [MR2172729](#)
- WASSERMAN, L. (2018). Topological data analysis. *Ann. Rev. Stat. Appl.* **5** 501–535. [MR3774757](#)

SPECTRAL METHOD AND REGULARIZED MLE ARE BOTH OPTIMAL FOR TOP- K RANKING¹

BY YUXIN CHEN^{*,2}, JIANQING FAN^{†,*,3}, CONG MA^{*} AND KAIZHENG WANG^{*}

Princeton University^{} and Fudan University[†]*

This paper is concerned with the problem of top- K ranking from pairwise comparisons. Given a collection of n items and a few pairwise comparisons across them, one wishes to identify the set of K items that receive the highest ranks. To tackle this problem, we adopt the logistic parametric model—the Bradley–Terry–Luce model, where each item is assigned a latent preference score, and where the outcome of each pairwise comparison depends solely on the relative scores of the two items involved. Recent works have made significant progress toward characterizing the performance (e.g., the mean square error for estimating the scores) of several classical methods, including the spectral method and the maximum likelihood estimator (MLE). However, where they stand regarding top- K ranking remains unsettled.

We demonstrate that under a natural random sampling model, the spectral method alone, or the regularized MLE alone, is minimax optimal in terms of the sample complexity—the number of paired comparisons needed to ensure exact top- K identification, for the fixed dynamic range regime. This is accomplished via optimal control of the entrywise error of the score estimates. We complement our theoretical studies by numerical experiments, confirming that both methods yield low entrywise errors for estimating the underlying scores. Our theory is established via a novel leave-one-out trick, which proves effective for analyzing both iterative and noniterative procedures. Along the way, we derive an elementary eigenvector perturbation bound for probability transition matrices, which parallels the Davis–Kahan $\sin \Theta$ theorem for symmetric matrices. This also allows us to close the gap between the ℓ_2 error upper bound for the spectral method and the minimax lower limit.

REFERENCES

- ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2017). Entrywise eigenvector analysis of random matrices with low expected rank. ArXiv preprint. Available at [ArXiv:1709.09565](https://arxiv.org/abs/1709.09565).
- AGARWAL, A., AGARWAL, S., ASSADI, S. and KHANNA, S. (2017). Learning with limited rounds of adaptivity: Coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In *Conference on Learning Theory* 39–75.
- AMMAR, A. and SHAH, D. (2011). Ranking: Compare, don’t score. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 776–783. DOI:10.1109/Allerton.2011.6120246.
- AMMAR, A. and SHAH, D. (2012). Efficient rank aggregation using partial data. In *SIGMETRICS* 40 355–366. ACM, New York.

MSC2010 subject classifications. Primary 62F07; secondary 62B10.

Key words and phrases. Top- K ranking, pairwise comparisons, spectral method, regularized MLE, entrywise perturbation, leave-one-out analysis, reversible Markov chains.

- BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems. RecSys '10* 119–126. ACM, New York.
- BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. [MR0070925](#)
- BUBECK, S. (2015). Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.* **8** 231–357.
- BUSA-FEKETE, R., SZÖRÉNYI, B., WENG, P., CHENG, W. and HÜLLERMEIER, E. (2013). Top- k selection based on adaptive sampling of noisy preferences. In *International Conference on Machine Learning*.
- CHEN, Y. and CANDES, E. (2016). The projected power method: An efficient algorithm for joint alignment from pairwise differences. *Comm. Pure Appl. Math.* To appear.
- CHEN, Y. and CANDÈS, E. J. (2017). Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Comm. Pure Appl. Math.* **70** 822–883. [MR3628877](#)
- CHEN, Y. and SUH, C. (2015). Spectral MLE: Top- K rank aggregation from pairwise comparisons. In *International Conference on Machine Learning* 371–380.
- CHEN, X., BENNETT, P. N., COLLINS-THOMPSON, K. and HORVITZ, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In *ACM International Conference on Web Search and Data Mining* 193–202. ACM, New York.
- CHEN, X., GOPI, S., MAO, J. and SCHNEIDER, J. (2017). Competitive analysis of the top- K ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 1245–1264. SIAM, Philadelphia, PA. [MR3627810](#)
- CHEN, Y., CHI, Y., FAN, J. and MA, C. (2018). Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. Available at [ArXiv:1803.07726](#).
- CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Supplement to “Spectral method and regularized MLE are both optimal for top- K ranking.” DOI:10.1214/18-AOS1745SUPP.
- CHUNG, F. R. K. (1997). *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics **92**. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the Amer. Math. Soc., Providence, RI. [MR1421568](#)
- DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. [MR0264450](#)
- DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the Web. In *International Conference on World Wide Web* 613–622.
- EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. [MR3748322](#)
- ELDRIDGE, J., BELKIN, M. and WANG, Y. (2017). Unperturbed: Spectral analysis beyond Davis–Kahan. ArXiv preprint. Available at [ArXiv:1706.06516](#).
- FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application. *J. Mach. Learn. Res.* **18** 1–42.
- FORD, L. R. JR. (1957). Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly* **64** 28–33. [MR0097876](#)
- HAJEK, B., OH, S. and XU, J. (2014). Minimax-optimal inference from partial rankings. In *Neural Information Processing Systems* 1475–1483.
- HECKEL, R., SHAH, N. B., RAMCHANDRAN, K. and WAINWRIGHT, M. J. (2016). Active ranking from pairwise comparisons and when parametric assumptions don’t help. ArXiv preprint. Available at [ArXiv:1606.08842](#).
- HUNTER, D. R. (2004). MM algorithms for generalized Bradley–Terry models. *Ann. Statist.* **32** 384–406. [MR2051012](#)
- JAMIESON, K. G. and NOWAK, R. D. (2011). Active ranking using pairwise comparisons. In *Neural Information Processing Systems* 2240–2248.

- JANG, M., KIM, S., SUH, C. and OH, S. (2016). Top- K ranking from pairwise comparisons: When spectral ranking is optimal. ArXiv preprint. Available at [arXiv:1603.04153](#).
- JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. [MR3851749](#)
- JIANG, X., LIM, L.-H., YAO, Y. and YE, Y. (2011). Statistical ranking and combinatorial Hodge theory. *Math. Program.* **127** 203–244. [MR2776715](#)
- KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. [MR2678022](#)
- KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. [MR3573302](#)
- KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII. Progress in Probability* **71** 397–423. Springer, Cham. [MR3565274](#)
- LU, Y. and NEGAHBAN, S. N. (2014). Individualized rank aggregation using nuclear norm regularization. ArXiv preprint. Available at [ArXiv:1410.0860](#).
- LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York; Chapman & Hall, London. [MR0108411](#)
- MA, C., WANG, K., CHI, Y. and CHEN, Y. (2017). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. ArXiv preprint. Available at [ArXiv:1711.10467](#).
- MASSE, K. (1997). Statistical models applied to the rating of sports teams. Technical Report, Bluefield College, Bluefield, VA.
- NEGAHBAN, S., OH, S. and SHAH, D. (2017). Rank centrality: Ranking from pairwise comparisons. *Oper. Res.* **65** 266–287. [MR3613103](#)
- NEGAHBAN, S., OH, S., THEKUMPARAMPIL, K. K. and XU, J. (2017). Learning from comparisons and choices. ArXiv preprint. Available at [ArXiv:1704.07228](#).
- PANANJADY, A., MAO, C., MUTHUKUMAR, V., WAINWRIGHT, M. J. and COURTADE, T. A. (2017). Worst-case vs average-case design for estimation from fixed pairwise comparisons. ArXiv preprint. Available at [ArXiv:1707.06217](#).
- RAJKUMAR, A. and AGARWAL, S. (2014). A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *International Conference on Machine Learning* I-118–I-126.
- RAJKUMAR, A. and AGARWAL, S. (2016). When can we rank well from comparisons of $O(n \log n)$ non-actively chosen pairs? In *Conference on Learning Theory* 1376–1401.
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](#)
- SHAH, N. B. and WAINWRIGHT, M. J. (2015). Simple, robust and optimal ranking from pairwise comparisons. ArXiv preprint. Available at [arXiv:1512.08949](#).
- SHAH, N. B., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. J. (2017). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inform. Theory* **63** 934–959. [MR3604649](#)
- SOUFIANI, H. A., CHEN, W. Z., PARKES, D. C. and XIA, L. (2013). Generalized method-of-moments for rank aggregation. In *Proceedings of the 26th International Conference on Neural Information Processing Systems. NIPS’13* 2706–2714.
- SUH, C., TAN, V. Y. F. and ZHAO, R. (2017). Adversarial top- K ranking. *IEEE Trans. Inform. Theory* **63** 2201–2225. [MR3626865](#)
- SUR, P., CHEN, Y. and CANDÈS, E. J. (2017). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled Chi-square. ArXiv preprint. Available at [ArXiv:1706.01191](#).

- TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.
- ZHONG, Y. and BOUMAL, N. (2017). Near-optimal bounds for phase synchronization. Available at [arXiv:1703.06605](https://arxiv.org/abs/1703.06605).

NEGATIVE ASSOCIATION, ORDERING AND CONVERGENCE OF RESAMPLING METHODS

BY MATHIEU GERBER*, NICOLAS CHOPIN^{†,1} AND NICK WHITELEY*

University of Bristol and CREST-ENSAE[†]*

We study convergence and convergence rates for resampling schemes. Our first main result is a general consistency theorem based on the notion of negative association, which is applied to establish the almost sure weak convergence of measures output from Kitagawa’s [J. Comput. Graph. Statist. **5** (1996) 1–25] stratified resampling method. Carpenter, Ckiffird and Fearnhead’s [IEE Proc. Radar Sonar Navig. **146** (1999) 2–7] systematic resampling method is similar in structure but can fail to converge depending on the order of the input samples. We introduce a new resampling algorithm based on a stochastic rounding technique of [In 42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001) (2001) 588–597 IEEE Computer Soc.], which shares some attractive properties of systematic resampling, but which exhibits negative association and, therefore, converges irrespective of the order of the input samples. We confirm a conjecture made by [J. Comput. Graph. Statist. **5** (1996) 1–25] that ordering input samples by their states in $\mathbb{R}$ yields a faster rate of convergence; we establish that when particles are ordered using the Hilbert curve in $\mathbb{R}^d$, the variance of the resampling error is $\mathcal{O}(N^{-(1+1/d)})$ under mild conditions, where N is the number of particles. We use these results to establish asymptotic properties of particle algorithms based on resampling schemes that differ from multinomial resampling.

REFERENCES

- AISTLEITNER, C. and DICK, J. (2015). Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arith.* **167** 143–171. [MR3312093](#)
- ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. [MR2758115](#)
- BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statistics* **98**. Springer, New York. [MR2195545](#)
- BERTAIL, P., CHAUTRU, E. and CLÉMENÇON, S. (2017). Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.* **44** 97–111. [MR3619696](#)
- BICKEL, P. J. and MILLAR, P. W. (1992). Uniform convergence of probability measures on classes of functions. *Statist. Sinica* **2** 1–15. [MR1152295](#)
- BROWN KRAMER, J., CUTLER, J. and RADCLIFFE, A. J. (2011). Negative dependence and Srinivasan’s sampling process. *Combin. Probab. Comput.* **20** 347–361. [MR2784632](#)
- CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. [MR2159833](#)

MSC2010 subject classifications. Primary 62G09, 62M20; secondary 60G35.

Key words and phrases. Negative association, resampling methods, particle filtering.

- CARPENTER, J., CLIFFORD, P. and FEARNHEAD, P. (1999). Improved particle filter for nonlinear problems. *IEE Proc. Radar Sonar Navig.* **146** 2–7.
- CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. [MR2153989](#)
- CORNEBISE, J., MOULINES, É. and OLSSON, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18** 461–480. [MR2461889](#)
- CRISAN, D. (2001). Particle filters—A theoretical perspective. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas and N. J. Gordon, eds.) 17–41. Springer, New York. [MR1847785](#)
- CRISAN, D. and DOUCET, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **50** 736–746. [MR1895071](#)
- DEL MORAL, P. and DOUCET, A. (2003). On a class of genealogical and interacting Metropolis models. In *Séminaire de Probabilités XXXVII. Lecture Notes in Math.* **1832** 415–446. Springer, Berlin. [MR2053058](#)
- DEVILLE, J.-C. and TILLÉ, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika* **85** 89–101. [MR1627234](#)
- DOUC, R., CAPPÉ, O. and MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis* 64–69. IEEE, New York.
- DOUC, R., MOULINES, É. and OLSSON, J. (2009). Optimality of the auxiliary particle filter. *Probab. Math. Statist.* **29** 1–28. [MR2552996](#)
- DOUCET, A., DE FREITAS, N. and GORDON, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- DUBHASHI, D., JONASSON, J. and RANJAN, D. (2007). Positive influence and negative dependence. *Combin. Probab. Comput.* **16** 29–41. [MR2286510](#)
- DUBHASHI, D. and RANJAN, D. (1998). Balls and bins: A study in negative dependence. *Random Structures Algorithms* **13** 99–124. [MR1642566](#)
- GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 509–579. [MR3351446](#)
- GERBER, M. and CHOPIN, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli* **23** 2951–2987. [MR3654796](#)
- GERBER, M., CHOPIN, N. and WHITELEY, N. (2019). Supplement to “Negative association, ordering and convergence of resampling methods.” DOI:[10.1214/18-AOS1746SUPP](#).
- GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Commun. Radar Signal Process.* **140** 107–113.
- GUARNIERO, P., JOHANSEN, A. M. and LEE, A. (2017). The iterated auxiliary particle filter. *J. Amer. Statist. Assoc.* **112** 1636–1647. [MR3750887](#)
- HE, Z. and OWEN, A. B. (2016). Extensible grids: Uniform sampling on a space filling curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 917–931. [MR3534356](#)
- JAFFARD, S. and NICOLAY, S. (2009). Pointwise smoothness of space-filling functions. *Appl. Comput. Harmon. Anal.* **26** 181–199. [MR2490213](#)
- JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. [MR0684886](#)
- JOHANSEN, A. M. and DOUCET, A. (2008). A note on auxiliary particle filters. *Statist. Probab. Lett.* **78** 1498–1504. [MR2528342](#)
- KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25. [MR1380850](#)
- KUIPERS, L. and NIEDERREITER, H. (1974). *Uniform Distribution of Sequences: Pure and Applied Mathematics*. Wiley-Interscience [John Wiley & Sons], New York–London–Sydney. [MR0419394](#)
- KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. [MR2211077](#)

- L'ECUYER, P. and LEMIEUX, C. (2000). Variance reduction via lattice rules. *Manage. Sci.* **46** 1214–1235.
- LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. [MR1649198](#)
- MADOW, W. G. and MADOW, L. H. (1944). On the theory of systematic sampling. I. *Ann. Math. Stat.* **15** 1–24. [MR0009836](#)
- NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics* **63**. SIAM, Philadelphia, PA. [MR1172997](#)
- PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. [MR1702328](#)
- PÓLYA, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **8** 171–181. [MR1544437](#)
- SCHRETTER, C., HE, Z., GERBER, M., CHOPIN, N. and NIEDERREITER, H. (2016). Van der Corput and golden ratio sequences along the Hilbert space-filling curve. In *Monte Carlo and Quasi-Monte Carlo Methods. Springer Proc. Math. Stat.* **163** 531–544. Springer, Cham. [MR3536680](#)
- SHAO, Q.-M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* **13** 343–356. [MR1777538](#)
- SRINIVASAN, A. (2001). Distributions on level-sets with applications to approximation algorithms. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 588–597. IEEE Computer Soc., Los Alamitos, CA. [MR1948748](#)
- STEIN, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29** 143–151. [MR0887702](#)
- TILLÉ, Y. (2006). *Sampling Algorithms*. Springer, New York. [MR2225036](#)
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- ZUMBUSCH, G. (2003). *Parallel Multilevel Methods*. Springer, Berlin.

ON DEEP LEARNING AS A REMEDY FOR THE CURSE OF DIMENSIONALITY IN NONPARAMETRIC REGRESSION

BY BENEDIKT BAUER AND MICHAEL KOHLER

Technische Universität Darmstadt

Assuming that a smoothness condition and a suitable restriction on the structure of the regression function hold, it is shown that least squares estimates based on multilayer feedforward neural networks are able to circumvent the curse of dimensionality in nonparametric regression. The proof is based on new approximation results concerning multilayer feedforward neural networks with bounded weights and a bounded number of hidden neurons. The estimates are compared with various other approaches by using simulated data.

REFERENCES

- ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning: Theoretical Foundations*. Cambridge Univ. Press, Cambridge. [MR1741038](#)
- BAGIROV, A. M., CLAUSEN, C. and KOHLER, M. (2009). Estimation of a regression function by maxima of minima of linear functions. *IEEE Trans. Inform. Theory* **55** 833–845. [MR2597271](#)
- BARRON, A. R. (1991). Complexity regularization with application to artificial neural networks. In *Nonparametric Functional Estimation and Related Topics (Spetses, 1990)*. NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **335** 561–576. Kluwer Academic, Dordrecht. [MR1154352](#)
- BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory* **39** 930–945. [MR1237720](#)
- BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.
- BAUER, B. and KOHLER, M. (2019). Supplement to “On deep learning as a remedy for the curse of dimensionality in nonparametric regression.” DOI:[10.1214/18-AOS1747SUPPA](#), DOI:[10.1214/18-AOS1747SUPPB](#).
- DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. Springer, New York. [MR1383093](#)
- ELDAN, R. and SHAMIR, O. (2015). The power of depth for feedforward neural networks. Arxiv preprint.
- FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. [MR0650892](#)
- GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York. [MR1920390](#)
- HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171](#)
- HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. [MR1134488](#)

MSC2010 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Curse of dimensionality, neural networks, nonparametric regression, rate of convergence.

- HAYKIN, S. O. (2008). *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, New York.
- HERTZ, J., KROGH, A. and PALMER, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, CA. [MR1096298](#)
- HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. [MR2382659](#)
- KOHLER, M. and KRZYŻAK, A. (2005). Adaptive regression estimation with multilayer feedforward neural networks. *J. Nonparametr. Stat.* **17** 891–913. [MR2192165](#)
- KOHLER, M. and KRZYŻAK, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Trans. Inform. Theory* **63** 1620–1630. [MR3625984](#)
- KONG, E. and XIA, Y. (2007). Variable selection for the single-index model. *Biometrika* **94** 217–229. [MR2367831](#)
- LAZZARO, D. and MONTEFUSCO, L. B. (2002). Radial basis functions for the multivariate interpolation of large scattered data sets. *J. Comput. Appl. Math.* **140** 521–536.
- LUGOSI, G. and ZEGER, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Trans. Inform. Theory* **41** 677–687. [MR1331260](#)
- MCCAFFREY, D. F. and GALLANT, A. R. (1994). Convergence rates for single hidden layer feedforward networks. *Neural Netw.* **7** 147–158.
- MHASKAR, H. N. and POGGIO, T. (2016). Deep vs. shallow networks: An approximation theory perspective. *Anal. Appl. (Singap.)* **14** 829–848. [MR3564936](#)
- MIELNICZUK, J. and TYRCHA, J. (1993). Consistency of multilayer perceptron regression estimators. *Neural Netw.* **6** 1019–1022.
- RIPLEY, B. D. (2008). *Pattern Recognition and Neural Networks*. Cambridge Univ. Press, Cambridge. Reprint of the 1996 original. [MR2451352](#)
- SCHMIDHUBER, J. (2015). Deep learning in neural networks: An overview. *Neural Netw.* **61** 85–117.
- SCHMIDT-HIEBER, J. (2017). Nonparametric regression using deep neural networks with ReLU activation function. Available at [arXiv:1708.06633v2](#).
- STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. [MR0673642](#)
- STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. [MR0790566](#)
- STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. [MR1272079](#)
- YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. [MR1951258](#)

CONVERGENCE RATES OF LEAST SQUARES REGRESSION ESTIMATORS WITH HEAVY-TAILED ERRORS

BY QIYANG HAN AND JON A. WELLNER¹

University of Washington

We study the performance of the least squares estimator (LSE) in a general nonparametric regression model, when the errors are independent of the covariates but may only have a p th moment ($p \geq 1$). In such a heavy-tailed regression setting, we show that if the model satisfies a standard “entropy condition” with exponent $\alpha \in (0, 2)$, then the L_2 loss of the LSE converges at a rate

$$\mathcal{O}_{\mathbf{P}}(n^{-\frac{1}{2+\alpha}} \vee n^{-\frac{1}{2} + \frac{1}{2p}}).$$

Such a rate cannot be improved under the entropy condition alone.

This rate quantifies both some positive and negative aspects of the LSE in a heavy-tailed regression setting. On the positive side, as long as the errors have $p \geq 1 + 2/\alpha$ moments, the L_2 loss of the LSE converges at the same rate as if the errors are Gaussian. On the negative side, if $p < 1 + 2/\alpha$, there are (many) hard models at any entropy level α for which the L_2 loss of the LSE converges at a strictly slower rate than other robust estimators.

The validity of the above rate relies crucially on the independence of the covariates and the errors. In fact, the L_2 loss of the LSE can converge arbitrarily slowly when the independence fails.

The key technical ingredient is a new multiplier inequality that gives sharp bounds for the “multiplier empirical process” associated with the LSE. We further give an application to the sparse linear regression model with heavy-tailed covariates and errors to demonstrate the scope of this new inequality.

REFERENCES

- [1] ALEXANDER, K. S. (1985). The nonexistence of a universal multiplier moment for the central limit theorem. In *Probability in Banach Spaces, V (Medford, Mass., 1984). Lecture Notes in Math.* **1153** 15–16. Springer, Berlin. [MR0821974](#)
- [2] ANDERSEN, N. T., GINÉ, E. and ZINN, J. (1988). The central limit theorem for empirical processes under local conditions: The case of Radon infinitely divisible limits without Gaussian component. *Trans. Amer. Math. Soc.* **308** 603–635. [MR0930076](#)
- [3] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. [MR2246369](#)
- [4] AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. [MR2906886](#)

MSC2010 subject classifications. Primary 60E15; secondary 62G05.

Key words and phrases. Multiplier empirical process, multiplier inequality, nonparametric regression, least squares estimation, sparse linear regression, heavy-tailed errors.

- [5] BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. [MR0963829](#)
- [6] 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](#)
- [7] BARAUD, Y., BIRGÉ, L. and SART, M. (2017). A new method for estimation and model selection: ρ -estimation. *Invent. Math.* **207** 425–517. [MR3595933](#)
- [8] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. [MR2166554](#)
- [9] BARTLETT, P. L. and MENDELSON, S. (2006). Empirical minimization. *Probab. Theory Related Fields* **135** 311–334. [MR2240689](#)
- [10] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. [MR1240719](#)
- [11] BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37** 157–183. [MR2488348](#)
- [12] BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. [MR3405602](#)
- [13] BUBECK, S., CESA-BIANCHI, N. and LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Inform. Theory* **59** 7711–7717. [MR3124669](#)
- [14] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- [15] CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. [MR3052407](#)
- [16] CATONI, O. (2016). PAC-Bayesian bounds for the Gram matrix and least squares regression with a random design. Preprint. Available at [arXiv:1603.05229](#).
- [17] 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](#)
- [18] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. [MR3576558](#)
- [19] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. [MR3445285](#)
- [20] GAO, X. and HUANG, J. (2010). Asymptotic analysis of high-dimensional LAD regression with Lasso. *Statist. Sinica* **20** 1485–1506. [MR2777333](#)
- [21] GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216. [MR2243881](#)
- [22] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U -statistics. In *High Dimensional Probability, II (Seattle, WA, 1999)*. *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. [MR1857312](#)
- [23] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **40**. Cambridge Univ. Press, New York. [MR3588285](#)
- [24] GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–998. [MR0757767](#)
- [25] GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. In *Probability and Banach Spaces (Zaragoza, 1985)*. *Lecture Notes in Math.* **1221** 50–113. Springer, Berlin. [MR0875007](#)
- [26] HAN, Q. and WELLNER, J. A. (2019). Supplement to “Convergence rates of least squares regression estimators with heavy-tailed errors.” DOI:[10.1214/18-AOS1748SUPP](#).

- [27] HAN, Q. and WELLNER, J. A. (2018). Robustness of shape-restricted regression estimators: An envelope perspective. Preprint. Available at [arXiv:1805.02542](https://arxiv.org/abs/1805.02542).
- [28] HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18, 40. [MR3491112](#)
- [29] JOLY, E., LUGOSI, G. and OLIVEIRA, R. I. (2017). On the estimation of the mean of a random vector. *Electron. J. Stat.* **11** 440–451. [MR3619312](#)
- [30] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. [MR2329442](#)
- [31] KOLTCHINSKII, V. and PANCHENKO, D. (2000). Rademacher processes and bounding the risk of function learning. In *High Dimensional Probability, II* (Seattle, WA, 1999). *Progress in Probability* **47** 443–457. Birkhäuser, Boston, MA. [MR1857339](#)
- [32] KOROSTEL'EV, A. P. and TSYBAKOV, A. B. (1992). Asymptotically minimax image reconstruction problems. In *Topics in Nonparametric Estimation. Adv. Soviet Math.* **12** 45–86. Amer. Math. Soc., Providence, RI. [MR1191691](#)
- [33] KOROSTEL'EV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. Springer, New York. [MR1226450](#)
- [34] LECUÉ, G. and MENDELSON, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc. (JEMS)* **19** 881–904. [MR3612870](#)
- [35] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann. Statist.* **46** 611–641. [MR3782379](#)
- [36] LEDOUX, M. and TALAGRAND, M. (1986). Conditions d'intégrabilité pour les multiplicateurs dans le TLC banachique. *Ann. Probab.* **14** 916–921. [MR0841593](#)
- [37] LEDOUX, M. and TALAGRAND, M. (2011). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin. [MR2814399](#)
- [38] LUGOSI, G. and MENDELSON, S. (2018). Regularization, sparse recovery, and median-of-means tournaments. *Bernoulli*. To appear.
- [39] LUGOSI, G. and MENDELSON, S. (2018). Risk minimization by median-of-means tournaments. *J. Eur. Math. Soc. (JEMS)*. To appear.
- [40] LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. [MR3909950](#)
- [41] MASON, D. M. (1983). The asymptotic distribution of weighted empirical distribution functions. *Stochastic Process. Appl.* **15** 99–109. [MR0694539](#)
- [42] MASSART, P. and NÉDÉLEC, É. (2006). Risk bounds for statistical learning. *Ann. Statist.* **34** 2326–2366. [MR2291502](#)
- [43] MASSART, P. and RIO, E. (1998). A uniform Marcinkiewicz–Zygmund strong law of large numbers for empirical processes. In *Asymptotic Methods in Probability and Statistics* (Ottawa, ON, 1997) 199–211. North-Holland, Amsterdam. [MR1661481](#)
- [44] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25. [MR3367000](#)
- [45] MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. [MR3565471](#)
- [46] MENDELSON, S. (2017). “Local” vs. “global” parameters—breaking the Gaussian complexity barrier. *Ann. Statist.* **45** 1835–1862. [MR3718154](#)
- [47] MENDELSON, S. (2017). On aggregation for heavy-tailed classes. *Probab. Theory Related Fields* **168** 641–674. [MR3663627](#)
- [48] MENDELSON, S. (2017). Extending the small-ball method. Preprint. Available at [arXiv:1709.00843](https://arxiv.org/abs/1709.00843).
- [49] MENDELSON, S. (2017). On multiplier processes under weak moment assumptions. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **2169** 301–318. Springer, Cham. [MR3645129](#)
- [50] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. [MR3378468](#)

- [51] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. [MR3161450](#)
- [52] POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199. [MR1128411](#)
- [53] PORTNOY, S. and KOENKER, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300. [MR1619189](#)
- [54] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. [MR2882274](#)
- [55] SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics. Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. [MR3396731](#)
- [56] SIVAKUMAR, V., BANERJEE, A. and RAVIKUMAR, P. K. (2015). Beyond sub-Gaussian measurements: High-dimensional structured estimation with sub-exponential designs. In *Advances in Neural Information Processing Systems* 2206–2214.
- [57] STRASSEN, V. and DUDLEY, R. M. (1969). The central limit theorem and ε -entropy. In *Probability and Information Theory (Proc. Internat. Sympos., McMaster Univ., Hamilton, Ont., 1968)* 224–231. Springer, Berlin. [MR0279872](#)
- [58] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics* **60**. Springer, Heidelberg. [MR3184689](#)
- [59] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [60] VAN DE GEER, S. (1987). A new approach to least-squares estimation, with applications. *Ann. Statist.* **15** 587–602. [MR0888427](#)
- [61] VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. [MR1056343](#)
- [62] 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](#)
- [63] VAN DE GEER, S. and WEGKAMP, M. (1996). Consistency for the least squares estimator in nonparametric regression. *Ann. Statist.* **24** 2513–2523. [MR1425964](#)
- [64] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge.
- [65] VAN DER VAART, A. and WELLNER, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.* **5** 192–203. [MR2792551](#)
- [66] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. [MR1385671](#)
- [67] YANG, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli* **7** 633–655. [MR1849372](#)
- [68] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. [MR1742500](#)
- [69] YAO, Q. W. (1993). Tests for change-points with epidemic alternatives. *Biometrika* **80** 179–191. [MR1225223](#)
- [70] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. [MR1902898](#)

CONVERGENCE COMPLEXITY ANALYSIS OF ALBERT AND CHIB'S ALGORITHM FOR BAYESIAN PROBIT REGRESSION

BY QIAN QIN AND JAMES P. HOBERT¹

University of Florida

The use of MCMC algorithms in high dimensional Bayesian problems has become routine. This has spurred so-called convergence complexity analysis, the goal of which is to ascertain how the convergence rate of a Monte Carlo Markov chain scales with sample size, n , and/or number of covariates, p . This article provides a thorough convergence complexity analysis of Albert and Chib's [*J. Amer. Statist. Assoc.* **88** (1993) 669–679] data augmentation algorithm for the Bayesian probit regression model. The main tools used in this analysis are drift and minorization conditions. The usual pitfalls associated with this type of analysis are avoided by utilizing centered drift functions, which are minimized in high posterior probability regions, and by using a new technique to suppress high-dimensionality in the construction of minorization conditions. The main result is that the geometric convergence rate of the underlying Markov chain is bounded below 1 both as $n \rightarrow \infty$ (with p fixed), and as $p \rightarrow \infty$ (with n fixed). Furthermore, the first computable bounds on the total variation distance to stationarity are byproducts of the asymptotic analysis.

REFERENCES

- ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. [MR1224394](#)
- BARAGATTI, M. and POMMERET, D. (2012). A study of variable selection using g -prior distribution with ridge parameter. *Comput. Statist. Data Anal.* **56** 1920–1934. [MR2892387](#)
- BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. [MR2114987](#)
- CHAKRABORTY, S. and KHARE, K. (2017). Convergence properties of Gibbs samplers for Bayesian probit regression with proper priors. *Electron. J. Stat.* **11** 177–210. [MR3604022](#)
- CHEN, M.-H. and SHAO, Q.-M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proc. Amer. Math. Soc.* **129** 293–302. [MR1694452](#)
- DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials (with discussion). *Statist. Sci.* **23** 151–178. [MR2446500](#)
- DURMUS, A. and MOULINES, E. (2016). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. [arXiv:1605.01559](#).
- FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. [MR2516823](#)
- FORT, G., MOULINES, E., ROBERTS, G. O. and ROSENTHAL, J. S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40** 123–146. [MR1953771](#)

Primary 60J05; secondary 65C05.

Key words and phrases. Drift condition, geometric ergodicity, high dimensional inference, large p -small n , Markov chain Monte Carlo, minorization condition.

- GUPTA, M. and IBRAHIM, J. G. (2007). Variable selection in regression mixture modeling for the discovery of gene regulatory networks. *J. Amer. Statist. Assoc.* **102** 867–880. [MR2411650](#)
- HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris’ ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser/Springer Basel AG, Basel. [MR2857021](#)
- JOHNDROW, J. E., SMITH, A., PILLAI, N. and DUNSON, D. B. (2018). MCMC for imbalanced categorical data. *J. Amer. Statist. Assoc.* To appear. Available at [arXiv:1605.05798](#).
- JONES, G. L. (2001). *Convergence Rates and Monte Carlo Standard Errors for Markov Chain Monte Carlo Algorithms*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, Univ. Florida. [MR2702583](#)
- 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](#)
- ŁATUSZYŃSKI, K., MIAOJEDOW, B. and NIEMIRO, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli* **19** 2033–2066. [MR3129043](#)
- MARCHEV, D. and HOBERT, J. P. (2004). Geometric ergodicity of van Dyk and Meng’s algorithm for the multivariate Student’s t model. *J. Amer. Statist. Assoc.* **99** 228–238. [MR2054301](#)
- MEYN, S. P. and TWEEDIE, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. [MR1304770](#)
- MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2509253](#)
- QIN, Q. and HOBERT, J. P. (2018). Supplement to “Convergence complexity analysis of Albert and Chib’s algorithm for Bayesian probit regression.” DOI:[10.1214/18-AOS1749SUPP](#).
- RAJARATNAM, B. and SPARKS, D. (2015). MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. [arXiv:1508.00947](#).
- ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. [MR1448322](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Markov-chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canad. J. Statist.* **26** 5–31. [MR1624414](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. [MR1858413](#)
- ROBERTS, G. O. and TWEEDIE, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80** 211–229. [MR1682243](#)
- ROBERTS, G. O. and TWEEDIE, R. L. (2001). Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *J. Appl. Probab.* **38A** 37–41. [MR1915532](#)
- ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. [MR1340509](#)
- ROY, V. and HOBERT, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 607–623. [MR2370071](#)
- ROY, V. and HOBERT, J. P. (2010). On Monte Carlo methods for Bayesian multivariate regression models with heavy-tailed errors. *J. Multivariate Anal.* **101** 1190–1202. [MR2595301](#)
- RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill Book Co., New York-Auckland-Düsseldorf. [MR0385023](#)
- SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** 93–133. [MR1003059](#)
- VATS, D. (2017). Geometric ergodicity of Gibbs samplers in Bayesian penalized regression models. *Electron. J. Stat.* **11** 4033–4064. [MR3714307](#)
- YANG, J. and ROSENTHAL, J. S. (2017). Complexity results for MCMC derived from quantitative bounds. [arXiv:1708.00829](#).
- YANG, A.-J. and SONG, X.-Y. (2009). Bayesian variable selection for disease classification using gene expression data. *Bioinformatics* **26** 215–222.

YANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. [MR3576552](#)

ON TESTING CONDITIONAL QUALITATIVE TREATMENT EFFECTS

BY CHENGCHUN SHI¹, RUI SONG¹ AND WENBIN LU²

North Carolina State University

Precision medicine is an emerging medical paradigm that focuses on finding the most effective treatment strategy tailored for individual patients. In the literature, most of the existing works focused on estimating the optimal treatment regime. However, there has been less attention devoted to hypothesis testing regarding the optimal treatment regime. In this paper, we first introduce the notion of conditional qualitative treatment effects (CQTE) of a set of variables given another set of variables and provide a class of equivalent representations for the null hypothesis of no CQTE. The proposed definition of CQTE does not assume any parametric form for the optimal treatment rule and plays an important role for assessing the incremental value of a set of new variables in optimal treatment decision making conditional on an existing set of prescriptive variables. We then propose novel testing procedures for no CQTE based on kernel estimation of the conditional contrast functions. We show that our test statistics have asymptotically correct size and nonnegligible power against some nonstandard local alternatives. The empirical performance of the proposed tests are evaluated by simulations and an application to an AIDS data set.

REFERENCES

- ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609–666. [MR3043344](#)
- ANDREWS, D. W. K. and SHI, X. (2014). Nonparametric inference based on conditional moment inequalities. *J. Econometrics* **179** 31–45. [MR3153647](#)
- ARMSTRONG, T. B. and CHAN, H. P. (2016). Multiscale adaptive inference on conditional moment inequalities. *J. Econometrics* **194** 24–43. [MR3523517](#)
- CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.* **19** 317–343. [MR2757118](#)
- CHANG, M., LEE, S. and WHANG, Y.-J. (2015). Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. *Econom. J.* **18** 307–346. [MR3452611](#)
- CHERNOZHUKOV, V., LEE, S. and ROSEN, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica* **81** 667–737. [MR3043345](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FAN, A., LU, W. and SONG, R. (2016). Sequential advantage selection for optimal treatment regime. *Ann. Appl. Stat.* **10** 32–53. [MR3480486](#)

MSC2010 subject classifications. 62G08, 62G10.

Key words and phrases. Conditional qualitative treatment effects, kernel estimation, nonstandard local alternatives, optimal treatment decision making.

- FAN, C., LU, W., SONG, R. and ZHOU, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1565–1582. [MR3731676](#)
- GINÉ, E., MASON, D. M. and ZAITSEV, A. YU. (2003). The L_1 -norm density estimator process. *Ann. Probab.* **31** 719–768. [MR1964947](#)
- GUNTER, L., ZHU, J. and MURPHY, S. A. (2011). Variable selection for qualitative interactions. *Stat. Methodol.* **8** 42–55. [MR2741508](#)
- HSU, Y.-C. (2017). Consistent test for conditional treatment effects. *Econom. J.* **20** 1–22. [MR3636959](#)
- LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052. [MR1015136](#)
- LU, W., ZHANG, H. H. and ZENG, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.* **22** 493–504. [MR3190671](#)
- LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. [MR3476615](#)
- MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. [MR2537201](#)
- MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. [MR1983752](#)
- QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. [MR2816351](#)
- ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *J. Educ. Psychol.* **66** 688–701.
- SHERGIN, V. V. (1990). The central limit theorem for finitely dependent random variables. In *Probability Theory and Mathematical Statistics, Vol. II* (Vilnius, 1989) 424–431. Mokslas, Vilnius. [MR1153895](#)
- SHI, C., SONG, R. and LU, W. (2019). Supplement to “On testing conditional qualitative treatment effects.” DOI:[10.1214/18-AOS1750SUPP](#).
- WAND, M. P. and JONES, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Amer. Statist. Assoc.* **88** 520–528. [MR1224377](#)
- WANG, H. (2009). Forward regression for ultra-high dimensional variable screening. *J. Amer. Statist. Assoc.* **104** 1512–1524. [MR2750576](#)
- WATKINS, C. J. C. H. and DAYAN, P. (1992). Q-learning. *Mach. Learn.* **8** 279–292.
- WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. [MR0640163](#)
- ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. [MR3040007](#)
- ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. [MR3094445](#)
- ZHANG, Y., LABER, E. B., TSIATIS, A. and DAVIDIAN, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics* **71** 895–904. [MR3436715](#)
- ZHANG, Y., LABER, E. B., TSIATIS, A. and DAVIDIAN, M. (2016). Interpretable dynamic treatment regimes. Preprint. Available at [arXiv:1606.01472](#).
- ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. [MR3010898](#)
- ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. [MR3367249](#)

DYNAMIC NETWORK MODELS AND GRAPHON ESTIMATION

BY MARIANNA PENSKY¹

University of Central Florida

In the present paper, we consider a dynamic stochastic network model. The objective is estimation of the tensor of connection probabilities $\mathbf{A}$ when it is generated by a Dynamic Stochastic Block Model (DSBM) or a dynamic graphon. In particular, in the context of the DSBM, we derive a penalized least squares estimator $\hat{\mathbf{A}}$ of $\mathbf{A}$ and show that $\hat{\mathbf{A}}$ satisfies an oracle inequality and also attains minimax lower bounds for the risk. We extend those results to estimation of $\mathbf{A}$ when it is generated by a dynamic graphon function. The estimators constructed in the paper are adaptive to the unknown number of blocks in the context of the DSBM or to the smoothness of the graphon function. The technique relies on the vectorization of the model and leads to much simpler mathematical arguments than the ones used previously in the stationary set up. In addition, all results in the paper are nonasymptotic and allow a variety of extensions.

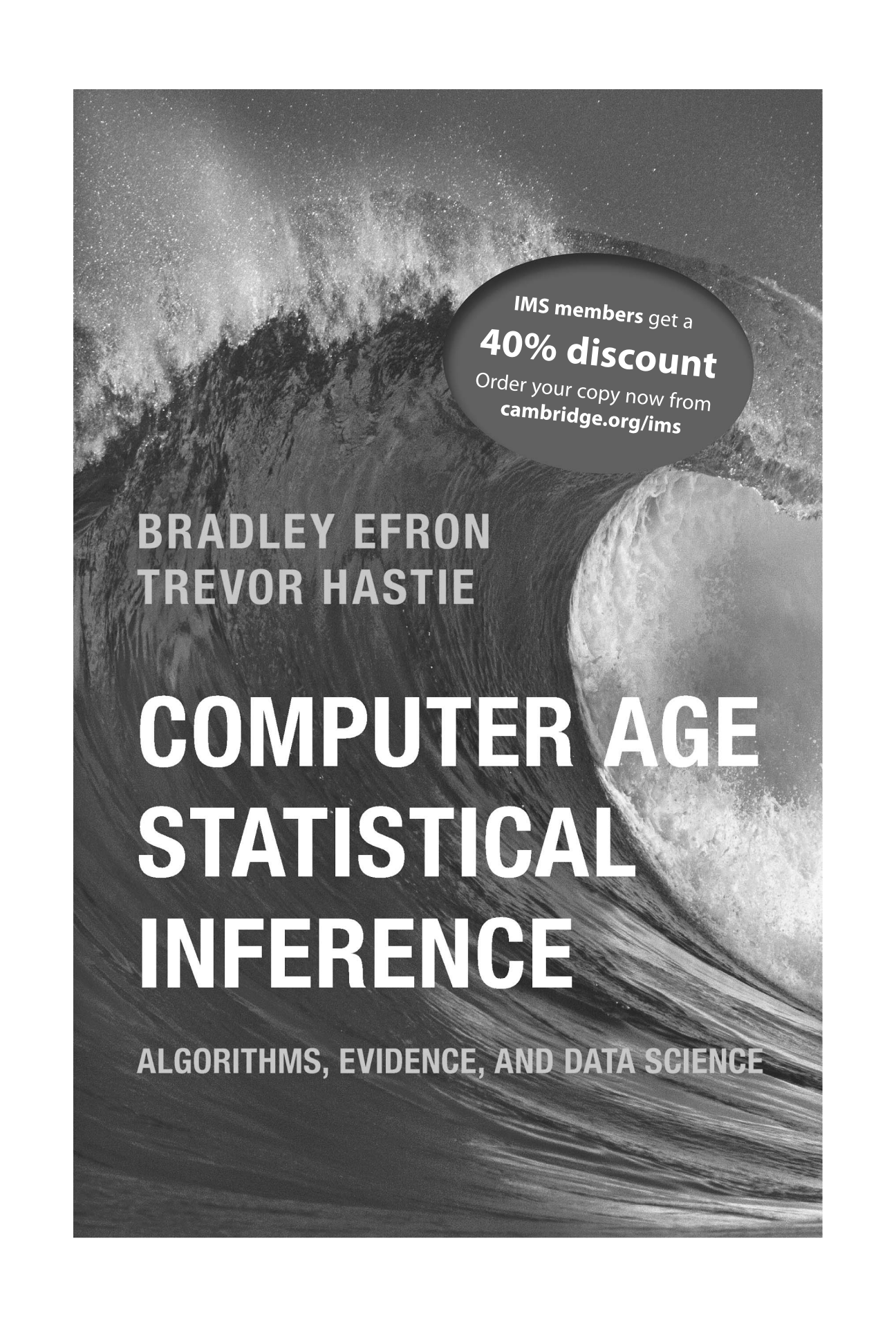
REFERENCES

- AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46** 149–179. [MR3766949](#)
- ANAGNOSTOPOULOS, A., LACKI, J., LATTANZI, S., LEONARDI, S. and MAHDIAN, M. (2016). Community detection on evolving graphs. In *NIPS* 3522–3530.
- BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
- BORGES, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2016). Consistent nonparametric estimation of heavy-tailed sparse graphs. Available at [arXiv:1508.06675v2](#).
- DURANTE, D. and DUNSON, D. B. (2016). Locally adaptive dynamic networks. *Ann. Appl. Stat.* **10** 2203–2232. [MR3592054](#)
- DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J. Amer. Statist. Assoc.* **112** 1516–1530. [MR3750873](#)
- GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. [MR3405606](#)
- GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* **17** Paper No. 161, 29. [MR3569248](#)
- GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45. [MR3687603](#)
- GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.
- GUPTA, A. K. and NAGAR, D. K. (2000). *Matrix Variate Distributions*. *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **104**. Chapman & Hall/CRC, Boca Raton, FL. [MR1738933](#)

MSC2010 subject classifications. Primary 60G05; secondary 05C80, 62F35.

Key words and phrases. Dynamic network, graphon, stochastic block model, nonparametric regression, minimax rate.

- HAN, Q., XU, K. S. and AIROLDI, E. M. (2015). Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on Machine Learning, Lille, France* 1511–1520.
- KLOPP, O. and PENSKEY, M. (2015). Sparse high-dimensional varying coefficient model: Nonasymptotic minimax study. *Ann. Statist.* **43** 1273–1299. [MR3346703](#)
- KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. [MR3611494](#)
- KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer Series in Statistics. Springer, New York. [MR2724362](#)
- KOLAR, M., SONG, L., AHMED, A. and XING, E. P. (2010). Estimating time-varying networks. *Ann. Appl. Stat.* **4** 94–123. [MR2758086](#)
- LEE, M., SHEN, H., HUANG, J. Z. and MARRON, J. S. (2010). Biclustering via sparse singular value decomposition. *Biometrics* **66** 1087–1095. [MR2758496](#)
- LEPSKI, O. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43** 1178–1242. [MR3346701](#)
- LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications **60**. Amer. Math. Soc., Providence, RI. [MR3012035](#)
- LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. [MR2274085](#)
- MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1119–1141. [MR3689311](#)
- MINHAS, S., HOFF, P. D. and WARDA, M. D. (2015). Relax, tensors are here: Dependencies in international processes. Available at [arXiv:1504.08218](#).
- OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.
- PENSKEY, M. (2019). Supplement to “Dynamic network models and graphon estimation.” DOI:[10.1214/18-AOS1751SUPP](#).
- XING, E. P., FU, W. and SONG, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **4** 535–566. [MR2758639](#)
- XU, K. S. (2015). Stochastic block transition models for dynamic networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS) 2015, San Diego, CA, USA, JMLR:W&CP* **38**.
- XU, K. S. and HERO, A. O. III (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **8** 552–562.
- YANG, T., CHI, Y., ZHU, S., GONG, Y. and JIN, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.* **82** 157–189. [MR3108191](#)
- ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. [MR3546450](#)



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

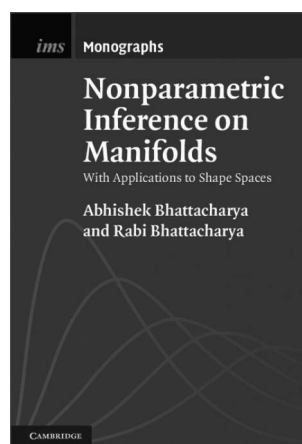
COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE



The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds *With Applications to Shape Spaces*

Abhishek Bhattacharya, Rabi Bhattacharya

This book introduces in a systematic manner a general nonparametric theory of statistics on manifolds, with emphasis on manifolds of shapes. The theory has important and varied applications in medical diagnostics, image analysis, and machine vision. An early chapter of examples establishes the effectiveness of the new methods and demonstrates how they outperform their parametric counterparts. Inference is developed for both intrinsic and extrinsic Fréchet means of probability distributions on manifolds, then applied to shape spaces defined as orbits of landmarks under a Lie group of transformations—in particular, similarity, reflection similarity, affine and projective transformations. In addition, nonparametric Bayesian theory is adapted and extended to manifolds for the purposes of density estimation, regression and classification. Ideal for statisticians who analyze manifold data and wish to develop their own methodology, this book is also of interest to probabilists, mathematicians, computer scientists and morphometricians with mathematical training.

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

Hardback price
US\$51.00
(non-member price
\$85.00)

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