THE ANNALS
of
STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

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.

ORTHOGONAL STATISTICAL LEARNING

BY DYLAN J. FOSTER[1,a] AND VASILIS SYRGKANIS[2,b]

[1]*Microsoft Research, New England,* [a]*dylanfoster@microsoft.com*

[2]*Department of Management Science and Engineering, Stanford University,* [b]*vsyrgk@stanford.edu*

We provide nonasymptotic excess risk guarantees for statistical learning in a setting where the population risk with respect to which we evaluate the target parameter depends on an unknown *nuisance parameter* that must be estimated from data. We analyze a two-stage sample splitting meta-algorithm that takes as input arbitrary estimation algorithms for the target parameter and nuisance parameter. We show that if the population risk satisfies a condition called *Neyman orthogonality*, the impact of the nuisance estimation error on the excess risk bound achieved by the meta-algorithm is of second order. Our theorem is agnostic to the particular algorithms used for the target and nuisance and only makes an assumption on their individual performance. This enables the use of a plethora of existing results from machine learning to give new guarantees for learning with a nuisance component. Moreover, by focusing on excess risk rather than parameter estimation, we can provide rates under weaker assumptions than in previous works and accommodate settings in which the target parameter belongs to a complex nonparametric class. We provide conditions on the metric entropy of the nuisance and target classes such that *oracle rates* of the same order, as if we knew the nuisance parameter, are achieved.

REFERENCES

[1] AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71** 1795–1843. MR2015420 https://doi.org/10.1111/1468-0262.00470

[2] AI, C. and CHEN, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *J. Econometrics* **141** 5–43. MR2411735 https://doi.org/10.1016/j.jeconom.2007.01.013

[3] AI, C. and CHEN, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *J. Econometrics* **170** 442–457. MR2970325 https://doi.org/10.1016/j.jeconom.2012.05.015

[4] ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

[5] ATHEY, S. and WAGER, S. (2017). Efficient policy learning. ArXiv preprint. Available at arXiv:1702.02896.

[6] AUDIBERT, J.-Y. (2008). Progressive mixture rules are deviation suboptimal. In *Advances in Neural Information Processing Systems* 41–48.

[7] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554 https://doi.org/10.1214/009053605000000282

[8] BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* **85** 233–298. MR3611771 https://doi.org/10.3982/ECTA12723

[9] BEYGELZIMER, A. and LANGFORD, J. (2009). The offset tree for learning with partial labels. In *Proceedings of the* 15*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 129–138. ACM, New York.

[10] BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. MR0663424

[11] BICKEL, P. J., KLAASSEN, C. A., BICKEL, P. J. and RITOV, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models* **4**. Johns Hopkins University Press, Baltimore.

[12] BLUNDELL, R., CHEN, X. and KRISTENSEN, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* **75** 1613–1669. MR2351452 https://doi.org/10.1111/j.1468-0262.2007.00808.x

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

[14] BOUSQUET, O., BOUCHERON, S. and LUGOSI, G. (2004). Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning* 169–207. Springer, Berlin.

[15] CASSEL, C. M., SÄRNDAL, C. E. and WRETMAN, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63** 615–620. MR0445666 https://doi.org/10.1093/biomet/63.3.615

[16] CHEN, X. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ*. **9** 39–84. MR3789729 https://doi.org/10.3982/QE722

[17] CHEN, X. and POUZO, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *J. Econometrics* **152** 46–60. MR2562763 https://doi.org/10.1016/j.jeconom.2009.02.002

[18] CHEN, X. and POUZO, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* **80** 277–321. MR2920758 https://doi.org/10.3982/ECTA7888

[19] CHEN, X. and POUZO, D. (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* **83** 1013–1079. MR3357484 https://doi.org/10.3982/ECTA10771

[20] CHEN, X. and WHITE, H. (1999). Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Trans*. *Inf*. *Theory* **45** 682–691. MR1677026 https://doi.org/10.1109/18.749011

[21] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J*. **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[22] CHERNOZHUKOV, V., ESCANCIANO, J. C., ICHIMURA, H., NEWEY, W. K. and ROBINS, J. M. (2022). Locally robust semiparametric estimation. *Econometrica* **90** 1501–1535. MR4467437 https://doi.org/10.3982/ecta16294

[23] CHERNOZHUKOV, V., GOLDMAN, M., SEMENOVA, V. and TADDY, M. (2017). Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels. ArXiv preprint. Available at arXiv:1712.09988.

[24] CHERNOZHUKOV, V., NEKIPELOV, D., SEMENOVA, V. and SYRGKANIS, V. (2018). Plug-in regularized estimation of high-dimensional parameters in nonlinear semiparametric models. ArXiv preprint. Available at arXiv:1806.04823.

[25] CHERNOZHUKOV, V., NEWEY, W. and ROBINS, J. (2018). Double/de-biased machine learning using regularized Riesz representers. ArXiv preprint. Available at arXiv:1802.08667.

[26] CHERNOZHUKOV, V., NEWEY, W. K., QUINTAS-MARTINEZ, V. and SYRGKANIS, V. (2021). Automatic debiased machine learning via neural nets for generalized linear regression. ArXiv preprint. Available at arXiv:2104.14737.

[27] CURTH, A., ALAA, A. M. and VAN DER SCHAAR, M. (2020). Estimating structural target functions using machine learning and influence functions. ArXiv preprint. Available at arXiv:2008.06461.

[28] DÍAZ, I. and VAN DER LAAN, M. J. (2013). Targeted data adaptive estimation of the causal dose-response curve. *J. Causal Inference* **1** 171–192. MR4289406 https://doi.org/10.1515/jci-2012-0005

[29] DUDÍK, M., LANGFORD, J. and LI, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the* 28*th International Conference on International Conference on Machine Learning* 1097–1104.

[30] FARRELL, M. H., LIANG, T. and MISRA, S. (2021). Deep neural networks for estimation and inference. *Econometrica* **89** 181–213. MR4220387 https://doi.org/10.3982/ecta16901

[31] FOSTER, D. J., KALE, S., LUO, H., MOHRI, M. and SRIDHARAN, K. (2018). Logistic regression: The importance of being improper. *Conf. Learn. Theory*.

[32] FOSTER, D. J. and SYRGKANIS, V. (2023). Supplement to "Orthogonal statistical learning." https://doi.org/10.1214/23-AOS2258SUPP

[33] FRIEDBERG, R., TIBSHIRANI, J., ATHEY, S. and WAGER, S. (2021). Local linear forests. *J. Comput. Graph. Statist*. **30** 503–517. MR4270520 https://doi.org/10.1080/10618600.2020.1831930

[34] HALL, P. and HOROWITZ, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Ann. Statist.* **33** 2904–2929. MR2253107 https://doi.org/10.1214/009053605000000714

[35] HIRSHBERG, D. A. and WAGER, S. (2021). Augmented minimax linear estimation. *Ann. Statist.* **49** 3206–3227. MR4352528 https://doi.org/10.1214/21-aos2080

[36] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York-Berlin. MR0620321

[37] KALLUS, N. and ZHOU, A. (2018). Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics* 1243–1251.

[38] KENNEDY, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and Their Applications in Public Health Research*. *ICSA Book Ser. Stat.* 141–167. Springer, Cham. MR3617956

[39] KENNEDY, E. H. (2020). Optimal doubly robust estimation of heterogeneous causal effects. ArXiv preprint. Available at arXiv:2004.14497.

[40] KENNEDY, E. H., LORCH, S. and SMALL, D. S. (2019). Robust causal inference with continuous instruments using the local instrumental variable curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 121–143. MR3904782 https://doi.org/10.1111/rssb.12300

[41] KENNEDY, E. H., MA, Z., MCHUGH, M. D. and SMALL, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1229–1245. MR3689316 https://doi.org/10.1111/rssb.12212

[42] KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562. MR0913573 https://doi.org/10.1214/aos/1176350609

[43] 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

[44] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

[45] KÜNZEL, S. R., SEKHON, J. S., BICKEL, P. J. and YU, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **116** 4156–4165.

[46] LECUÉ, G. and MENDELSON, S. (2016). Learning subgaussian classes: Upper and minimax bounds. In *Topics in Learning Theory* Societe Mathematique de France, Paris, France.

[47] LEVIT, B. Y. (1976). On the efficiency of a class of non-parametric estimates. *Theory Probab. Appl.* **20** 723–740.

[48] LIANG, T., RAKHLIN, A. and SRIDHARAN, K. (2015). Learning with square loss: Localization through offset Rademacher complexity. In *Proceedings of the* 28*th Conference on Learning Theory* 1260–1285.

[49] MAURER, A. and PONTIL, M. (2009). Empirical Bernstein bounds and sample variance penalization. In *The* 22*nd Conference on Learning Theory* (*COLT*).

[50] MENDELSON, S. (2011). Discrepancy, chaining and subgaussian processes. *Ann. Probab.* **39** 985–1026. MR2789581 https://doi.org/10.1214/10-AOP575

[51] MENDELSON, S. (2014). Learning without concentration. In *Conference on Learning Theory* (*COLT*) 25–39.

[52] MENDELSON, S. and NEEMAN, J. (2010). Regularization in kernel learning. *Ann. Statist.* **38** 526–565. MR2590050 https://doi.org/10.1214/09-AOS728

[53] NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237 https://doi.org/10.2307/2951752

[54] NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578. MR2000257 https://doi.org/10.1111/1468-0262.00459

[55] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics*: *The Harald Cramér Volume* (*Edited by Ulf Grenander*) 213–234. Almqvist & Wiksell, Stockholm. MR0112201

[56] NEYMAN, J. (1979). $C(\alpha)$ tests and their use. *Sankhyā Ser. A* **41** 1–21. MR0615037

[57] NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

[58] NING, Y., PENG, S. and IMAI, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika* **107** 533–554. MR4138975 https://doi.org/10.1093/biomet/asaa020

[59] OPRESCU, M., SYRGKANIS, V. and WU, Z. S. (2019). Orthogonal random forest for causal inference. In *International Conference on Machine Learning* 4932–4941.

[60] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **13**. Springer, New York-Berlin. MR0675954

[61] QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864

[62] RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679

[63] ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A. Freedman* 335–421. IMS, Beachwood, OH.

[64] ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

[65] ROBINS, J. M. and ROTNITZKY, A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer". *Statist. Sinica* **11** 920–936.

[66] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[67] ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

[68] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[69] RUBIN, D. and VAN DER LAAN, M. J. (2005). A general imputation methodology for nonparametric regression with censored data.

[70] RUBIN, D. and VAN DER LAAN, M. J. (2007). A doubly robust censoring unbiased transformation. *Int. J. Biostat.* **3** 4. MR2306842 https://doi.org/10.2202/1557-4679.1052

[71] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Rejoinder-adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1135–1146.

[72] SEMENOVA, V. and CHERNOZHUKOV, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econom. J.* **24** 264–289. MR4281225 https://doi.org/10.1093/ectj/utaa027

[73] SREBRO, N., SRIDHARAN, K. and TEWARI, A. (2010). Smoothness, low noise and fast rates. In *Advances in Neural Information Processing Systems* 2199–2207.

[74] SWAMINATHAN, A. and JOACHIMS, T. (2015). Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning* 814–823.

[75] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. *Springer Series in Statistics*. Springer, New York. MR2233926

[76] VAN DER LAAN, M. J. and DUDOIT, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples.

[77] VAN DER LAAN, M. J., DUDOIT, S. and VAN DER VAART, A. W. (2006). The cross-validated adaptive epsilon-net estimator. *Statist. Decisions* **24** 373–395. MR2305113 https://doi.org/10.1524/stnd.2006.24.3.373

[78] VAN DER LAAN, M. J. and LUEDTKE, A. R. (2014). Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome.

[79] VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** 25. MR2349918 https://doi.org/10.2202/1544-6115.1309

[80] VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, Berlin.

[81] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning*: *Causal Inference for Observational and Experimental Data*. *Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

[82] VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** 11. MR2306500 https://doi.org/10.2202/1557-4679.1043

[83] VAN DER VAART, A. (2000). *Asymptotic Statistics* **3**. Cambridge Univ. Press, Cambridge.

[84] VAN DER VAART, A. and VAN DER LAAN, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2** 9. MR2306498 https://doi.org/10.2202/1557-4679.1014

[85] VAPNIK, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York. MR1367965 https://doi.org/10.1007/978-1-4757-2440-0

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

[87] WANG, L., ROTNITZKY, A. and LIN, X. (2010). Nonparametric regression with missing outcomes using weighted kernel estimating equations. *J. Amer. Statist. Assoc.* **105** 1135–1146. MR2752609 https://doi.org/10.1198/jasa.2010.tm08463

[88] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

[89] 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 https://doi.org/10.1080/01621459.2012.695674

[90] ZHENG, W. and VAN DER LAAN, M. J. (2010). Asymptotic theory for cross-validated targeted maximum likelihood estimation.

[91] ZHOU, X., MAYER-HAMBLETT, N., KHAN, U. and KOSOROK, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *J. Amer. Statist. Assoc.* **112** 169–187. MR3646564 https://doi.org/10.1080/01621459.2015.1093947

[92] ZHOU, Z., ATHEY, S. and WAGER, S. (2023). Offline multi-action policy learning: Generalization and optimization. *Oper. Res.* **71** 148–183. MR4560193 https://doi.org/10.1287/opre.2022.2271

OPTIMAL HIGH-DIMENSIONAL AND NONPARAMETRIC DISTRIBUTED TESTING UNDER COMMUNICATION CONSTRAINTS

BY BOTOND SZABÓ[1,a], LASSE VUURSTEEN[2,b] AND HARRY VAN ZANTEN[3,c]

[1]*Department of Decision Sciences, Bocconi Institute for Data Science and Analytics (BIDSA), Bocconi University,*
[a]*botond.szabo@unibocconi.it*

[2]*Delft Institute of Applied Mathematics (DIAM), Delft University of Technology,* [b]*l.vuursteen@tudelft.nl*

[3]*Department of Mathematics, Vrije Universiteit Amsterdam,* [c]*j.h.van.zanten@vu.nl*

We derive minimax testing errors in a distributed framework where the data is split over multiple machines and their communication to a central machine is limited to *b* bits. We investigate both the *d*- and infinite-dimensional signal detection problem under Gaussian white noise. We also derive distributed testing algorithms reaching the theoretical lower bounds.

Our results show that distributed testing is subject to fundamentally different phenomena that are not observed in distributed estimation. Among our findings we show that testing protocols that have access to shared randomness can perform strictly better in some regimes than those that do not. We also observe that consistent nonparametric distributed testing is always possible, even with as little as *one* bit of communication, and the corresponding test outperforms the best local test using only the information available at a single local machine. Furthermore, we also derive adaptive nonparametric distributed testing strategies and the corresponding theoretical lower bounds.

REFERENCES

[1] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference under information constraints I: Lower bounds from chi-square contraction. *IEEE Trans. Inf. Theory* **66** 7835–7855. MR4191030 https://doi.org/10.1109/TIT.2020.3028440

[2] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference under information constraints II: Communication constraints and shared randomness. *IEEE Trans. Inf. Theory* **66** 7856–7877. MR4191031 https://doi.org/10.1109/TIT.2020.3028439

[3] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Distributed signal detection under communication constraints. In *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 41–63. PMLR.

[4] AHLSWEDE, R. and CSISZÁR, I. (1986). Hypothesis testing with communication constraints. *IEEE Trans. Inf. Theory* **32** 533–542. MR0852123 https://doi.org/10.1109/TIT.1986.1057194

[5] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[6] BALAKRISHNAN, S. and WASSERMAN, L. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. MR3953439 https://doi.org/10.1214/18-AOS1729

[7] BARNES, L. P., HAN, Y. and ÖZGÜR, A. (2020). Lower bounds for learning distributions under communication constraints via Fisher information. *J. Mach. Learn. Res.* **21** Paper No. 236, 30. MR4209522

[8] BENNETT, J., CARBERY, A., CHRIST, M. and TAO, T. (2008). The Brascamp–Lieb inequalities: Finiteness, structure and extremals. *Geom. Funct. Anal.* **17** 1343–1415. MR2377493 https://doi.org/10.1007/s00039-007-0619-6

[9] BERGER, T. and ZHANG, Z. (1994). On the CEO problem. In *Proceedings of* 1994 *IEEE International Symposium on Information Theory* 201. https://doi.org/10.1109/ISIT.1994.394767

[10] BRASCAMP, H. J. and LIEB, E. H. (1976). Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Adv. Math.* **20** 151–173. MR0412366 https://doi.org/10.1016/0001-8708(76)90184-5

[11] BRAVERMAN, M., GARG, A., MA, T., NGUYEN, H. L. and WOODRUFF, D. P. (2016). Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 1011–1020. ACM, New York. MR3536632 https://doi.org/10.1145/2897518.2897582

[12] CAI, T. T. and WEI, H. (2020). Distributed Gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms. cs, math, stat, available at arXiv:2001.08877.

[13] CAI, T. T. and WEI, H. (2022). Distributed adaptive Gaussian mean estimation with unknown variance: interactive protocol helps adaptation. arXiv preprint. Available at arXiv:2001.08877.

[14] CAI, T. T. and WEI, H. (2022). Distributed nonparametric function estimation: Optimal rate of convergence and cost of adaptation. *Ann. Statist.* **50** 698–725. MR4404917 https://doi.org/10.1214/21-aos2124

[15] CARLEN, E. A. and CORDERO-ERAUSQUIN, D. (2009). Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geom. Funct. Anal.* **19** 373–405. MR2545242 https://doi.org/10.1007/s00039-009-0001-y

[16] DUCHI, J. C., JORDAN, M. I., WAINWRIGHT, M. J. and ZHANG, Y. (2014). Optimality guarantees for distributed statistical estimation. cs, math, stat, available at arXiv:1405.0782.

[17] HAN, T. S. and AMARI, S. (1998). Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory* **44** 2300–2324. MR1658791 https://doi.org/10.1109/18.720540

[18] HAN, Y., ÖZGÜR, A. and WEISSMAN, T. (2018). Geometric lower bounds for distributed parameter estimation under communication constraints. In *Conference on Learning Theory* 3163–3188. PMLR.

[19] 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 https://doi.org/10.1007/978-0-387-21580-8

[20] KREIDL, O. P., TSITSIKLIS, J. N. and ZOUMPOULIS, S. I. (2011). On decentralized detection with partial information sharing among sensors. *IEEE Trans. Signal Process.* **59** 1759–1765. https://doi.org/10.1109/TSP.2010.2099223

[21] LIEB, E. H. (1990). Gaussian kernels have only Gaussian maximizers. *Invent. Math.* **102** 179–208. MR1069246 https://doi.org/10.1007/BF01233426

[22] LIU, J., COURTADE, T. A., CUFF, P. W. and VERDÚ, S. (2016). Brascamp–Lieb inequality and its reverse: An information theoretic view. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 1048–1052. IEEE, Barcelona, Spain. https://doi.org/10.1109/ISIT.2016.7541459

[23] RAO, A. and YEHUDAYOFF, A. (2020). *Communication Complexity*: *And Applications*. Cambridge Univ. Press, Cambridge.

[24] SHAMIR, O. (2014). Fundamental limits of online and distributed algorithms for statistical learning and estimation. *Adv. Neural Inf. Process. Syst.* **27** 163–171.

[25] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962 https://doi.org/10.1214/aos/1032181163

[26] SZABÓ, B. and VAN ZANTEN, H. (2020). Adaptive distributed methods under communication constraints. *Ann. Statist.* **48** 2347–2380. MR4134798 https://doi.org/10.1214/19-AOS1890

[27] SZABÓ, B. and VAN ZANTEN, H. (2022). Distributed function estimation: adaptation using minimal communication. arXiv preprint. Available at arXiv:2003.12838.

[28] SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2022). Optimal distributed composite testing in high-dimensional Gaussian models with 1-bit communication. *IEEE Trans. Inf. Theory* **68** 4070–4084. MR4433269

[29] SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2023). Supplement to "Optimal high-dimensional and nonparametric distributed testing under communication constraints." https://doi.org/10.1214/23-AOS2269SUPP

[30] SZABÓ, B. and ZAMAN, A. (2022). Distributed nonparametric estimation under communication constraints. Preprint.

[31] TARIGHATI, A., GROSS, J. and JALDÉN, J. (2017). Decentralized hypothesis testing in energy harvesting wireless sensor networks. *IEEE Trans. Signal Process.* **65** 4862–4873. MR3689096 https://doi.org/10.1109/TSP.2017.2716909

[32] TENNEY, R. R. and SANDELL, N. R. JR. (1981). Detection with distributed sensors. *IEEE Trans. Aerosp. Electron. Syst.* **17** 501–510. MR0624526 https://doi.org/10.1109/TAES.1981.309178

[33] TSITSIKLIS, J. N. (1988). Decentralized detection by a large number of sensors. *Math. Control Signals Systems* **1** 167–182. MR0936332 https://doi.org/10.1007/BF02551407

[34] VERSHYNIN, R. (2018). *High-Dimensional Probability*: *An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

[35] XU, A. and RAGINSKY, M. (2017). Information-theoretic lower bounds on Bayes risk in decentralized estimation. *IEEE Trans. Inf. Theory* **63** 1580–1600. MR3625982 https://doi.org/10.1109/TIT.2016.2646342

[36] YAO, A. C.-C. (1979). Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*. STOC'79 209–213. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/800135.804414

[37] ZHANG, Y., DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *NIPS* 2328–2336. Citeseer.

[38] ZHU, Y. and LAFFERTY, J. (2018). Distributed nonparametric regression under communication constraints. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 6009–6017. PMLR.

OPTIMAL PERMUTATION ESTIMATION IN CROWDSOURCING PROBLEMS

BY EMMANUEL PILLIAT[1,a], ALEXANDRA CARPENTIER[2,b] AND NICOLAS VERZELEN[3,c]

[1]*IMAG, Université Montpellier,* [a]*emmanuel.pilliat@umontpellier.fr*
[2]*Institut für Mathematik, Universität Potsdam,* [b]*carpentier@uni-potsdam.de*
[3]*MISTEA, INRAE, Université Montpellier,* [c]*nicolas.verzelen@inrae.fr*

Motivated by crowdsourcing applications, we consider a model where we have partial observations from a bivariate isotonic $n \times d$ matrix with an unknown permutation π^* acting on its rows. Focusing on the twin problems of recovering the permutation π^* and estimating the unknown matrix, we introduce a polynomial-time procedure achieving the minimax risk for these two problems, this for all possible values of n, d, and all possible sampling efforts. Along the way we establish that, in some regimes, recovering the unknown permutation π^* is considerably simpler than estimating the matrix.

REFERENCES

[1] BALLINGER, T. P. and WILCOX, N. T. (1997). Decisions, error and heterogeneity. *Econ. J.* **107** 1090–1105.
[2] 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 https://doi.org/10.2307/2334029
[3] BRAVERMAN, M. and MOSSEL, E. (2008). Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 268–276. ACM, New York. MR2485312
[4] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272
[5] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865
[6] CHATTERJEE, S. and MUKHERJEE, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans. Inf. Theory* **65** 3525–3539. MR3959003 https://doi.org/10.1109/TIT.2019.2893911
[7] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Optimal full ranking from pairwise comparisons. *Ann. Statist.* **50** 1775–1805. MR4441140 https://doi.org/10.1214/22-aos2175
[8] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745
[9] DAWID, A. P. and SKENE, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **28** 20–28.
[10] FLAMMARION, N., MAO, C. and RIGOLLET, P. (2019). Optimal rates of statistical seriation. *Bernoulli* **25** 623–653. MR3892331 https://doi.org/10.3150/17-bej1000
[11] KYNG, R., RAO, A. and SACHDEVA, S. (2015). Fast, provable algorithms for isotonic regression in all l_p-norms. *Adv. Neural Inf. Process. Syst.* **28**.
[12] LIU, A. and MOITRA, A. (2020). Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation. In *Conference on Learning Theory* 2780–2829. PMLR.
[13] LIU, H., GAO, C. and SAMWORTH, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *Ann. Statist.* **49** 1081–1112. MR4255120 https://doi.org/10.1214/20-aos1994
[14] LUCE, R. D. (2012). *Individual choice behavior: A theoretical analysis.* Courier Corporation.
[15] MAO, C., PANANJADY, A. and WAINWRIGHT, M. J. (2020). Towards optimal estimation of bivariate isotonic matrices with unknown permutations. *Ann. Statist.* **48** 3183–3205. MR4185805 https://doi.org/10.1214/19-AOS1925
[16] MAO, C., WEED, J. and RIGOLLET, P. (2018). Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 27. Proceedings of Machine Learning Research PMLR. MR3857331

[17] McLAUGHLIN, D. H. and LUCE, R. D. (1965). Stochastic transitivity and cancellation of preferences between bitter-sweet solutions. *Psychonomic Science* **2** 89–90.

[18] PANANJADY, A., MAO, C., MUTHUKUMAR, V., WAINWRIGHT, M. J. and COURTADE, T. A. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.* **48** 1072–1097. MR4102688 https://doi.org/10.1214/19-AOS1838

[19] PANANJADY, A. and SAMWORTH, R. J. (2022). Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Ann. Statist.* **50** 324–350. MR4382019 https://doi.org/10.1214/21-aos2107

[20] PILLIAT, E., CARPENTIER, A. and VERZELEN, N. (2023). Supplement to "Optimal Permutation Estimation in CrowdSourcing problems." https://doi.org/10.1214/23-AOS2271SUPP

[21] SHAH, N., BALAKRISHNAN, S., BRADLEY, J., PAREKH, A., RAMCHANDRAN, K. and WAINWRIGHT, M. (2015). Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. In *Artificial intelligence and statistics* 856–865. PMLR.

[22] SHAH, N. B., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. J. (2017). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inf. Theory* **63** 934–959. MR3604649 https://doi.org/10.1109/TIT.2016.2634418

[23] SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2019). Feeling the bern: Adaptive estimators for Bernoulli probabilities of pairwise comparisons. *IEEE Trans. Inf. Theory* **65** 4854–5874. MR3988527 https://doi.org/10.1109/TIT.2019.2903249

[24] SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2021). A permutation-based model for crowd labeling: Optimal estimation and robustness. *IEEE Trans. Inf. Theory* **67** 4162–4184. MR4289370 https://doi.org/10.1109/tit.2020.3045613

[25] THURSTONE, L. (1927). A law of comparative judgment. *Psychol. Rev.* **34**.

[26] VIGNA, S. (2016). Spectral ranking. *Netw. Sci.* **4** 433–445.

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

[28] ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

TOTAL POSITIVITY IN MULTIVARIATE EXTREMES

BY FRANK RÖTTGER[1,a], SEBASTIAN ENGELKE[1,b] AND PIOTR ZWIERNIK[2,c]

[1]*Research Center for Statistics, Université de Genève,* [a]*frank.roettger@unige.ch,* [b]*sebastian.engelke@unige.ch*

[2]*Department of Statistical Sciences, University of Toronto,* [c]*piotr.zwiernik@utoronto.ca*

Positive dependence is present in many real world data sets and has appealing stochastic properties that can be exploited in statistical modeling and in estimation. In particular, the notion of multivariate total positivity of order 2 (MTP$_2$) is a convex constraint and acts as an implicit regularizer in the Gaussian case. We study positive dependence in multivariate extremes and introduce EMTP$_2$, an extremal version of MTP$_2$. This notion turns out to appear prominently in extremes, and in fact, it is satisfied by many classical models. For a Hüsler–Reiss distribution, the analogue of a Gaussian distribution in extremes, we show that it is EMTP$_2$ if and only if its precision matrix is a Laplacian of a connected graph. We propose an estimator for the parameters of the Hüsler–Reiss distribution under EMTP$_2$ as the solution of a convex optimization problem with Laplacian constraint. We prove that this estimator is consistent and typically yields a sparse model with possibly nondecomposable extremal graphical structure. Applying our methods to a data set of Danube River flows, we illustrate this regularization and the superior performance compared to existing methods.

REFERENCES

AGRAWAL, R., ROY, U. and UHLER, C. (2020). Covariance matrix estimation under total positivity for portfolio selection. *J. Financ. Econom.*. https://doi.org/10.1093/jjfinec/nbaa018

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

ASENOVA, S., MAZO, G. and SEGERS, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes* **24** 461–500. MR4301085 https://doi.org/10.1007/s10687-021-00407-5

BANERJEE, O., EL GHAOUI, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243

BERGE, C. (1997). *Topological Spaces*: *Including a Treatment of Multi-Valued Functions*, *Vector Spaces and Convexity*. Dover, Mineola, NY. Translated from the French original by E. M. Patterson, Reprint of the 1963 translation. MR1464690

COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

COOLEY, D., NAVEAU, P. and PONCET, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics*. *Lect. Notes Stat.* **187** 373–390. Springer, New York. MR2283264 https://doi.org/10.1007/0-387-36062-X_17

COOLEY, D. and THIBAUD, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika* **106** 587–604. MR3992391 https://doi.org/10.1093/biomet/asz028

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290 https://doi.org/10.1007/BF00533086

DELLACHERIE, C., MARTINEZ, S. and SAN MARTIN, J. (2014). *Inverse M-Matrices and Ultrametric Matrices*. *Lecture Notes in Math*. **2118**. Springer, Cham. MR3289211 https://doi.org/10.1007/978-3-319-10298-6

DEVRIENDT, K. (2022). Effective resistance is more than distance: Laplacians, simplices and the Schur complement. *Linear Algebra Appl*. **639** 24–49. MR4364997 https://doi.org/10.1016/j.laa.2022.01.002

DEZA, M. M. and LAURENT, M. (1997). *Geometry of Cuts and Metrics. Algorithms and Combinatorics* **15**. Springer, Berlin. MR1460488 https://doi.org/10.1007/978-3-642-04295-9

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103** 303–317. MR3509888 https://doi.org/10.1093/biomet/asw008

DOMBRY, C. and EYI-MINKO, F. (2013). Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab*. **18** no. 7, 21. MR3024101 https://doi.org/10.1214/EJP.v18-1991

DREES, H. and SABOURIN, A. (2021). Principal component analysis for multivariate extremes. *Electron. J. Stat*. **15** 908–943. MR4255291 https://doi.org/10.1214/21-ejs1803

DUVAL, A. M., KLIVANS, C. J. and MARTIN, J. L. (2009). Simplicial matrix-tree theorems. *Trans. Amer. Math. Soc*. **361** 6073–6114. MR2529925 https://doi.org/10.1090/S0002-9947-09-04898-3

EGILMEZ, H. E., PAVEZ, E. and ORTEGA, A. (2017). Graph learning from data under Laplacian and structural constraints. *IEEE J. Sel. Top. Signal Process*. **11** 825–841. https://doi.org/10.1109/JSTSP.2017.2726975

EINMAHL, J. H. J., KIRILIOUK, A., KRAJINA, A. and SEGERS, J. (2016). An *M*-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **78** 275–298. MR3453656 https://doi.org/10.1111/rssb.12114

EINMAHL, J. H. J. and SEGERS, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist*. **37** 2953–2989. MR2541452 https://doi.org/10.1214/08-AOS677

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **82** 871–932. With discussions. MR4136498

ENGELKE, S., HITZ, A. S., GNECCO, N. and HENTSCHEL, M. (2022). graphicalExtremes: Statistical methodology for graphical extreme value models. R package version 0.2.0.

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl*. **8** 241–270. MR4243547 https://doi.org/10.1146/annurev-statistics-040620-041554

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 239–265. MR3299407 https://doi.org/10.1111/rssb.12074

ENGELKE, S., OPITZ, T. and WADSWORTH, J. (2019). Extremal dependence of random scale constructions. *Extremes* **22** 623–666. MR4031852 https://doi.org/10.1007/s10687-019-00353-3

ENGELKE, S. and VOLGUSHEV, S. (2022). Structure learning for extremal tree models. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **84** 2055–2087. MR4515566

ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat*. **38** 1466–1474. MR0217826 https://doi.org/10.1214/aoms/1177698701

FALLAT, S., LAURITZEN, S., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Ann. Statist*. **45** 1152–1184. MR3662451 https://doi.org/10.1214/16-AOS1478

FELSENSTEIN, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet*. **25** 471–492.

FIEDLER, M. (1998). Some characterizations of symmetric inverse *M*-matrices. In *Proceedings of the Sixth Conference of the International Linear Algebra Society* (*Chemnitz*, 1996) **275/276** 179–187. Linear Algebra Appl. MR1628388 https://doi.org/10.1016/S0024-3795(97)10022-2

FOMICHOV, V. and IVANOVS, J. (2023). Spherical clustering in detection of groups of concomitant extremes. *Biometrika* **110** 135–153. MR4565448 https://doi.org/10.1093/biomet/asac020

FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys*. **22** 89–103. MR0309498

GNECCO, N., MEINSHAUSEN, N., PETERS, J. and ENGELKE, S. (2021). Causal discovery in heavy-tailed models. *Ann. Statist*. **49** 1755–1778. MR4298880 https://doi.org/10.1214/20-aos2021

GOWER, J. C. (1985). Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Appl*. **67** 81–97. MR0787367 https://doi.org/10.1016/0024-3795(85)90187-9

HASHORVA, E. (2012). Exact tail asymptotics in bivariate scale mixture models. *Extremes* **15** 109–128. MR2891312 https://doi.org/10.1007/s10687-011-0129-7

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **66** 497–546. With discussions and reply by the authors. MR2088289 https://doi.org/10.1111/j.1467-9868.2004.02050.x

HENTSCHEL, M., ENGELKE, S. and SEGERS, J. (2022). Statistical inference for Hüsler–Reiss graphical models through matrix completions. arXiv preprint arXiv:2210.14292.

HOLBROOK, A. (2018). Differentiating the pseudo determinant. *Linear Algebra Appl.* **548** 293–304. MR3783061 https://doi.org/10.1016/j.laa.2018.03.018

HOLLAND, P. W. and ROSENBAUM, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14** 1523–1543. MR0868316 https://doi.org/10.1214/aos/1176350174

HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

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. MR0599685 https://doi.org/10.1016/0047-259X(80)90065-2

KLEIN, D. J. and RANDIĆ, M. (1993). Resistance distance. *J. Math. Chem.* **12** 81–95. MR1219566 https://doi.org/10.1007/BF01164627

KRIJNEN, W. P. (2004). Positive loadings and factor correlations from positive covariance matrices. *Psychometrika* **69** 655–660. MR2272470 https://doi.org/10.1007/BF02289861

KRUPSKII, P., HUSER, R. and GENTON, M. G. (2018). Factor copula models for replicated spatial data. *J. Amer. Statist. Assoc.* **113** 467–479. MR3803479 https://doi.org/10.1080/01621459.2016.1261712

LARSSON, M. and RESNICK, S. I. (2012). Extremal dependence measure and extremogram: The regularly varying case. *Extremes* **15** 231–256. MR2915582 https://doi.org/10.1007/s10687-011-0135-9

LAURITZEN, S. and SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46** 2251–2278. MR3845017 https://doi.org/10.1214/17-AOS1618

LAURITZEN, S., UHLER, C. and ZWIERNIK, P. (2019). Maximum likelihood estimation in Gaussian models under total positivity. *Ann. Statist.* **47** 1835–1863. MR3953437 https://doi.org/10.1214/17-AOS1668

LAURITZEN, S., UHLER, C. and ZWIERNIK, P. (2021). Total positivity in exponential families with application to binary variables. *Ann. Statist.* **49** 1436–1459. MR4298870 https://doi.org/10.1214/20-aos2007

LAURITZEN, S. and ZWIERNIK, P. (2020). GOLAZO: Flexible regularised likelihood estimation using the GO-LAZO approach. Available from https://github.com/pzwiernik/golazo.

LAURITZEN, S. and ZWIERNIK, P. (2022). Locally associated graphical models and mixed convex exponential families. *Ann. Statist.* **50** 3009–3038. MR4505375 https://doi.org/10.1214/22-aos2219

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

LAWLEY, D. N. and MAXWELL, A. E. (1962). Factor analysis as a statistical method. *J. R. Stat. Soc., Ser. D, Stat.* **12** 209–229.

LAYER, M. and RHODES, J. A. (2017). Phylogenetic trees and Euclidean embeddings. *J. Math. Biol.* **74** 99–111. MR3590677 https://doi.org/10.1007/s00285-016-1018-0

LIU, Y. and KOZUBOWSKI, T. J. (2015). A folded Laplace distribution. *J. Stat. Distrib. Appl.* **2** 1–17.

MARSHALL, A. W. and OLKIN, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11** 168–177. MR0682807

MHALLA, L., CHAVEZ-DEMOULIN, V. and DUPUIS, D. J. (2020). Causal mechanism of extreme river discharges in the upper Danube basin network. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 741–764. MR4133145

MICCHELLI, C. A. (1986). Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx.* **2** 11–22. MR0891767 https://doi.org/10.1007/BF01893414

MUROTA, K. (2009). Recent developments in discrete convex analysis. In *Research Trends in Combinatorial Optimization* 219–260. Springer, Berlin. MR2513318 https://doi.org/10.1007/978-3-540-76796-1_11

NEWMAN, C. M. (1983). A general central limit theorem for FKG systems. *Comm. Math. Phys.* **91** 75–80. MR0719811

NEWMAN, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In *Inequalities in Statistics and Probability (Lincoln, Neb., 1982). Institute of Mathematical Statistics Lecture Notes—Monograph Series* **5** 127–140. IMS, Hayward, CA. MR0789244 https://doi.org/10.1214/lnms/1215465639

PAPASTATHOPOULOS, I. and STROKORB, K. (2016). Conditional independence among max-stable laws. *Statist. Probab. Lett.* **108** 9–15. MR3420420 https://doi.org/10.1016/j.spl.2015.08.008

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

PITT, L. D. (1982). Positively correlated normal variables are associated. *Ann. Probab.* **10** 496–499. MR0665603

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766 https://doi.org/10.1214/11-EJS631

RESNICK, S. I. (2008). *Extreme Values, Regular Variation and Point Processes. Springer Series in Operations Research and Financial Engineering*. Springer, New York. Reprint of the 1987 original. MR2364939

ROBEVA, E., STURMFELS, B., TRAN, N. and UHLER, C. (2021). Maximum likelihood estimation for totally positive log-concave densities. *Scand. J. Stat.* **48** 817–844. MR4303559 https://doi.org/10.1111/sjos.12462

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668 https://doi.org/10.3150/bj/1161614952

ROSSELL, D. and ZWIERNIK, P. (2021). Dependence in elliptical partial correlation graphs. *Electron. J. Stat.* **15** 4236–4263. MR4312203 https://doi.org/10.1214/21-ejs1891

RÖTTGER, F., ENGELKE, S. and ZWIERNIK, P. (2023). Supplement to "Total positivity in multivariate extremes." https://doi.org/10.1214/23-AOS2272SUPP

SCHLATHER, M. and TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90** 139–156. MR1966556 https://doi.org/10.1093/biomet/90.1.139

SCHOENBERG, I. J. (1935). Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert" [MR1503246]. *Ann. of Math.* (2) **36** 724–732. MR1503248 https://doi.org/10.2307/1968654

SEGERS, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Adv. in Appl. Probab.* **52** 855–878. MR4153591 https://doi.org/10.1017/apr.2020.22

SEMPLE, C. and STEEL, M. (2003). *Phylogenetics. Oxford Lecture Series in Mathematics and Its Applications* **24**. Oxford Univ. Press, Oxford. MR2060009

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. MR3338330 https://doi.org/10.1016/j.laa.2014.04.020

SMITH, R. L., TAWN, J. A. and COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84** 249–268. MR1467045 https://doi.org/10.1093/biomet/84.2.249

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson, A Bradford Book. MR1815675

STEEL, M. (2016). *Phylogeny—Discrete and Random Processes in Evolution. CBMS-NSF Regional Conference Series in Applied Mathematics* **89**. SIAM, Philadelphia, PA. MR3601108 https://doi.org/10.1137/1.9781611974485.ch1

STELLATO, B., BANJAC, G., GOULART, P., BEMPORAD, A. and BOYD, S. (2020). OSQP: An operator splitting solver for quadratic programs. *Math. Program. Comput.* **12** 637–672. MR4163557 https://doi.org/10.1007/s12532-020-00179-2

STURMFELS, B., UHLER, C. and ZWIERNIK, P. (2020). Brownian motion tree models are toric. *Kybernetika* (*Prague*) **56** 1154–1175. MR4199908 https://doi.org/10.14736/kyb-2020-6-1154

TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77** 245–253.

TRAN, N. M., BUCK, J. and KLÜPPELBERG, C. (2021). Estimating a latent tree for extremes. arXiv preprint arXiv:2102.06197.

VINICIUS, Z. and PALOMAR, D. (2019). spectralGraphTopology: Learning graphs from data via spectral constraints. Available from https://cran.r-project.org/package=spectralGraphTopology.

VINICIUS, Z., YING, J. and PALOMAR, D. (2021). sparseGraph: Estimating graphs with nonconvex, sparse promoting regularizations. Available from https://github.com/mirca/sparseGraph/.

WADSWORTH, J. L. and TAWN, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika* **99** 253–272. MR2931252 https://doi.org/10.1093/biomet/asr080

WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

WANG, Y., ROY, U. and UHLER, C. (2020). Learning high-dimensional Gaussian graphical models under total positivity without adjustment of tuning parameters. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 2698–2708. PMLR.

YING, J., CARDOSO, J. M. and PALOMAR, D. (2021). Minimax estimation of Laplacian constrained precision matrices. In *International Conference on Artificial Intelligence and Statistics* 3736–3744. PMLR.

ZHAO, L., WANG, Y., KUMAR, S. and PALOMAR, D. P. (2019). Optimization algorithms for graph Laplacian estimation via ADMM and MM. *IEEE Trans. Signal Process.* **67** 4231–4244. MR3999755 https://doi.org/10.1109/TSP.2019.2925602

ZWIERNIK, P. (2019). Latent tree models. In *Handbook of Graphical Models. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 265–288. CRC Press, Boca Raton, FL. MR3888000

A POWER ANALYSIS FOR MODEL-X KNOCKOFFS WITH ℓ_p-REGULARIZED STATISTICS

BY ASAF WEINSTEIN[1,a], WEIJIE J. SU[2,b], MAŁGORZATA BOGDAN[3,c], RINA FOYGEL BARBER[4,d] AND EMMANUEL J. CANDÈS[5,e]

[1]*Department of Statistics and Data Science, Hebrew University of Jerusalem,* [a]*asaf.weinstein@mail.huji.ac.il*

[2]*Department of Statistics and Data Science, University of Pennsylvania,* [b]*suw@wharton.upenn.edu*

[3]*Department of Mathematics, University of Wroclaw,* [c]*malgorzata.bogdan@uwr.edu.pl*

[4]*Department of Statistics, University of Chicago,* [d]*rina@uchicago.edu*

[5]*Department of Statistics and Department of Mathematics, Stanford University,* [e]*candes@stanford.edu*

Variable selection properties of procedures utilizing penalized-likelihood estimates is a central topic in the study of high-dimensional linear regression problems. Existing literature emphasizes the quality of ranking of the variables by such procedures as reflected in the receiver operating characteristic curve or in prediction performance. Specifically, recent works have harnessed modern theory of approximate message-passing (AMP) to obtain, in a particular setting, exact asymptotic predictions of the type I, type II error tradeoff for selection procedures that rely on ℓ_p-regularized estimators.

In practice, effective ranking by itself is often not sufficient because some calibration for Type I error is required. In this work, we study theoretically the power of selection procedures that similarly rank the features by the size of an ℓ_p-regularized estimator, but further use Model-X knockoffs to control the false discovery rate in the realistic situation where no prior information about the signal is available. In analyzing the power of the resulting procedure, we extend existing results in AMP theory to handle the pairing between original variables and their knockoffs. This is used to derive exact asymptotic predictions for power. We apply the general results to compare the power of the knockoffs versions of Lasso and thresholded-Lasso selection, and demonstrate that in the i.i.d. covariate setting under consideration, tuning by cross-validation on the augmented design matrix is nearly optimal. We further demonstrate how the techniques allow to analyze also the Type S error, and a corresponding notion of power, when selections are supplemented with a decision on the sign of the coefficient.

REFERENCES

[1] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

[2] BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* **47** 2504–2537. MR3988764 https://doi.org/10.1214/18-AOS1755

[3] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

[4] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inf. Theory* **58** 1997–2017. MR2951312 https://doi.org/10.1109/TIT.2011.2174612

[5] BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438 https://doi.org/10.1093/biomet/93.3.491

[6] BLANCHARD, G. and ROQUAIN, É. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2871. MR2579914

[7] 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 https://doi.org/10.1214/15-AOAS842

[8] BOGDAN, M., VAN DEN BERG, E., SU, W. J. and CANDÈS, E. J. (2013). Statistical estimation and testing via the sorted ℓ_1 norm. ArXiv preprint. Available at arXiv:1310.1969.

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

[10] CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

[11] DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. MR2546388 https://doi.org/10.1098/rsta.2009.0152

[12] DOSSAL, C. (2012). A necessary and sufficient condition for exact sparse recovery by ℓ_1 minimization. *C. R. Math. Acad. Sci. Paris* **350** 117–120. MR2887848 https://doi.org/10.1016/j.crma.2011.12.014

[13] GELMAN, A. and TUERLINCKX, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Comput. Statist.* **15** 373–390.

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

[15] SU, W., BOGDAN, M. and CANDÈS, E. (2017). False discoveries occur early on the Lasso path. *Ann. Statist.* **45** 2133–2150. MR3718164 https://doi.org/10.1214/16-AOS1521

[16] TARDIVEL, P. J. C. and BOGDAN, M. (2022). On the sign recovery by least absolute shrinkage and selection operator, thresholded least absolute shrinkage and selection operator, and thresholded basis pursuit denoising. *Scand. J. Stat.* **49** 1636–1668. MR4544815

[17] TUKEY, J. W. (1960). Conclusions vs decisions. *Technometrics* **2** 423–433. MR0116435 https://doi.org/10.2307/1266451

[18] TUKEY, J. W. (1991). The philosophy of multiple comparisons. *Statist. Sci.* 100–116.

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

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

[21] WANG, S., WENG, H. and MALEKI, A. (2020). Which bridge estimator is the best for variable selection? *Ann. Statist.* **48** 2791–2823. MR4152121 https://doi.org/10.1214/19-AOS1906

[22] WANG, W. and JANSON, L. (2022). A high-dimensional power analysis of the conditional randomization test and knockoffs. *Biometrika* **109** 631–645. MR4472839 https://doi.org/10.1093/biomet/asab052

[23] WEINSTEIN, A., BARBER, R. and CANDÈS, E. J. (2017). A power and prediction analysis for knockoffs with lasso statistics. ArXiv preprint. Available at arXiv:1712.06465.

[24] WEINSTEIN, A., SU, W. J., BOGDAN, M., FOYGEL BARBER, R. and CANDÈS, E. J. (2023). Supplement to "A power analysis for model-X knockoffs with ℓ_p-regularized statistics." https://doi.org/10.1214/23-AOS2274SUPP

[25] WENG, H., MALEKI, A. and ZHENG, L. (2018). Overcoming the limitations of phase transition by higher order analysis of regularization techniques. *Ann. Statist.* **46** 3099–3129. MR3851766 https://doi.org/10.1214/17-AOS1651

[26] YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. MR2756192

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

[28] ZHOU, S. (2009). Thresholding procedures for high dimensional variable selection and statistical estimation. In *Advances in Neural Information Processing Systems* 2304–2312.

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

A GENERAL CHARACTERIZATION OF OPTIMAL TIE-BREAKER DESIGNS

BY HARRISON H. LI[a] AND ART B. OWEN[b]

Department of Statistics, Stanford University, [a]*hli90722@stanford.edu,* [b]*owen@stanford.edu*

Tie-breaker designs trade off a measure of statistical efficiency against a short-term gain from preferentially assigning a binary treatment to subjects with higher values of a running variable x. The efficiency measure can be any continuous function of the expected information matrix in a two-line regression model. The short-term gain is expressed as the covariance between the running variable and the treatment indicator. We investigate how to choose design functions $p(x)$ specifying the probability of treating a subject with running variable x in order to optimize these competing objectives, under external constraints on the number of subjects receiving treatment. Our results include sharp existence and uniqueness guarantees, while accommodating the ethically appealing requirement that $p(x)$ be nondecreasing in x. Under this condition, there is always an optimal treatment probability function $p(x)$ that is constant on the sets $(-\infty, t)$ and (t, ∞) for some threshold t and generally discontinuous at $x = t$. When the running variable distribution is not symmetric or the fraction of subjects receiving the treatment is not $1/2$, our optimal designs improve upon a D-optimality objective without sacrificing short-term gain, compared to a typical three-level tie-breaker design that fixes treatment probabilities at 0, $1/2$ and 1. We illustrate our optimal designs with data from Head Start, an early childhood government intervention program.

REFERENCES

ABDULKADIROGLU, A., ANGRIST, J. D., NARITA, Y. and PATHAK, P. A. (2017). Impact evaluation in matching markets with general tie-breaking. Technical Report, National Bureau of Economic Research.

AIKEN, L. S., WEST, S. G., SCHWALM, D. E., CARROLL, J. L. and HSIUNG, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Eval. Rev.* **22** 207–244.

ANGRIST, J., AUTOR, D. and PALLAIS, A. (2020). Marginal effects of merit aid for low-income students. Technical Report, National Bureau of Economic Research.

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67. MR0655670 https://doi.org/10.1093/biomet/69.1.61

ATKINSON, A. C. (2014). Selecting a biased-coin design. *Statist. Sci.* **29** 144–163. MR3201860 https://doi.org/10.1214/13-STS449

ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS. Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

BANDYOPADHYAY, U. and BISWAS, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika* **88** 409–419. MR1844841 https://doi.org/10.1093/biomet/88.2.409

BISWAS, A. and BHATTACHARYA, R. (2018). A class of covariate-adjusted response-adaptive allocation designs for multitreatment binary response trials. *J. Biopharm. Statist.* **28** 809–823.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

BRENT, R. P. (1973). *Algorithms for Minimization Without Derivatives. Prentice-Hall Series in Automatic Computation*. Prentice-Hall, Inc., Englewood Cliffs, NJ. MR0339493

CAMPBELL, D. T. (1969). Reforms as experiments. *Amer. Psychol.* **24** 409.

CATTANEO, M. D., TITIUNIK, R. and VAZQUEZ-BARE, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. *J. Policy Anal. Manage.* **36** 643–681.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.* **24** 586–602. MR0058932 https://doi.org/10.1214/aoms/1177728915

CLYDE, M. and CHALONER, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *J. Amer. Statist. Assoc.* **91** 1236–1244. MR1424621 https://doi.org/10.2307/2291742

COOK, R. D. and WONG, W. K. (1994). On the equivalence of constrained and compound optimal designs. *J. Amer. Statist. Assoc.* **89** 687–692. MR1294092

DANTZIG, G. B. and WALD, A. (1951). On the fundamental lemma of Neyman and Pearson. *Ann. Math. Stat.* **22** 87–93. MR0039962 https://doi.org/10.1214/aoms/1177729695

DE OLIVEIRA, O. (2018). The implicit function theorem for maps that are only differentiable: An elementary proof. *Real Anal. Exchange* **43** 429–443. MR3942588 https://doi.org/10.14321/realanalexch.43.2.0429

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660 https://doi.org/10.1093/biomet/58.3.403

GELMAN, A. and IMBENS, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *J. Bus. Econom. Statist.* **37** 447–456. MR3968384 https://doi.org/10.1080/07350015.2017.1366909

GOLDBERGER, A. S. (1972). Selection bias in evaluating treatment effects: Some formal illustrations. Technical Report No. Discussion paper 128–72, Institute for Research on Poverty, Univ. Wisconsin–Madison.

GOLDENSHLUGER, A. and ZEEVI, A. (2013). A linear response bandit problem. *Stoch. Syst.* **3** 230–261. MR3353472 https://doi.org/10.1214/11-SSY032

HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials. Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2245329 https://doi.org/10.1002/047005588X

HU, J., ZHU, H. and HU, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *J. Amer. Statist. Assoc.* **110** 357–367. MR3338508 https://doi.org/10.1080/01621459.2014.903846

JACOB, R., ZHU, P. and SOMERS, M. (2012). A practical guide to regression discontinuity. MDRC.

KLUGER, D. M. and OWEN, A. B. (2023). Kernel regression analysis of tie-breaker designs. *Electron. J. Stat.* **17** 243–290. MR4536118 https://doi.org/10.1214/23-ejs2102

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379–398. MR0440812 https://doi.org/10.1080/02331937408842203

LÄUTER, E. (1976). Optimal multipurpose designs for regression models. *Math. Operationsforsch. Statist.* **7** 51–68. MR0413385 https://doi.org/10.1080/02331937608842317

LEE, C. M.-S. (1987). Constrained optimal designs for regression models. *Comm. Statist. Theory Methods* **16** 765–783. MR0886553 https://doi.org/10.1080/03610928708829401

LEE, C. M.-S. (1988). Constrained optimal designs. *J. Statist. Plann. Inference* **18** 377–389. MR0926640 https://doi.org/10.1016/0378-3758(88)90114-0

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LIPSEY, M. W., CORDRAY, D. S. and BERGER, D. E. (1981). Evaluation of a juvenile diversion program: Using multiple lines of evidence. *Eval. Rev.* **5** 283–306.

LUDWIG, J. and MILLER, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Q. J. Econ.* **122** 159–208.

METELKINA, A. and PRONZATO, L. (2017). Information-regret compromise in covariate-adaptive treatment allocation. *Ann. Statist.* **45** 2046–2073. MR3718161 https://doi.org/10.1214/16-AOS1518

MORRISON, T. P. and OWEN, A. B. (2022). Optimality in multivariate tie-breaker designs. Technical Report, Stanford Univ. Available at arXiv:2202.10030.

NEYMAN, J. and PEARSON, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A* **231** 289–337.

OWEN, A. B. and VARIAN, H. (2020). Optimizing the tie-breaker regression discontinuity design. *Electron. J. Stat.* **14** 4004–4027. MR4167610 https://doi.org/10.1214/20-EJS1765

R CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. MR2483911 https://doi.org/10.1214/08-STS269

STIGLER, S. M. (1971). Optimal experimental design for polynomial regression. *J. Amer. Statist. Assoc.* **66** 311–318.

SVERDLOV, O., ROSENBERGER, W. F. and RYEZNIK, Y. (2013). Utility of covariate-adjusted response-adaptive randomization in survival trials. *Stat. Biopharm. Res.* **5** 38–53.

THISTLETHWAITE, D. L. and CAMPBELL, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J. Educ. Psychol.* **51** 309.

TROCHIM, W. M. K. and CAPPELLERI, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Control. Clin. Trials* **13** 190–212.

WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** 123–130. MR0356388

ZHANG, L.-X. and HU, F. (2009). A new family of covariate-adjusted response adaptive designs and their properties. *Appl. Math. J. Chinese Univ. Ser. B* **24** 1–13. MR2486489 https://doi.org/10.1007/s11766-009-0001-6

ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Ann. Statist.* **35** 1166–1182. MR2341702 https://doi.org/10.1214/009053606000001424

COMPLEXITY ANALYSIS OF BAYESIAN LEARNING OF HIGH-DIMENSIONAL DAG MODELS AND THEIR EQUIVALENCE CLASSES

BY QUAN ZHOU[a] AND HYUNWOONG CHANG[b]

Department of Statistics, Texas A&M University, [a]*quan@stat.tamu.edu,* [b]*hwchang@stat.tamu.edu*

Structure learning via MCMC sampling is known to be very challenging because of the enormous search space and the existence of Markov equivalent DAGs. Theoretical results on the mixing behavior are lacking. In this work, we prove the rapid mixing of a random walk Metropolis–Hastings algorithm, which reveals that the complexity of Bayesian learning of sparse equivalence classes grows only polynomially in n and p, under some high-dimensional assumptions. A series of high-dimensional consistency results is obtained, including the strong selection consistency of an empirical Bayes model for structure learning. Our proof is based on two new results. First, we derive a general mixing time bound on finite-state spaces, which can be applied to local MCMC schemes for other model selection problems. Second, we construct high-probability search paths on the space of equivalence classes with node degree constraints by proving a combinatorial property of DAG comparisons. Simulation studies on the proposed MCMC sampler are conducted to illustrate the main theoretical findings.

REFERENCES

[1] AGRAWAL, R. and UHLER, C. (2018). Minimal I-MAP MCMC for scalable structure discovery in causal DAG models. In *International Conference on Machine Learning* 89–98.

[2] ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. MR1439312 https://doi.org/10.1214/aos/1031833662

[3] ARAGAM, B., AMINI, A. and ZHOU, Q. (2019). Globally optimal score-based learning of directed acyclic graphs in high-dimensions. In *Advances in Neural Information Processing Systems* 4450–4462.

[4] BANERJEE, S., CASTILLO, I. and GHOSAL, S. (2021). Bayesian inference in high-dimensional models. arXiv preprint. Available at arXiv:2101.04491.

[5] BANERJEE, S. and GHOSAL, S. (2015). Bayesian structure learning in graphical models. *J. Multivariate Anal.* **136** 147–162. MR3321485 https://doi.org/10.1016/j.jmva.2015.01.015

[6] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758

[7] CAO, X., KHARE, K. and GHOSH, M. (2019). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. *Ann. Statist.* **47** 319–348. MR3909935 https://doi.org/10.1214/18-AOS1689

[8] CASTELLETTI, F., CONSONNI, G., DELLA VEDOVA, M. L. and PELUSO, S. (2018). Learning Markov equivalence classes of directed acyclic graphs: An objective Bayes approach. *Bayesian Anal.* **13** 1235–1260. MR3855370 https://doi.org/10.1214/18-BA1101

[9] CASTELO, R. and KOČKA, T. (2003). On inclusion-driven learning of Bayesian networks. *J. Mach. Learn. Res.* **4** 527–574. MR2072261 https://doi.org/10.1162/153244304773936045

[10] CHICKERING, D. M. (2002). Learning equivalence classes of Bayesian-network structures. *J. Mach. Learn. Res.* **2** 445–498. MR1929415 https://doi.org/10.1162/153244302760200696

[11] CHICKERING, D. M. (2002). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554. MR1991085 https://doi.org/10.1162/153244303321897717

[12] DRTON, M., FOYGEL, R. and SULLIVANT, S. (2011). Global identifiability of linear structural equation models. *Ann. Statist.* **39** 865–886. MR2816341 https://doi.org/10.1214/10-AOS859

[13] DRTON, M. and MAATHUIS, M. H. (2017). Structure learning in graphical modeling. *Annu. Rev. Stat. Appl.* **4** 365–393.

[14] EATON, D. and MURPHY, K. (2007). Bayesian structure learning using dynamic programming and MCMC. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* 101–108.

[15] ELLIS, B. and WONG, W. H. (2008). Learning causal Bayesian network structures from experimental data. *J. Amer. Statist. Assoc.* **103** 778–789. MR2524009 https://doi.org/10.1198/016214508000000193

[16] FRIEDMAN, N. and KOLLER, D. (2003). Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. *Mach. Learn.* **50** 95–125.

[17] GAO, B. and CUI, Y. (2015). Learning directed acyclic graphical structures with genetical genomics data. *Bioinformatics* **31** 3953–3960. https://doi.org/10.1093/bioinformatics/btv513

[18] GAO, C., VAN DER VAART, A. W. and ZHOU, H. H. (2020). A general framework for Bayes structured linear models. *Ann. Statist.* **48** 2848–2878. MR4152123 https://doi.org/10.1214/19-AOS1909

[19] GEIGER, D. and HECKERMAN, D. (2002). Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *Ann. Statist.* **30** 1412–1440. MR1936324 https://doi.org/10.1214/aos/1035844981

[20] GIUDICI, P. and CASTELO, R. (2003). Improving Markov chain Monte Carlo model search for data mining. *Mach. Learn.* **50** 127–158.

[21] GOUDIE, R. J. B. and MUKHERJEE, S. (2016). A Gibbs sampler for learning DAGs. *J. Mach. Learn. Res.* **17** Paper No. 30, 39. MR3491124

[22] GRZEGORCZYK, M. and HUSMEIER, D. (2008). Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Mach. Learn.* **71** 265.

[23] HE, Y., JIA, J. and YU, B. (2013). Reversible MCMC on Markov equivalence classes of sparse directed acyclic graphs. *Ann. Statist.* **41** 1742–1779. MR3127848 https://doi.org/10.1214/13-AOS1125

[24] HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–417. MR1765176 https://doi.org/10.1214/ss/1009212519

[25] JEONG, S. and GHOSAL, S. (2021). Posterior contraction in sparse generalized linear models. *Biometrika* **108** 367–379. MR4259137 https://doi.org/10.1093/biomet/asaa074

[26] JOHNSON, V. E. and ROSSELL, D. (2012). Bayesian model selection in high-dimensional settings. *J. Amer. Statist. Assoc.* **107** 649–660. MR2980074 https://doi.org/10.1080/01621459.2012.682536

[27] KAHALE, N. (1997). A semidefinite bound for mixing rates of Markov chains. *Random Structures Algorithms* **11** 299–313. MR1608821 https://doi.org/10.1002/(SICI)1098-2418(199712)11:4<299::AID-RSA2>3.0.CO;2-U

[28] KALISCH, M. and BÜHLMANN, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.* **8** 613–636.

[29] KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models: Principles and Techniques. Adaptive Computation and Machine Learning.* MIT Press, Cambridge, MA. MR2778120

[30] KUIPERS, J. and MOFFA, G. (2017). Partition MCMC for inference on acyclic digraphs. *J. Amer. Statist. Assoc.* **112** 282–299. MR3646571 https://doi.org/10.1080/01621459.2015.1133426

[31] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459 https://doi.org/10.1214/09-AOS720

[32] LAURITZEN, S. L. (1992). Propagation of probabilities, means, and variances in mixed graphical association models. *J. Amer. Statist. Assoc.* **87** 1098–1108. MR1209568

[33] LEE, K., LEE, J. and LIN, L. (2019). Minimax posterior convergence rates and model selection consistency in high-dimensional DAG models based on sparse Cholesky factors. *Ann. Statist.* **47** 3413–3437. MR4025747 https://doi.org/10.1214/18-AOS1783

[34] LIU, S., SUZUKI, T., RELATOR, R., SESE, J., SUGIYAMA, M. and FUKUMIZU, K. (2017). Support consistency of direct sparse-change learning in Markov networks. *Ann. Statist.* **45** 959–990. MR3662445 https://doi.org/10.1214/16-AOS1470

[35] MAATHUIS, M. H., COLOMBO, D., KALISCH, M. and BÜHLMANN, P. (2010). Predicting causal effects in large-scale systems from observational data. *Nat. Methods* **7** 247–248. https://doi.org/10.1038/nmeth0410-247

[36] MADIGAN, D., YORK, J. and ALLARD, D. (1995). Bayesian graphical models for discrete data. *Int. Stat. Rev.* 215–232.

[37] MADIGAN, D., ANDERSSON, S. A., PERLMAN, M. D. and VOLINSKY, C. T. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Comm. Statist. Theory Methods* **25** 2493–2519.

[38] MARTIN, R., MESS, R. and WALKER, S. G. (2017). Empirical Bayes posterior concentration in sparse high-dimensional linear models. *Bernoulli* **23** 1822–1847. MR3624879 https://doi.org/10.3150/15-BEJ797

[39] MEEK, C. (1997). Graphical models: Selecting causal and statistical models. Ph.D. thesis, Carnegie Mellon Univ., Pittsburgh, PA.

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

[41] MUNTEANU, P. and BENDOU, M. (2001). The EQ framework for learning equivalence classes of Bayesian networks. In *Proceedings* 2001 *IEEE International Conference on Data Mining* 417–424. IEEE, San Jose, CA.

[42] NANDY, P., HAUSER, A. and MAATHUIS, M. H. (2018). High-dimensional consistency in score-based and hybrid structure learning. *Ann. Statist.* **46** 3151–3183. MR3851768 https://doi.org/10.1214/17-AOS1654

[43] NARISETTY, N. N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. MR3210987 https://doi.org/10.1214/14-AOS1207

[44] NIINIMÄKI, T. M., PARVIAINEN, P. and KOIVISTO, M. (2011). Partial order MCMC for structure discovery in Bayesian networks. In *Proceedings of the Twenty-Seventh Conference Conference on Uncertainty in Artificial Intelligence* (*UAI*-11) 557–564. AUAI Press, Barcelona, Spain.

[45] PATI, D., BHATTACHARYA, A., PILLAI, N. S. and DUNSON, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *Ann. Statist.* **42** 1102–1130. MR3210997 https://doi.org/10.1214/14-AOS1215

[46] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*. *The Morgan Kaufmann Series in Representation and Reasoning*. Morgan Kaufmann, San Mateo, CA. MR0965765

[47] PELUSO, S. and CONSONNI, G. (2020). Compatible priors for model selection of high-dimensional Gaussian DAGs. *Electron. J. Stat.* **14** 4110–4132. MR4170698 https://doi.org/10.1214/20-EJS1768

[48] PENA, J. M. (2007). Approximate counting of graphical models via MCMC. In *AISTATS* 355–362.

[49] PERLMAN, M. D. (2001). Graphical model search via essential graphs. In *Algebraic Methods in Statistics and Probability* (*Notre Dame*, *IN*, 2000). *Contemp. Math.* **287** 255–265. Amer. Math. Soc., Providence, RI. MR1873680 https://doi.org/10.1090/conm/287/04790

[50] RASKUTTI, G., YU, B. and WAINWRIGHT, M. J. (2008). Model selection in Gaussian graphical models: High-dimensional consistency of ℓ_1-regularized MLE. *Adv. Neural Inf. Process. Syst.* **21**.

[51] SCUTARI, M., GRAAFLAND, C. E. and GUTIÉRREZ, J. M. (2019). Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. *Internat. J. Approx. Reason.* **115** 235–253. MR4018631 https://doi.org/10.1016/j.ijar.2019.10.003

[52] SINCLAIR, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1** 351–370. MR1211324 https://doi.org/10.1017/S0963548300000390

[53] SOLUS, L., WANG, Y. and UHLER, C. (2021). Consistency guarantees for greedy permutation-based causal inference algorithms. *Biometrika* **108** 795–814. MR4341352 https://doi.org/10.1093/biomet/asaa104

[54] SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR1815675

[55] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. *Information Science and Statistics*. Springer, London. MR3183760

[56] SU, C. and BORSUK, M. E. (2016). Improving structure MCMC for Bayesian networks through Markov blanket resampling. *J. Mach. Learn. Res.* **17** Paper No. 118, 20. MR3543524

[57] SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled lasso. *J. Mach. Learn. Res.* **14** 3385–3418. MR3144466

[58] TALWAR, K. (2019). Computational separations between sampling and optimization. *Adv. Neural Inf. Process. Syst.* **32**.

[59] UHLER, C., RASKUTTI, G., BÜHLMANN, P. and YU, B. (2013). Geometry of the faithfulness assumption in causal inference. *Ann. Statist.* **41** 436–463. MR3099109 https://doi.org/10.1214/12-AOS1080

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

[61] VERMA, T. and PEARL, J. (1991). Equivalence and synthesis of causal models. Technical report, UCLA, Computer Science Department.

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

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

[64] YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

[65] 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 https://doi.org/10.1214/15-AOS1417

[66] ZANELLA, G. (2020). Informed proposals for local MCMC in discrete spaces. *J. Amer. Statist. Assoc.* **115** 852–865. MR4107684 https://doi.org/10.1080/01621459.2019.1585255

[67] ZHOU, Q. and CHANG, H. (2023). Supplement to "Complexity analysis of Bayesian learning of high-dimensional DAG models and their equivalence classes." https://doi.org/10.1214/23-AOS2280SUPP

[68] ZHOU, Q., YANG, J., VATS, D., ROBERTS, G. O. and ROSENTHAL, J. S. (2022). Dimension-free mixing for high-dimensional Bayesian variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1751–1784. MR4515557 https://doi.org/10.1111/rssb.12546

[69] ZHUO, B. and GAO, C. (2021). Mixing time of Metropolis–Hastings for Bayesian community detection. *J. Mach. Learn. Res.* **22** Paper No. 10, 89. MR4253703

OPTIMAL REACH ESTIMATION AND METRIC LEARNING

BY EDDIE AAMARI[1,a], CLÉMENT BERENFELD[2,c] AND CLÉMENT LEVRARD[1,b]

[1]*LPSM–CNRS, Université Paris Cité, Sorbonne Université,* [a]*aamari@lpsm.paris,* [b]*levrard@lpsm.paris*
[2]*Institut für Mathematik, Universität Potsdam,* [c]*berenfeld@uni-potsdam.de*

We study the estimation of the *reach*, an ubiquitous regularity parameter in manifold estimation and geometric data analysis. Given an i.i.d. sample over an unknown d-dimensional $\mathcal{C}^k$-smooth submanifold M of $\mathbb{R}^D$, we provide optimal nonasymptotic bounds for the estimation of its reach. We build upon a formulation of the reach in terms of maximal curvature on one hand and geodesic metric distortion on the other. The derived rates are adaptive, with rates depending on whether the reach of M arises from curvature or from a bottleneck structure. In the process we derive optimal geodesic metric estimation bounds.

REFERENCES

[1] AAMARI, E., AARON, C. and LEVRARD, C. (2021). Minimax boundary estimation and estimation with boundary. Preprint. Available at arXiv:2108.03135.

[2] AAMARI, E., BERENFELD, C. and LEVRARD, C. (2023). Supplement to "Optimal reach estimation and metric learning." https://doi.org/10.1214/23-AOS2281SUPP

[3] AAMARI, E., KIM, J., CHAZAL, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2019). Estimating the reach of a manifold. *Electron. J. Stat.* **13** 1359–1399. MR3938326 https://doi.org/10.1214/19-ejs1551

[4] AAMARI, E. and LEVRARD, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204. MR3909931 https://doi.org/10.1214/18-AOS1685

[5] AIZENBUD, Y. and SOBER, B. (2021). Non-parametric estimation of manifolds from noisy data. Preprint. Available at arXiv:2105.04754.

[6] ARIAS-CASTRO, E. and CHAU, P. A. (2020). Minimax estimation of distances on a surface and minimax manifold learning in the isometric-to-convex setting. Preprint. Available at arXiv:2011.12478.

[7] ARIAS-CASTRO, E. and LE GOUIC, T. (2019). Unconstrained and curvature-constrained shortest-path distances and their approximation. *Discrete Comput. Geom.* **62** 1–28. MR3959920 https://doi.org/10.1007/s00454-019-00060-7

[8] ARIAS-CASTRO, E. and PELLETIER, B. (2013). On the convergence of maximum variance unfolding. *J. Mach. Learn. Res.* **14** 1747–1770. MR3104494

[9] ATTALI, D., LIEUTIER, A. and SALINAS, D. (2013). Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes. *Comput. Geom.* **46** 448–465. MR3003901 https://doi.org/10.1016/j.comgeo.2012.02.009

[10] BALAKRISHNAN, S., RINALDO, A., SHEEHY, D., SINGH, A. and WASSERMAN, L. (2012). Minimax rates for homology inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (N. D. Lawrence and M. Girolami, eds.). *Proceedings of Machine Learning Research* **22** 64–72. PMLR, La Palma, Canary Islands.

[11] BERENFELD, C., HARVEY, J., HOFFMANN, M. and SHANKAR, K. (2022). Estimating the reach of a manifold via its convexity defect function. *Discrete Comput. Geom.* **67** 403–438. MR4376572 https://doi.org/10.1007/s00454-021-00290-8

[12] BOISSONNAT, J.-D. and GHOSH, A. (2014). Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.* **51** 221–267. MR3148657 https://doi.org/10.1007/s00454-013-9557-2

[13] BOISSONNAT, J.-D., LIEUTIER, A. and WINTRAECKEN, M. (2019). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *J. Appl. Comput. Topol.* **3** 29–58. MR3988212 https://doi.org/10.1007/s41468-019-00029-8

[14] BOURGAIN, J. (1985). On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.* **52** 46–52. MR0815600 https://doi.org/10.1007/BF02776078

[15] CHATPATANASIRI, R., KORSRILABUTR, T., TANGCHANACHAIANAN, P. and KIJSIRIKUL, B. (2010). A new kernelization framework for Mahalanobis distance learning algorithms. *Neurocomputing* **73** 1570–1579.

[16] CHAZAL, F., COHEN-STEINER, D. and LIEUTIER, A. (2006). A sampling theory for compact sets in Euclidean space. In *Computational Geometry* (*SCG*'06) 319–326. ACM, New York. MR2389341 https://doi.org/10.1145/1137856.1137904

[17] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric inference for probability measures. *Found. Comput. Math.* **11** 733–751. MR2859954 https://doi.org/10.1007/s10208-011-9098-0

[18] CHAZAL, F., GLISSE, M., LABRUÈRE, C. and MICHEL, B. (2014). Convergence rates for persistence diagram estimation in topological data analysis. In *Proceedings of the 31st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.). *Proceedings of Machine Learning Research* **32** 163–171. PMLR, Bejing, China.

[19] CHAZAL, F. and LIEUTIER, A. (2005). The "λ-medial axis". *Graph. Models* **67** 304–331. https://doi.org/10.1016/j.gmod.2005.01.002

[20] CHAZAL, F. and MICHEL, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Front. Artif. Intell.* **4** 667963. https://doi.org/10.3389/frai.2021.667963

[21] CHENNURU VANKADARA, L. and VON LUXBURG, U. (2018). Measures of distortion for machine learning. *Adv. Neural Inf. Process. Syst.* **31**.

[22] CHOLAQUIDIS, A., FRAIMAN, R. and MORENO, L. (2023). Universally consistent estimation of the reach. *J. Statist. Plann. Inference* **225** 110–120. MR4520497 https://doi.org/10.1016/j.jspi.2022.11.007

[23] CLARKE, F. H., STERN, R. J. and WOLENSKI, P. R. (1995). Proximal smoothness and the lower-C^2 property. *J. Convex Anal.* **2** 117–144. MR1363364

[24] COX, M. A. and COX, T. F. (2008). Multidimensional scaling. In *Handbook of Data Visualization* 315–347. Springer, Berlin.

[25] DIVOL, V. (2021). Minimax adaptive estimation in manifold inference. *Electron. J. Stat.* **15** 5888–5932. MR4355700 https://doi.org/10.1214/21-ejs1934

[26] FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078 https://doi.org/10.2307/1993504

[27] FEFFERMAN, C., IVANOV, S., LASSAS, M. and NARAYANAN, H. (2019). Fitting a manifold of large reach to noisy data. Preprint. Available at arXiv:1910.05084.

[28] GARCÍA TRILLOS, N., SANZ-ALONSO, D. and YANG, R. (2019). Local regularization of noisy point clouds: Improved global geometric estimates and data analysis. *J. Mach. Learn. Res.* **20** Paper No. 136, 37 pp. MR4002890

[29] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Minimax manifold estimation. *J. Mach. Learn. Res.* **13** 1263–1291. MR2930639

[30] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939 https://doi.org/10.1214/12-AOS994

[31] GLOBERSON, A. and ROWEIS, S. (2005). Metric learning by collapsing classes. *Adv. Neural Inf. Process. Syst.* **18**.

[32] GUILLAUMIN, M., VERBEEK, J. and SCHMID, C. (2009). Is that you? Metric learning approaches for face identification. In 2009 *IEEE* 12*th International Conference on Computer Vision* 498–505. IEEE, New York.

[33] HINTON, G. E. and ROWEIS, S. (2002). Stochastic neighbor embedding. *Adv. Neural Inf. Process. Syst.* **15**.

[34] KIM, A. K. H. and ZHOU, H. H. (2015). Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.* **9** 1562–1582. MR3376117 https://doi.org/10.1214/15-EJS1039

[35] KWOK, J. T. and TSANG, I. W. (2003). Learning with idealized kernels. In *Proceedings of the* 20*th International Conference on Machine Learning* (*ICML*-03) 400–407.

[36] NGUYEN, B., MORELL, C. and DE BAETS, B. (2017). Supervised distance metric learning through maximization of the Jeffrey divergence. *Pattern Recognit.* **64** 215–225.

[37] POLIQUIN, R. A., ROCKAFELLAR, R. T. and THIBAULT, L. (2000). Local differentiability of distance functions. *Trans. Amer. Math. Soc.* **352** 5231–5249. MR1694378 https://doi.org/10.1090/S0002-9947-00-02550-2

[38] PUCHKIN, N. and SPOKOINY, V. (2022). Structure-adaptive manifold estimation. *J. Mach. Learn. Res.* **23** Paper No. 40, 62 pp. MR4420765

[39] SUÁREZ, J. L., GARCÍA, S. and HERRERA, F. (2021). A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. *Neurocomputing* **425** 300–322.

[40] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323. https://doi.org/10.1126/science.290.5500.2319

[41] WASSERMAN, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.* **5** 501–535. MR3774757 https://doi.org/10.1146/annurev-statistics-031017-100045

[42] XING, E., JORDAN, M., RUSSELL, S. J. and NG, A. (2002). Distance metric learning with application to clustering with side-information. *Adv. Neural Inf. Process. Syst.* **15**.

[43] YANG, L. and JIN, R. (2006). Distance metric learning: A comprehensive survey. *Michigan State Univ.* **2** 4.

UNIVERSAL RANK INFERENCE VIA RESIDUAL SUBSAMPLING WITH APPLICATION TO LARGE NETWORKS

BY XIAO HAN[1,a], QING YANG[1,b] AND YINGYING FAN[2,c]

[1]*International Institute of Finance, School of Management, University of Science and Technology of China,*
[a]*xhan011@ustc.edu.cn*, [b]*yangq@ustc.edu.cn*

[2]*Data Sciences and Operations Department, University of Southern California,* [c]*fanyingy@marshall.usc.edu*

Determining the precise rank is an important problem in many large-scale applications with matrix data exploiting low-rank plus noise models. In this paper, we suggest a universal approach to rank inference via residual subsampling (RIRS) for testing and estimating rank in a wide family of models, including many popularly used network models such as the degree corrected mixed membership model as a special case. Our procedure constructs a test statistic via subsampling entries of the residual matrix after extracting the spiked components. The test statistic converges in distribution to the standard normal under the null hypothesis, and diverges to infinity with asymptotic probability one under the alternative hypothesis. The effectiveness of RIRS procedure is justified theoretically, utilizing the asymptotic expansions of eigenvectors and eigenvalues for large random matrices recently developed in (*J. Amer. Statist. Assoc.* **117** (2022) 996–1009) and (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** (2022) 630–653). The advantages of the newly suggested procedure are demonstrated through several simulation and real data examples.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] ADAMIC, L. A. and GLANCE, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43 ACM, New York, USA. https://doi.org/10.1145/1134271.1134277

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

[4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

[5] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. Available at arXiv:1705.05305.

[6] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

[7] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

[8] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

[9] CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 https://doi.org/10.1080/01621459.2016.1246365

[10] DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817 https://doi.org/10.1007/s11222-007-9046-7

[11] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990

[12] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505

[13] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. Available at arXiv:1710.00862.

[14] HAN, X., YANG, Q. and FAN, Y. (2023). Supplement to "Universal rank inference via residual subsampling with application to large networks." https://doi.org/10.1214/23-AOS2282SUPP

[15] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Comput. Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[16] JIN, J., KE, Z. T. and LUO, S. (2021). Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.* **49** 3408–3433. MR4352535 https://doi.org/10.1214/21-aos2089

[17] JIN, J., KE, Z. T., LUO, S. and WANG, M. (2020). Estimating the number of communities by stepwise goodness-of-fit. Available at arXiv:2009.09177.

[18] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[19] LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **12** 93–115. MR2953099 https://doi.org/10.1177/1471082X1001200105

[20] LE, C. M. and LEVINA, E. (2022). Estimating the number of communities in networks by spectral methods. Available at arXiv:1507.00827.

[21] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

[22] LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. MR4108931 https://doi.org/10.1093/biomet/asaa006

[23] MCDAID, A. F., MURPHY, T. B., FRIEL, N. and HURLEY, N. J. (2013). Improved Bayesian inference for the stochastic block model with application to large networks. *Comput. Statist. Data Anal.* **60** 12–31. MR3007016 https://doi.org/10.1016/j.csda.2012.10.021

[24] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[25] SALDAÑA, D. F., YU, Y. and FENG, Y. (2017). How many communities are there? *J. Comput. Graph. Statist.* **26** 171–181. MR3610418 https://doi.org/10.1080/10618600.2015.1096790

[26] WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

[27] WANG, Y. X. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528. MR3650391 https://doi.org/10.1214/16-AOS1457

[28] ZHANG, L. and AMINI, A. A. (2020). Adjusted chi-square test for degree-corrected block models. Available at arXiv:2012.15047.

[29] ZHAO, Y., LEVINA, E. and ZHU, J. (2011). Community extraction for social networks. *Proc. Natl. Acad. Sci. USA* **108** 7321–7326. https://doi.org/10.1073/pnas.1006642108

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

PAIRWISE INTERACTION FUNCTION ESTIMATION OF STATIONARY GIBBS POINT PROCESSES USING BASIS EXPANSION

BY ISMAÏLA BA[1,a], JEAN-FRANÇOIS COEURJOLLY[2,b] AND
FRANCISCO CUEVAS-PACHECO[3,c]

[1]*Department of Mathematics, Université du Québec à Montréal (UQAM),* [a]*ba.ismaila@courrier.uqam.ca*

[2]*Laboratory Jean Kuntzmann, Department DATA, Univ. Grenoble Alpes,* [b]*jean-francois.coeurjolly@univ-grenoble-alpes.fr*

[3]*Departamento de Matemática, Universidad Técnica Federico Santa María,* [c]*francisco.cuevas@usm.cl*

The class of Gibbs point processes (GPP) is a large class of spatial point processes able to model both clustered and repulsive point patterns. They are specified by their conditional intensity, which for a point pattern **x** and a location u, is roughly speaking the probability that an event occurs in an infinitesimal ball around u given the rest of the configuration is **x**. The most simple and natural class of models is the class of pairwise interaction point processes where the conditional intensity depends on the number of points and pairwise distances between them. This paper is concerned with the problem of estimating the pairwise interaction function nonparametrically. We propose to estimate it using an orthogonal series expansion of its logarithm. Such an approach has numerous advantages compared to existing ones. The estimation procedure is simple, fast and completely data-driven. We provide asymptotic properties such as consistency and asymptotic normality and show the efficiency of the procedure through simulation experiments and illustrate it with several data sets.

REFERENCES

[1] BA, I. and COEURJOLLY, J.-F. (2022). Inference for low- and high-dimensional inhomogeneous Gibbs point processes. *Scand. J. Stat.* https://doi.org/10.1111/sjos.12616

[2] BA, I., COEURJOLLY, J.-F. and CUEVAS-PACHECO, F. (2023). Supplement to "Pairwise interaction function estimation of stationary Gibbs point processes using basis expansion." https://doi.org/10.1214/23-AOS2284SUPP

[3] BADDELEY, A., COEURJOLLY, J.-F., RUBAK, E. and WAAGEPETERSEN, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika* **101** 377–392. MR3215354 https://doi.org/10.1093/biomet/ast060

[4] BADDELEY, A., GREGORI, P., MATEU, J., STOICA, R. and STOYAN, D. (2006). *Case Studies in Spatial Point Process Modeling* **185**. Springer.

[5] BADDELEY, A. and NAIR, G. (2012). Fast approximation of the intensity of Gibbs point processes. *Electron. J. Stat.* **6** 1155–1169. MR2988442 https://doi.org/10.1214/12-EJS707

[6] BADDELEY, A., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns*: *Methodology and Applications with R*. Chapman and Hall/CRC.

[7] BADDELEY, A. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Stat.* **42** 283–322. MR1794056 https://doi.org/10.1111/1467-842X.00128

[8] BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 617–666. MR2210685 https://doi.org/10.1111/j.1467-9868.2005.00519.x

[9] BERTHELSEN, K. K. and MØLLER, J. (2003). Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scand. J. Stat.* **30** 549–564.

[10] BERTHELSEN, K. K. and MØLLER, J. (2008). Non-parametric Bayesian inferences for inhomogeneous Markov point processes. *Aust. N. Z. J. Stat.* **50** 257–272. MR2455432 https://doi.org/10.1111/j.1467-842X.2008.00516.x

[11] BERTIN, E., BILLIOT, J.-M. and DROUILHET, R. (1999). Existence of "nearest-neighbour" spatial Gibbs models. *Adv. in Appl. Probab.* **31** 895–909. MR1747447 https://doi.org/10.1239/aap/1029955249

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

[13] BILLIOT, J.-M., COEURJOLLY, J.-F. and DROUILHET, R. (2008). Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electron. J. Stat.* **2** 234–264. MR2399195 https://doi.org/10.1214/07-EJS160

[14] BILLIOT, J.-M. and GOULARD, M. (2001). An estimation method of the pair potential function for Gibbs point processes on spheres. *Scand. J. Stat.* **28** 185–203. MR1844356 https://doi.org/10.1111/1467-9469.00231

[15] BOGNAR, M. A. and COWLES, M. K. (2004). Bayesian inference for pairwise interacting point processes. *Stat. Comput.* **14** 109–117. MR2037941 https://doi.org/10.1023/B:STCO.0000021409.73461.b9

[16] CHOIRUDDIN, A., COEURJOLLY, J.-F. and WAAGEPETERSEN, R. (2021). Information criteria for inhomogeneous spatial point processes. *Aust. N. Z. J. Stat.* **63** 119–143. MR4296145 https://doi.org/10.1111/anzs.12327

[17] COEURJOLLY, J.-F., CUEVAS-PACHECO, F. and WAAGEPETERSEN, R. (2019). Second-order variational equations for spatial point processes with a view to pair correlation function estimation. *Spat. Stat.* **30** 103–115. MR3936810 https://doi.org/10.1016/j.spasta.2019.03.001

[18] COEURJOLLY, J.-F., DEREUDRE, D., DROUILHET, R. and LAVANCIER, F. (2012). Takacs–Fiksel method for stationary marked Gibbs point processes. *Scand. J. Stat.* **39** 416–443. MR2971630 https://doi.org/10.1111/j.1467-9469.2011.00738.x

[19] COEURJOLLY, J.-F. and DROUILHET, R. (2010). Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard–Jones model. *Electron. J. Stat.* **4** 677–706. MR2678967 https://doi.org/10.1214/09-EJS494

[20] COEURJOLLY, J.-F., GUAN, Y., KHANMOHAMMADI, M. and WAAGEPETERSEN, R. (2016). Towards optimal Takacs–Fiksel estimation. *Spat. Stat.* **18** 396–411. MR3575498 https://doi.org/10.1016/j.spasta.2016.08.002

[21] COEURJOLLY, J.-F. and LAVANCIER, F. (2017). Parametric estimation of pairwise Gibbs point processes with infinite range interaction. *Bernoulli* **23** 1299–1334. MR3606767 https://doi.org/10.3150/15-BEJ779

[22] COEURJOLLY, J.-F. and LAVANCIER, F. (2018). Intensity approximation for pairwise interaction Gibbs point processes using determinantal point processes. *Electron. J. Stat.* **12** 3181–3203. MR3858696 https://doi.org/10.1214/18-ejs1477

[23] COEURJOLLY, J.-F. and RUBAK, E. (2013). Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scand. J. Stat.* **40** 669–684. MR3145111 https://doi.org/10.1111/sjos.12017

[24] DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II*: *General Theory and Structure*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR2371524 https://doi.org/10.1007/978-0-387-49835-5

[25] DANIEL, J., HORROCKS, J. and UMPHREY, G. J. (2018). Penalized composite likelihoods for inhomogeneous Gibbs point process models. *Comput. Statist. Data Anal.* **124** 104–116. MR3787616 https://doi.org/10.1016/j.csda.2018.02.005

[26] DAVID, S. V. and HEINRICH, L. (2010). Central limit theorem for the integrated squared error of the empirical second-order product density and goodness-of-fit tests for stationary point processes. Preprint Nr. 17/2010, Institut für Mathematik, Univ. Augsburg.

[27] DEREUDRE, D. (2019). Introduction to the theory of Gibbs point processes. In *Stochastic Geometry. Lecture Notes in Math.* **2237** 181–229. Springer, Cham. MR3931586

[28] DEREUDRE, D. and LAVANCIER, F. (2017). Consistency of likelihood estimation for Gibbs point processes. *Ann. Statist.* **45** 744–770. MR3650399 https://doi.org/10.1214/16-AOS1466

[29] DIGGLE, P. J., GATES, D. J. and STIBBARD, A. (1987). A nonparametric estimator for pairwise-interaction point processes. *Biometrika* **74** 763–770. MR0919844 https://doi.org/10.1093/biomet/74.4.763

[30] DIGGLE, P. J. and GRATTON, R. J. (1984). Monte Carlo methods of inference for implicit statistical models. *J. Roy. Statist. Soc. Ser. B* **46** 193–227. MR0781880

[31] EFROMOVICH, S. (1999). *Nonparametric Curve Estimation*: *Methods*, *Theory*, *and Applications*. *Springer Series in Statistics*. Springer, New York. MR1705298

[32] FIKSEL, T. (1984). Estimation of parametrized pair potentials of marked and nonmarked Gibbsian point processes. *Elektron. Informationsverarb. Kybernet.* **20** 270–278. MR0760146

[33] FIKSEL, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *Statistics* **19** 77–86. MR0921627 https://doi.org/10.1080/02331888808802074

[34] GABRIEL, E., ROWLINGSON, B. S. and DIGGLE, P. J. (2013). stpp: An R package for plotting, simulating and analyzing spatio-temporal point patterns. *J. Stat. Softw.* **53** 1–29.

[35] GAO, X. and SONG, P. X.-K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *J. Amer. Statist. Assoc.* **105** 1531–1540. MR2796569 https://doi.org/10.1198/jasa.2010.tm09414

[36] GATES, D. J. and WESTCOTT, M. (1986). Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.* **38** 123–135. MR0837242 https://doi.org/10.1007/BF02482505

[37] GEORGII, H.-O. (1979). *Canonical Gibbs Measures. Lecture Notes in Math.* **760**. Springer, Berlin. MR0551621

[38] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR2807681 https://doi.org/10.1515/9783110250329

[39] GEYER, C. J. and MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.* **21** 359–373. MR1310082

[40] HAAR, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* **69** 331–371. MR1511592 https://doi.org/10.1007/BF01456326

[41] HALL, P. (1986). On the rate of convergence of orthogonal series density estimators. *J. Roy. Statist. Soc. Ser. B* **48** 115–122. MR0848057

[42] HALL, P. (1987). Cross-validation and the smoothing of orthogonal series density estimators. *J. Multivariate Anal.* **21** 189–206. MR0884096 https://doi.org/10.1016/0047-259X(87)90001-7

[43] HANSEN, J. P. and MCDONALD, I. R. (1976). *Theory of Simple Liquids: By Jean Pierre Hansen and Ian R. Mcdonald*. Academic Press, San Diego.

[44] HEIKKINEN, J. and PENTTINEN, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli* **5** 1119–1136. MR1735787 https://doi.org/10.2307/3318562

[45] IFTIMI, A., VAN LIESHOUT, M.-C. and MONTES, F. (2018). A multi-scale area-interaction model for spatio-temporal point patterns. *Spat. Stat.* **26** 38–55. MR3846278 https://doi.org/10.1016/j.spasta.2018.06.001

[46] ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns. Statistics in Practice*. Wiley, Chichester. MR2384630

[47] JALILIAN, A., GUAN, Y. and WAAGEPETERSEN, R. (2019). Orthogonal series estimation of the pair correlation function of a spatial point process. *Statist. Sinica* **29** 769–787. MR3931387

[48] JENSEN, J. L. and KÜNSCH, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46** 475–486. MR1309718

[49] JENSEN, J. L. and MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461. MR1111528

[50] MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability* **100**. CRC Press/CRC, Boca Raton, FL. MR2004226

[51] MYLLYMÄKI, M. and MRKVIČKA, T. (2019). GET: Global envelopes in R. ArXiv preprint. Available at arXiv:1911.06583.

[52] NGUYEN, X.-X. and ZESSIN, H. (1979). Integral and differential characterizations of the Gibbs process. *Math. Nachr.* **88** 105–115. MR0543396 https://doi.org/10.1002/mana.19790880109

[53] NGUYEN, X.-X. and ZESSIN, H. (1979). Ergodic theorems for spatial processes. *Z. Wahrsch. Verw. Gebiete* **48** 133–158. MR0534841 https://doi.org/10.1007/BF01886869

[54] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. MR0143282 https://doi.org/10.1214/aoms/1177704472

[55] PERCUS, J. (1964). The pair distribution function in classical statistical mechanics. In *The Equilibrium Theory of Classical Fluids* (H. L. Frish and J. L. Lebowitz, eds.)

[56] RAEISI, M., BONNEU, F. and GABRIEL, E. (2021). A spatio-temporal multi-scale model for Geyer saturation point process: Application to forest fire occurrences. *Spat. Stat.* **41** Paper No. 100492. MR4207952 https://doi.org/10.1016/j.spasta.2021.100492

[57] RIPLEY, B. and SILVERMAN, B. (1978). Quick tests for spatial interaction. *Biometrika* **65** 641–642.

[58] RUELLE, D. (1969). *Statistical Mechanics: Rigorous Results*. W. A. Benjamin, Inc., New York-Amsterdam. MR0289084

[59] RUELLE, D. (1970). Superstable interactions in classical statistical mechanics. *Comm. Math. Phys.* **18** 127–159. MR0266565

[60] STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, 1st ed. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

[61] STOYAN, D. and STOYAN, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, Chichester. MR1297125

[62] STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **62** 467–475. MR0383493 https://doi.org/10.1093/biomet/62.2.467

[63] STUCKI, K. and SCHUHMACHER, D. (2014). Bounds for the probability generating functional of a Gibbs point process. *Adv. in Appl. Probab.* **46** 21–34. MR3189046 https://doi.org/10.1239/aap/1396360101

[64] SZEGÖ, G. (1939). *Orthogonal Polynomials. American Mathematical Society Colloquium Publications*, *Vol.* 23. Amer. Math. Soc., New York. MR0000077

[65] TAKACS, R. (1986). Estimator for the pair-potential of a Gibbsian point process. *Statistics* **17** 429–433. MR0849741 https://doi.org/10.1080/02331888608801956

[66] WAHBA, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.* **9** 146–156. MR0600541

[67] WASSERMAN, L. (2006). *All of Nonparametric Statistics. Springer Texts in Statistics*. Springer, New York. MR2172729

[68] WATSON, G. S. (1969). Density estimation by orthogonal series. *Ann. Math. Stat.* **40** 1496–1498. MR0242332 https://doi.org/10.1214/aoms/1177697523

[69] WHITTLE, P. (1958). On the smoothing of probability density functions. *J. Roy. Statist. Soc. Ser. B* **20** 334–343. MR0099098

EFFICIENCY OF ESTIMATORS FOR LOCALLY ASYMPTOTICALLY NORMAL QUANTUM STATISTICAL MODELS

BY AKIO FUJIWARA[1,a] AND KOICHI YAMAGATA[2,b]

[1]*Department of Mathematics, Osaka University,* [a]*fujiwara@math.sci.osaka-u.ac.jp*

[2]*Principles of Informatics Research Division, National Institute of Informatics,* [b]*yamagata@nii.ac.jp*

We herein establish an asymptotic representation theorem for locally asymptotically normal quantum statistical models. This theorem enables us to study the asymptotic efficiency of quantum estimators, such as quantum regular estimators and quantum minimax estimators, leading to a universal tight lower bound beyond the i.i.d. assumption. This formulation complements the theory of quantum contiguity developed in the previous paper [Fujiwara and Yamagata, *Bernoulli* **26** (2020) 2105–2141], providing a solid foundation of the theory of weak quantum local asymptotic normality.

REFERENCES

[1] ANDO, T. (1978). *Topics on Operator Inequalities*. Hokkaido Univ., Research Institute of Applied Electricity, Division of Applied Mathematics, Sapporo. MR0482378

[2] FUJIWARA, A. (2006). Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *J. Phys. A* **39** 12489–12504. MR2265836 https://doi.org/10.1088/0305-4470/39/40/014

[3] FUJIWARA, A. and YAMAGATA, K. (2020). Noncommutative Lebesgue decomposition and contiguity with applications in quantum statistics. *Bernoulli* **26** 2105–2142. MR4091103 https://doi.org/10.3150/19-BEJ1185

[4] FUJIWARA, A. and YAMAGATA, K. (2023). Supplement to "Efficiency of estimators for locally asymptotically normal quantum statistical models." https://doi.org/10.1214/23-AOS2285SUPP

[5] GILL, R. D. and GUȚĂ, M. I. (2013). On asymptotic quantum statistical inference. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes*. *Inst. Math. Stat.* (*IMS*) *Collect.* **9** 105–127. IMS, Beachwood, OH. MR3186752 https://doi.org/10.1214/12-IMSCOLL909

[6] GILL, R. D. and MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* **61** 042312.

[7] GUȚĂ, M. and JENČOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. MR2346393 https://doi.org/10.1007/s00220-007-0340-1

[8] GUȚĂ, M. and KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* (3) **73** 052108, 15. MR2229156 https://doi.org/10.1103/PhysRevA.73.052108

[9] HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. I*: *Theory of Statistics* 175–194. Univ. California Press, Berkeley, CA. MR0400513

[10] HIAI, F. (2021). *Lectures on Selected Topics in Von Neumann Algebras*. EMS Series of Lectures in Mathematics. EMS Press, Berlin. MR4331436 https://doi.org/10.4171/ELM/32

[11] HOLEVO, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. *Quaderni/Monographs* **1**. Edizioni della Normale, Pisa. MR2797301 https://doi.org/10.1007/978-88-7642-378-9

[12] HUSIMI, K. (1940). Some formal properties of the density matrix. *Proc. Phys.-Math. Soc. Jpn.* **22** 264–314.

[13] JAKŠIĆ, V., PAUTRAT, Y. and PILLET, C.-A. (2010). A non-commutative Lévy–Cramér continuity theorem. *Markov Process. Related Fields* **16** 59–78. MR2664336

[14] JAKŠIĆ, V., PAUTRAT, Y. and PILLET, C.-A. (2010). A quantum central limit theorem for sums of independent identically distributed random variables. *J. Math. Phys.* **51** 015208, 8. MR2605841 https://doi.org/10.1063/1.3285287

[15] JUDGE, G. G. and BOCK, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. *Studies in Mathematical and Managerial Economics*, *Vol.* 25. North-Holland, Amsterdam–Oxford. MR0483199

[16] KADISON, R. V. and RINGROSE, J. R. (1997). *Fundamentals of the Theory of Operator Algebras. Vol. II*: *Advanced Theory*. *Graduate Studies in Mathematics* **16**. Amer. Math. Soc., Providence, RI. MR1468230 https://doi.org/10.1090/gsm/016

[17] KAHN, J. and GUȚĂ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764 https://doi.org/10.1007/s00220-009-0787-3

[18] KUBO, F. and ANDO, T. (1979/80). Means of positive linear operators. *Math. Ann.* **246** 205–224. MR0563399 https://doi.org/10.1007/BF01371042

[19] LAHIRY, S. and NUSSBAUM, M. (2021). Local asymptotic normality and optimal estimation of low-rank quantum systems. Available at arXiv:2111.03279.

[20] REED, M. and SIMON, B. (1979). *Methods of Modern Mathematical Physics. I*: *Functional Analysis*. Academic Press, New York.

[21] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[22] YAMAGATA, K. (2011). Efficiency of quantum state tomography for qubits. *Int. J. Quantum Inf.* **9** 1167–1183. MR2931457 https://doi.org/10.1142/S0219749911007551

[23] YAMAGATA, K., FUJIWARA, A. and GILL, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** 2197–2217. MR3127863 https://doi.org/10.1214/13-AOS1147

INFERENCE IN ISING MODELS ON DENSE REGULAR GRAPHS

BY YUANZHE XU[a] AND SUMIT MUKHERJEE[b]

Department of Statistics, Columbia University, [a]*yuanzhe.xu@columbia.edu,* [b]*sm3949@columbia.edu*

In this paper, we derive the limit of experiments for one-parameter Ising models on dense regular graphs. In particular, we show that the limiting experiment is Gaussian in the "low temperature" regime, and non-Gaussian in the "critical" regime. We also derive the limiting distributions of the maximum likelihood and maximum pseudolikelihood estimators, and study limiting power for tests of hypothesis against contiguous alternatives. To the best of our knowledge, this is the first attempt at establishing the classical limits of experiments for Ising models (and more generally, Markov random fields).

REFERENCES

[1] BASAK, A. and MUKHERJEE, S. (2017). Universality of the mean-field for the Potts model. *Probab. Theory Related Fields* **168** 557–600. MR3663625 https://doi.org/10.1007/s00440-016-0718-0

[2] BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2019). Exact recovery in the Ising blockmodel. *Ann. Statist.* **47** 1805–1834. MR3953436 https://doi.org/10.1214/17-AOS1620

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

[4] BESAG, J. (1975). Statistical analysis of non-lattice data. *J. R. Stat. Soc., Ser. D, Stat.* **24** 179–195.

[5] BHATTACHARYA, B. B. and MUKHERJEE, S. (2018). Inference in Ising models. *Bernoulli* **24** 493–525. MR3706767 https://doi.org/10.3150/16-BEJ886

[6] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626 https://doi.org/10.1016/j.aim.2008.07.008

[7] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)* **176** 151–219. MR2925382 https://doi.org/10.4007/annals.2012.176.1.2

[8] CHATTERJEE, S. (2007). Estimation in spin glasses: A first step. *Ann. Statist.* **35** 1931–1946. MR2363958 https://doi.org/10.1214/009053607000000109

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

[10] COMETS, F. and GIDAS, B. (1991). Asymptotics of maximum likelihood estimators for the Curie–Weiss model. *Ann. Statist.* **19** 557–578. MR1105836 https://doi.org/10.1214/aos/1176348111

[11] DEB, N., MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2023). Detecting structured signals in Ising models. *Ann. Appl. Probab.* To appear.

[12] DEB, N. and MUKHERJEE, S. (2023). Fluctuations in mean-field Ising models. *Ann. Appl. Probab.* **33** 1961–2003. MR4583662 https://doi.org/10.1214/22-aap1857

[13] DEMBO, A. and MONTANARI, A. (2010). Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20** 565–592. MR2650042 https://doi.org/10.1214/09-AAP627

[14] ELLIS, R. S. and NEWMAN, C. M. (1978). The statistics of Curie–Weiss models. *J. Stat. Phys.* **19** 149–161. MR0503332 https://doi.org/10.1007/BF01012508

[15] GHOSAL, P. and MUKHERJEE, S. (2020). Joint estimation of parameters in Ising model. *Ann. Statist.* **48** 785–810. MR4102676 https://doi.org/10.1214/19-AOS1822

[16] GRAY, R. M. et al. (2006). Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory* **2** 155–239.

[17] HARRIS, A. B. (1974). Effect of random defects on the critical behaviour of Ising models. *J. Phys. C, Solid State Phys.* **7** 1671.

[18] ISING, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift Für Physik* **31** 253–258.

[19] KABLUCHKO, Z., LÖWE, M. and SCHUBERT, K. (2019). Fluctuations of the magnetization for Ising models on dense Erdős–Rényi random graphs. *J. Stat. Phys.* **177** 78–94. MR4003721 https://doi.org/10.1007/s10955-019-02358-5

[20] LE CAM, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol.* 1: *Theory of Statistics* 245–261. Univ. California Press, Berkeley, CA. MR0415819

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

[22] MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2018). Global testing against sparse alternatives under Ising models. *Ann. Statist.* **46** 2062–2093. MR3845011 https://doi.org/10.1214/17-AOS1612

[23] ONSAGER, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* (2) **65** 117–149. MR0010315

[24] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*, *Vol.* 3. Cambridge Univ. Press, Cambridge.

[25] XU, Y. and MUKHERJEE, S. (2022). Ising models on dense regular graphs. arXiv preprint. Available at arXiv:2210.13178v1.

[26] XU, Y. and MUKHERJEE, S. (2023). Supplement to "Inference in Ising models on dense regular graphs." https://doi.org/10.1214/23-AOS2286SUPP

AUTOREGRESSIVE APPROXIMATIONS TO NONSTATIONARY TIME SERIES WITH INFERENCE AND APPLICATIONS

BY XIUCAI DING[1,a] AND ZHOU ZHOU[2,b]

[1]*Department of Statistics, University of California,* [a]*xcading@ucdavis.edu*

[2]*Department of Statistical Sciences, University of Toronto,* [b]*zhou@utstat.utoronto.ca*

Understanding the time-varying structure of complex temporal systems is one of the main challenges of modern time-series analysis. In this paper, we show that every uniformly-positive-definite-in-covariance and sufficiently short-range dependent nonstationary and nonlinear time series can be well approximated globally by a white-noise-driven autoregressive (AR) process of slowly diverging order. To our best knowledge, it is the first time such a structural approximation result is established for general classes of nonstationary time series. A high-dimensional $\mathcal{L}^2$ test and an associated multiplier bootstrap procedure are proposed for the inference of the AR approximation coefficients. In particular, an adaptive stability test is proposed to check whether the AR approximation coefficients are time-varying, a frequently encountered question for practitioners and researchers of time series. As an application, globally optimal sffollowing hort-term forecasting theory and methodology for a wide class of locally stationary time series are established via the method of sieves.

REFERENCES

[1] BAXTER, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.* **10** 137–144. MR0149584 https://doi.org/10.7146/math.scand.a-10520

[2] BAXTER, G. (1963). A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois J. Math.* **7** 97–103. MR0145285

[3] BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series*: *Theory and Methods*. *Springer Series in Statistics*. Springer, New York. MR0868859 https://doi.org/10.1007/978-1-4899-0004-3

[4] CAI, T. T., LIU, W. and ZHOU, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488. MR3476606 https://doi.org/10.1214/13-AOS1171

[5] CHEN, L. and FANG, X. (2011). Multivariate normal approximation by Stein's method: The concentration inequality approach. ArXiv preprint. Available at arXiv:1111.4073.

[6] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, *Vol. 6B* (J. J. Heckman and E. E. Leamer, eds.). Chapter 76.

[7] CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021. MR3161455 https://doi.org/10.1214/13-AOS1182

[8] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 https://doi.org/10.1214/aos/1015957480

[9] DAHLHAUS, R. (2012). Locally stationary processes. In *Handbook of Statistics* **30** 351–413. Elsevier, Amsterdam.

[10] DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

[11] DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352 https://doi.org/10.1214/009053606000000227

[12] DAS, S. and POLITIS, D. N. (2021). Predictive inference for locally stationary time series with an application to climate data. *J. Amer. Statist. Assoc.* **116** 919–934. MR4270034 https://doi.org/10.1080/01621459.2019.1708368

[13] DEMKO, S., MOSS, W. F. and SMITH, P. W. (1984). Decay rates for inverses of band matrices. *Math. Comp.* **43** 491–499. MR0758197 https://doi.org/10.2307/2008290

[14] DETTE, H., PREUSS, P. and VETTER, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768 https://doi.org/10.1198/jasa.2011.tm10811

[15] DETTE, H. and WU, W. (2022). Prediction in locally stationary time series. *J. Bus. Econom. Statist.* **40** 370–381. MR4356579 https://doi.org/10.1080/07350015.2020.1819296

[16] DETTE, H., WU, W. and ZHOU, Z. (2019). Change point analysis of correlation in non-stationary time series. *Statist. Sinica* **29** 611–643. MR3931381

[17] DING, X. and ZHOU, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *Ann. Statist.* **48** 2455–2477. MR4134802 https://doi.org/10.1214/19-AOS1894

[18] DING, X. and ZHOU, Z. (2023). Supplement to "AutoRegressive approximations to nonstationary time series with inference and applications." https://doi.org/10.1214/23-AOS2288SUPP

[19] DING, X. and ZHOU, Z. (2021). Simultaneous sieve inference for time-inhomogeneous nonlinear time series regression. ArXiv preprint. Available at arXiv:2112.08545.

[20] DWIVEDI, Y. and SUBBA RAO, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91. MR2790673 https://doi.org/10.1111/j.1467-9892.2010.00685.x

[21] FANG, X. (2016). A multivariate CLT for bounded decomposable random vectors with the best known rate. *J. Theoret. Probab.* **29** 1510–1523. MR3571252 https://doi.org/10.1007/s10959-015-0619-7

[22] FRYZLEWICZ, P., VAN BELLEGEM, S. and VON SACHS, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.* **55** 737–764. MR2028615 https://doi.org/10.1007/BF02523391

[23] JIN, L., WANG, S. and WANG, H. (2015). A new non-parametric stationarity test of time series in the time domain. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 893–922. MR3414133 https://doi.org/10.1111/rssb.12091

[24] KANG, X., DENG, X., TSUI, K.-W. and POURAHMADI, M. (2020). On variable ordination of modified Cholesky decomposition for estimating time-varying covariance matrices. *Int. Stat. Rev.* **88** 616–641. MR4180670 https://doi.org/10.1111/insr.12357

[25] KARMAKAR, S., RICHTER, S. and WU, W. B. (2022). Simultaneous inference for time-varying models. *J. Econometrics* **227** 408–428. MR4384679 https://doi.org/10.1016/j.jeconom.2021.03.002

[26] KLEY, T., PREUSS, P. and FRYZLEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **13** 3710–3774. MR4013749 https://doi.org/10.1214/19-ejs1606

[27] KOLMOGOROFF, A. N. (1941). Stationary sequences in Hilbert's space. *Moscow Univ. Math. Bull.* **2** 1–40. MR0009098

[28] KOLMOGOROV, A. N. (1941). Interpolation and extrapolation of stationary random sequences. *Izv. Ross. Akad. Nauk Ser. Mat.* **5** 3–14.

[29] MAYER, U., ZÄHLE, H. and ZHOU, Z. (2020). Functional weak limit theorem for a local empirical process of non-stationary time series and its application. *Bernoulli* **26** 1891–1911. MR4091095 https://doi.org/10.3150/19-BEJ1174

[30] NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 879–904. MR3124795 https://doi.org/10.1111/rssb.12015

[31] PAPARODITIS, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. MR2597589 https://doi.org/10.3150/08-BEJ179

[32] PAPARODITIS, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Amer. Statist. Assoc.* **105** 839–851. MR2724865 https://doi.org/10.1198/jasa.2010.tm08243

[33] PAPARODITIS, E. and PREUSS, P. (2016). On local power properties of frequency domain-based tests for stationarity. *Scand. J. Stat.* **43** 664–682. MR3543316 https://doi.org/10.1111/sjos.12197

[34] POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** 677–690. MR1723786 https://doi.org/10.1093/biomet/86.3.677

[35] PREUSS, P., VETTER, M. and DETTE, H. (2013). A test for stationarity based on empirical processes. *Bernoulli* **19** 2715–2749. MR3160569 https://doi.org/10.3150/12-BEJ472

[36] ROBINSON, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25** 2054–2083. MR1474083 https://doi.org/10.1214/aos/1069362387

[37] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.* **23** 470–472. MR0049525 https://doi.org/10.1214/aoms/1177729394

[38] ROUEFF, F. and SÁNCHEZ-PÉREZ, A. (2018). Prediction of weakly locally stationary processes by auto-regression. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** 1215–1239. MR3867204 https://doi.org/10.30757/alea.v15-45

[39] SHERK, J. (2014). *Not Looking for Work*: *Why Labor Force Participation Has Fallen During the Recovery*. Available at https://www.heritage.org/jobs-and-labor/report/not-looking-work-why-labor-force-participation-has-fallen-during-the-recovery.

[40] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

[41] VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614 https://doi.org/10.1214/12-AOS1043

[42] WIENER, N. (1949). *Extrapolation*, *Interpolation and Smoothing of Stationary Time Series*. Wiley, New York.

[43] WOLD, H. (1954). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell, Stockholm. 2d ed, With an appendix by Peter Whittle. MR0061344

[44] WU, W. and ZHOU, Z. (2019). Multiscale jump testing and estimation under complex temporal dynamics. ArXiv preprint. Available at arXiv:1909.06307.

[45] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

[46] WU, W. B. and MIELNICZUK, J. (2010). A new look at measuring dependence. In *Dependence in Probability and Statistics*. *Lect. Notes Stat.* **200** 123–142. Springer, Berlin. MR2731829 https://doi.org/10.1007/978-3-642-14104-1_7

[47] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

[48] YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. MR2719856

[49] ZHAO, Z. (2015). Inference for local autocorrelations in locally stationary models. *J. Bus. Econom. Statist.* **33** 296–306. MR3337064 https://doi.org/10.1080/07350015.2014.948177

[50] ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

[51] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636

[52] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 513–531. MR2758526 https://doi.org/10.1111/j.1467-9868.2010.00743.x

OPTIMAL DISCRIMINANT ANALYSIS IN HIGH-DIMENSIONAL LATENT FACTOR MODELS

BY XIN BING[1,a] AND MARTEN WEGKAMP[2,b]

[1]*Department of Statistical Sciences, University of Toronto,* [a]*xin.bing@utoronto.ca*

[2]*Department of Mathematics & Department of Statistics and Data Science, Cornell University,* [b]*mhw73@cornell.edu*

In high-dimensional classification problems, a commonly used approach is to first project the high-dimensional features into a lower-dimensional space, and base the classification on the resulting lower-dimensional projections. In this paper, we formulate a latent-variable model with a hidden low-dimensional structure to justify this two-step procedure and to guide which projection to choose. We propose a computationally efficient classifier that takes certain principal components (PCs) of the observed features as projections, with the number of retained PCs selected in a data-driven way. A general theory is established for analyzing such two-step classifiers based on any projections. We derive explicit rates of convergence of the excess risk of the proposed PC-based classifier. The obtained rates are further shown to be optimal up to logarithmic factors in the minimax sense. Our theory allows the lower dimension to grow with the sample size and is also valid even when the feature dimension (greatly) exceeds the sample size. Extensive simulations corroborate our theoretical findings. The proposed method also performs favorably relative to other existing discriminant methods on three real data examples.

REFERENCES

[1] ABRAMOVICH, F. and PENSKY, M. (2019). Classification with many classes: Challenges and pluses. *J. Multivariate Anal.* **174** 104536. MR3995262 https://doi.org/10.1016/j.jmva.2019.104536

[2] ALON, U., BARKAI, N., NOTTERMAN, D. A., GISH, K., YBARRA, S., MACK, D. and LEVINE, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **96** 6745–6750.

[3] ANTONIADIS, A., LAMBERT-LACROIX, S. and LEBLANC, F. (2003). Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics* **19** 563–570.

[4] AZIZYAN, M., SINGH, A. and WASSERMAN, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. In *Advances in Neural Information Processing Systems* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26**. Curran Associates, Red Hook.

[5] BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313 https://doi.org/10.1214/11-AOS966

[6] BAI, J. and NG, S. (2008). Forecasting economic time series using targeted predictors. *J. Econometrics* **146** 304–317. MR2465175 https://doi.org/10.1016/j.jeconom.2008.08.010

[7] BAIR, E., HASTIE, T., PAUL, D. and TIBSHIRANI, R. (2006). Prediction by supervised principal components. *J. Amer. Statist. Assoc.* **101** 119–137. MR2252436 https://doi.org/10.1198/016214505000000628

[8] BARKER, M. and RAYENS, W. (2003). Partial least squares for discrimination. *J. Chemom.* **17** 166–173.

[9] BARSHAN, E., GHODSI, A., AZIMIFAR, Z. and JAHROMI, M. Z. (2011). Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognit.* **44** 1357–1371.

[10] BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429. MR3001131 https://doi.org/10.3982/ECTA9626

[11] BIAU, G., BUNEA, F. and WEGKAMP, M. H. (2005). Functional classification in Hilbert spaces. *IEEE Trans. Inf. Theory* **51** 2163–2172. MR2235289 https://doi.org/10.1109/TIT.2005.847705

[12] BING, X., BUNEA, F., NING, Y. and WEGKAMP, M. (2020). Adaptive estimation in structured factor models with applications to overlapping clustering. *Ann. Statist.* **48** 2055–2081. MR4134786 https://doi.org/10.1214/19-AOS1877

[13] BING, X., BUNEA, F., STRIMAS-MACKEY, S. and WEGKAMP, M. (2021). Prediction under latent factor regression: Adaptive PCR, interpolating predictors and beyond. *J. Mach. Learn. Res.* **22** Paper No. 177. MR4318533 https://doi.org/10.22405/2226-8383-2021-22-1-177-187

[14] BING, X. and WEGKAMP, M. (2022). Interpolating discriminant functions in high-dimensional Gaussian latent mixtures. Available at arXiv:2210.14347.

[15] BING, X. and WEGKAMP, M. (2023). Supplement to "Optimal discriminant analysis in high-dimensional latent factor models." https://doi.org/10.1214/23-AOS2289SUPP

[16] BING, X. and WEGKAMP, M. H. (2019). Adaptive estimation of the rank of the coefficient matrix in high-dimensional multivariate response regression models. *Ann. Statist.* **47** 3157–3184. MR4025738 https://doi.org/10.1214/18-AOS1774

[17] BOULESTEIX, A.-L. (2004). PLS dimension reduction for classification with microarray data. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 33. MR2101480 https://doi.org/10.2202/1544-6115.1075

[18] CAI, T. and LIU, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *J. Amer. Statist. Assoc.* **106** 1566–1577. MR2896857 https://doi.org/10.1198/jasa.2011.tm11199

[19] CAI, T. T. and ZHANG, L. (2019). High dimensional linear discriminant analysis: Optimality, adaptive algorithm and missing data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 675–705. MR3997097

[20] CAI, T. T. and ZHANG, L. (2021). A convex optimization approach to high-dimensional sparse quadratic discriminant analysis. *Ann. Statist.* **49** 1537–1568. MR4298872 https://doi.org/10.1214/20-aos2012

[21] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[22] CHIAROMONTE, F. and MARTINELLI, J. (2002). Dimension reduction strategies for analyzing global gene expression data with a response. *Math. Biosci.* **176** 123–144. MR1869195 https://doi.org/10.1016/S0025-5564(01)00106-7

[23] CLEMMENSEN, L., HASTIE, T., WITTEN, D. and ERSBØLL, B. (2011). Sparse discriminant analysis. *Technometrics* **53** 406–413. MR2850472 https://doi.org/10.1198/TECH.2011.08118

[24] DAI, J. J., LIEU, L. and ROCKE, D. (2006). Dimension reduction for classification with gene expression microarray data. *Stat. Appl. Genet. Mol. Biol.* **5** Art. 6. MR2221299 https://doi.org/10.2202/1544-6115.1147

[25] DETTLING, M. (2004). BagBoosting for tumor classification with gene expression data. *Bioinformatics* **20** 3583–3593. https://doi.org/10.1093/bioinformatics/bth447

[26] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

[27] FAN, J. and FAN, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* **36** 2605–2637. MR2485009 https://doi.org/10.1214/07-AOS504

[28] FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

[29] FAN, J., XUE, L. and YAO, J. (2017). Sufficient forecasting using factor models. *J. Econometrics* **201** 292–306. MR3717565 https://doi.org/10.1016/j.jeconom.2017.08.009

[30] GHOSH, D. (2001). Singular value decomposition regression models for classification of tumors from microarray experiments. In *Biocomputing* 2002 18–29. World Scientific, Singapore.

[31] HADEF, H. and DJEBABRA, M. (2019). Proposal method for the classification of industrial accident scenarios based on the improved principal components analysis (improved PCA). *Prod. Eng.* **13** 53–60.

[32] HAHN, P. R., CARVALHO, C. M. and MUKHERJEE, S. (2013). Partial factor modeling: Predictor-dependent shrinkage for linear regression. *J. Amer. Statist. Assoc.* **108** 999–1008. MR3174679 https://doi.org/10.1080/01621459.2013.779843

[33] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[34] HOTELLING, H. (1957). The relations of the newer multivariate statistical methods to factor analysis. *Br. J. Stat. Psychol.* **10** 69–79.

[35] IZENMAN, A. J. (2008). *Modern Multivariate Statistical Techniques*: *Regression*, *Classification*, *and Manifold Learning*. *Springer Texts in Statistics*. Springer, New York. MR2445017 https://doi.org/10.1007/978-0-387-78189-1

[36] JIN, D., HENRY, P., SHAN, J. and CHEN, J. (2021). Classification of cannabis strains in the Canadian market with discriminant analysis of principal components using genome-wide single nucleotide polymorphisms. *PLoS ONE* **16** e0253387.

[37] LI, H. (2016). Accurate and efficient classification based on common principal components analysis for multivariate time series. *Neurocomputing* **171** 744–753.

[38] MA, Z., LIU, Z., ZHAO, Y., ZHANG, L., LIU, D., REN, T., ZHANG, X. and LI, S. (2020). An unsupervised crop classification method based on principal components isometric binning. *ISPRS Int.l J. Geo-Inf.* **9** 648.

[39] MAI, Q., YANG, Y. and ZOU, H. (2019). Multiclass sparse discriminant analysis. *Statist. Sinica* **29** 97–111. MR3889359

[40] MAI, Q., ZOU, H. and YUAN, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42. MR2899661 https://doi.org/10.1093/biomet/asr066

[41] MALLARY, C., BERG, C., BUCK, J. R., TANDON, A. and ANDONIAN, A. (2022). Acoustic rainfall detection with linear discriminant functions of principal components. *J. Acoust. Soc. Am.* **151** A149–A149.

[42] NGUYEN, D. V. and ROCKE, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* **18** 39–50.

[43] SHAO, J., WANG, Y., DENG, X. and WANG, S. (2011). Sparse linear discriminant analysis by thresholding for high dimensional data. *Ann. Statist.* **39** 1241–1265. MR2816353 https://doi.org/10.1214/10-AOS870

[44] SINGH, D., FEBBO, P. G., ROSS, K., JACKSON, D. G., MANOLA, J., LADD, C., TAMAYO, P., RENSHAW, A. A., D'AMICO, A. V. et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1** 203–209. https://doi.org/10.1016/s1535-6108(02)00030-2

[45] STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

[46] STOCK, J. H. and WATSON, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *J. Bus. Econom. Statist.* **20** 147–162. MR1963257 https://doi.org/10.1198/073500102317351921

[47] TARIGAN, B. and VAN DE GEER, S. A. (2006). Classifiers of support vector machine type with l_1 complexity regularization. *Bernoulli* **12** 1045–1076. MR2274857 https://doi.org/10.3150/bj/1165269150

[48] TIBSHIRANI, R., HASTIE, T., NARASIMHAN, B. and CHU, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. USA* **99** 6567–6572.

[49] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

[50] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *Ann. Statist.* **41** 2905–2947. MR3161452 https://doi.org/10.1214/13-AOS1151

[51] WEGKAMP, M. and YUAN, M. (2011). Support vector machines with a reject option. *Bernoulli* **17** 1368–1385. MR2854776 https://doi.org/10.3150/10-BEJ320

[52] WITTEN, D. M. and TIBSHIRANI, R. (2011). Penalized classification using Fisher's linear discriminant. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 753–772. MR2867457 https://doi.org/10.1111/j.1467-9868.2011.00783.x

DISPERSAL DENSITY ESTIMATION ACROSS SCALES

BY MARC HOFFMANN[1,a] AND MATHIAS TRABS[2,b]

[1]*CEREMADE, Université Paris-Dauphine PSL,* [a]*hoffmann@ceremade.dauphine.fr*

[2]*Department of Mathematics, Karlsruhe Institute of Technology,* [b]*trabs@kit.edu*

We consider a space structured population model generated by two-point clouds: a homogeneous Poisson process M with intensity $n \to \infty$ as a model for a parent generation together with a Cox point process N as offspring generation, with conditional intensity given by the convolution of M with a scaled dispersal density $\sigma^{-1} f(\cdot/\sigma)$. Based on a realisation of M and N, we study the nonparametric estimation of f and the estimation of the physical scale parameter $\sigma > 0$ simultaneously for all regimes $\sigma = \sigma_n$. We establish that the optimal rates of convergence do not depend monotonously on the scale and we construct minimax estimators accordingly whether σ is known or considered as a nuisance, in which case we can estimate it and achieve asymptotic minimaxity by plug-in. The statistical reconstruction exhibits a competition between a direct and a deconvolution problem. Our study reveals in particular the existence of a least favorable intermediate inference scale, a phenomenon that seems to be new.

REFERENCES

ARNAUD, J.-F., VIARD, F., DELESCLUSE, M. and CUGUEN, J. (2003). Evidence for gene flow via seed dispersal from crop to wild relatives in Beta vulgaris (Chenopodiaceae): Consequences for the release of genetically modified crop species with weedy lineages. *Proc. R. Soc. Lond., B Biol. Sci.* **270** 1565–1571.

BACCELLI, F., KAUFFMANN, B. and VEITCH, D. (2009). Inverse problems in queueing theory and Internet probing. *Queueing Syst.* **63** 59–107. MR2576007 https://doi.org/10.1007/s11134-009-9150-9

BELOMESTNY, D. and GOLDENSHLUGER, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *Ann. Statist.* **49** 615–649. MR4255101 https://doi.org/10.1214/20-aos1969

BERENFELD, C. and HOFFMANN, M. (2021). Density estimation on an unknown submanifold. *Electron. J. Stat.* **15** 2179–2223. MR4255312 https://doi.org/10.1214/21-ejs1826

BINGHAM, N. H. and DUNHAM, B. (1997). Estimating diffusion coefficients from count data: Einstein–Smoluchowski theory revisited. *Ann. Inst. Statist. Math.* **49** 667–679. MR1621845 https://doi.org/10.1023/A:1003214209227

BLANGHAPS, N., NOV, Y. and WEISS, G. (2013). Sojourn time estimation in an M/G/∞ queue with partial information. *J. Appl. Probab.* **50** 1044–1056. MR3161372 https://doi.org/10.1239/jap/1389370098

BONNET, A., LACOUR, C., PICARD, F. and RIVOIRARD, V. (2022). Uniform deconvolution for Poisson point processes. *J. Mach. Learn. Res.* **23** Paper No. 194. MR4577147

BROWN, M. (1970). An $M/G/\infty$ estimation problem. *Ann. Math. Stat.* **41** 651–654. MR0261722 https://doi.org/10.1214/aoms/1177697107

CARSTENSEN, L., SANDELIN, A., WINTHER, O. and HANSEN, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinform.* **11** 1–19.

CHOROWSKI, J. (2018). Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies. *Bernoulli* **24** 2934–2990. MR3779707 https://doi.org/10.3150/17-BEJ950

COMTE, F. and LACOUR, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 601–627. MR2853732 https://doi.org/10.1111/j.1467-9868.2011.00775.x

DATTNER, I., REISS, M. and TRABS, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** 143–192. MR3449779 https://doi.org/10.3150/14-BEJ626

DELAIGLE, A. and MEISTER, A. (2011). Nonparametric function estimation under Fourier-oscillating noise. *Statist. Sinica* **21** 1065–1092. MR2827515 https://doi.org/10.5705/ss.2009.082

DUVAL, C. and HOFFMANN, M. (2011). Statistical inference across time scales. *Electron. J. Stat.* **5** 2004–2030. MR2870155 https://doi.org/10.1214/11-EJS660

GOLDENSHLUGER, A. (2016). Nonparametric estimation of the service time distribution in the M/G/∞ queue. *Adv. in Appl. Probab.* **48** 1117–1138. MR3595768 https://doi.org/10.1017/apr.2016.67

GOLDENSHLUGER, A. (2018). The $M/G/\infty$ estimation problem revisited. *Bernoulli* **24** 2531–2568. MR3779694 https://doi.org/10.3150/17-BEJ936

GOLDENSHLUGER, A. and JACOBOVIC, R. (2021). Smoluchowski processes and nonparametric estimation of functionals of particle displacement distributions from count data. ArXiv preprint. Available at arXiv:2108.06954.

GOLDENSHLUGER, A. and KOOPS, D. T. (2019). Nonparametric estimation of service time characteristics in infinite-server queues with nonstationary Poisson input. *Stoch. Syst.* **9** 183–207. MR4032680 https://doi.org/10.1287/stsy.2018.0026

GROENEBOOM, P. and JONGBLOED, G. (2003). Density estimation in the uniform deconvolution model. *Stat. Neerl.* **57** 136–157.

GUSTO, G. and SCHBATH, S. (2005). FADO: A statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 24. MR2170440 https://doi.org/10.2202/1544-6115.1119

HALL, P. and PARK, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 861–875. MR2102469 https://doi.org/10.1111/j.1467-9868.2004.B5725.x

HEUERTZ, M., VEKEMANS, X., HAUSMAN, J.-F., PALADA, M. and HARDY, O. J. (2003). Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Mol. Ecol.* **12** 2483–2495. https://doi.org/10.1046/j.1365-294x.2003.01923.x

HOFFMANN, M. and TRABS, M. (2023). Supplement to "Dispersal density estimation across scales." https://doi.org/10.1214/23-AOS2290SUPP

HUNT, X. J., REYNAUD-BOURET, P., RIVOIRARD, V., SANSONNET, L. and WILLETT, R. (2019). A data-dependent weighted LASSO under Poisson noise. *IEEE Trans. Inf. Theory* **65** 1589–1613. MR3923187 https://doi.org/10.1109/TIT.2018.2869578

ISAGI, Y., KANAZASHI, T., SUZUKI, W., TANAKA, H. and ABE, T. (2000). Microsatellite analysis of the regeneration process of Magnolia obovata Thunb. *Heredity* **84** 143–151.

LAVOREL, S., GARDNER, R. H. and O'NEILL, R. V. (1995). Dispersal of annual plants in hierarchically structured landscapes. *Landsc. Ecol.* **10** 277–289.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

MEISTER, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Probl.* **24** 015003. MR2384762 https://doi.org/10.1088/0266-5611/24/1/015003

NEUMANN, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955–1968. MR2396948 https://doi.org/10.1016/j.jmva.2006.09.012

NICKL, R., REISS, M., SÖHL, J. and TRABS, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields* **164** 61–108. MR3449386 https://doi.org/10.1007/s00440-014-0607-3

RANDOL, B. (1969). On the asymptotic behavior of the Fourier transform of the indicator function of a convex set. *Trans. Amer. Math. Soc.* **139** 279–285. MR0251450 https://doi.org/10.2307/1995320

VAN DER VAART, A. (2002). The statistical work of Lucien Le Cam. *Ann. Statist.* **30** 631–682.

MINIMAX RATE OF DISTRIBUTION ESTIMATION ON UNKNOWN SUBMANIFOLDS UNDER ADVERSARIAL LOSSES

BY RONG TANG[1,a] AND YUN YANG[2,b]

[1]*Department of Mathematics, The Hong Kong University of Science and Technology,* [a]*martang@ust.hk*

[2]*Department of Statistics, University of Illinois Urbana-Champaign,* [b]*yy84@illinois.edu*

Statistical inference from high-dimensional data with low-dimensional structures has recently attracted a lot of attention. In machine learning, deep generative modelling approaches implicitly estimate distributions of complex objects by creating new samples from the underlying distribution, and have achieved great success in generating synthetic realistic-looking images and texts. A key step in these approaches is the extraction of latent features or representations (encoding) that can be used for accurately reconstructing the original data (decoding). In other words, low-dimensional manifold structure is implicitly assumed and utilized in the distribution modelling and estimation. To understand the benefit of low-dimensional manifold structure in generative modelling, we build a general minimax framework for distribution estimation on unknown submanifold under adversarial losses, with suitable smoothness assumptions on the target distribution and the manifold. The established minimax rate elucidates how various problem characteristics, including intrinsic dimensionality of the data and smoothness levels of the target distribution and the manifold, affect the fundamental limit of high-dimensional distribution estimation. To prove the minimax upper bound, we construct an estimator based on a mixture of locally fitted generative models, which is motivated by the partition of unity technique from differential geometry and is necessary to cover cases where the underlying data manifold does not admit a global parametrization. We also propose a data-driven adaptive estimator that is shown to simultaneously attain within a logarithmic factor of the optimal rate over a large collection of distribution classes.

REFERENCES

[1] AAMARI, E. and LEVRARD, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204. MR3909931 https://doi.org/10.1214/18-AOS1685

[2] ARJOVSKY, M. and BOTTOU, L. (2017). Towards principled methods for training generative adversarial networks. ArXiv preprint. Available at arXiv:1701.04862.

[3] ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein GAN. ArXiv preprint. Available at arXiv:1701.07875.

[4] AYDIN, O. U., TAHA, A. A., HILBERT, A., KHALIL, A. A., GALINOVIC, I., FIEBACH, J. B., FREY, D. and MADAI, V. I. (2021). On the usage of average Hausdorff distance for segmentation performance assessment: Hidden error when used for ranking. *Eur. Radiol. Exp.*

[5] BERENFELD, C. and HOFFMANN, M. (2021). Density estimation on an unknown submanifold. *Electron. J. Stat.* **15** 2179–2223. MR4255312 https://doi.org/10.1214/21-ejs1826

[6] BIAU, G., CADRE, B., SANGNIER, M. and TANIELIAN, U. (2020). Some theoretical properties of GANs. *Ann. Statist.* **48** 1539–1566. MR4124334 https://doi.org/10.1214/19-AOS1858

[7] BICKEL, P. J. and LI, B. (2007). Local polynomial regression on unknown manifolds. In *Complex Datasets and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 177–186. IMS, Beachwood, OH. MR2459188 https://doi.org/10.1214/074921707000000148

[8] BIŃKOWSKI, M., SUTHERLAND, D. J., ARBEL, M. and GRETTON, A. (2018). Demystifying mmd gans. ArXiv preprint. Available at arXiv:1801.01401.

[9] CAFFARELLI, L. A. (1996). Boundary regularity of maps with convex potentials. II. *Ann. of Math.* (2) **144** 453–496. MR1426885 https://doi.org/10.2307/2118564

[10] CAILLERIE, C., CHAZAL, F., DEDECKER, J. and MICHEL, B. (2011). Deconvolution for the Wasserstein metric and geometric inference. *Electron. J. Stat.* **5** 1394–1423. MR2851684 https://doi.org/10.1214/11-EJS646

[11] CAMASTRA, F. and VINCIARELLI, A. (2002). Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 1404–1407.

[12] CARTER, K. M., RAICH, R. and HERO, A. O. III (2010). On local intrinsic dimension estimation and its applications. *IEEE Trans. Signal Process.* **58** 650–663. MR2750463 https://doi.org/10.1109/TSP.2009.2031722

[13] CHAE, M., KIM, D., KIM, Y. and LIN, L. (2023). A likelihood approach to nonparametric estimation of a singular distribution using deep generative models. *J. Mach. Learn. Res.* **24** Paper No. 77. MR4582499

[14] CHEN, Y., GAO, Q. and WANG, X. (2022). Inferential Wasserstein generative adversarial networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 83–113. MR4400391 https://doi.org/10.1111/rssb.12476

[15] CORNEA, O., LUPTON, G., OPREA, J. and TANRÉ, D. (2003). *Lusternik–Schnirelmann Category. Mathematical Surveys and Monographs* **103**. Amer. Math. Soc., Providence, RI. MR1990857 https://doi.org/10.1090/surv/103

[16] DIVOL, V. (2022). Measure estimation on manifolds: An optimal transport approach. *Probab. Theory Related Fields* **183** 581–647. MR4421180 https://doi.org/10.1007/s00440-022-01118-z

[17] DO CARMO, M. P. (1992). *Riemannian Geometry. Mathematics*: *Theory & Applications*. Birkhäuser, Inc., Boston, MA. MR1138207 https://doi.org/10.1007/978-1-4757-2201-7

[18] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

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

[20] FARAHMAND, A. M., SZEPESVÁRI, C. and AUDIBERT, J.-Y. (2007). Manifold-adaptive dimension estimation. In *Proceedings of the* 24*th International Conference on Machine Learning* 265–272.

[21] FOX, R. H. (1941). On the Lusternik–Schnirelmann category. *Ann. of Math.* (2) **42** 333–370. MR0004108 https://doi.org/10.2307/1968905

[22] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Minimax manifold estimation. *J. Mach. Learn. Res.* **13** 1263–1291. MR2930639

[23] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939 https://doi.org/10.1214/12-AOS994

[24] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2020). Generative adversarial networks. *Commun. ACM* **63** 139–144.

[25] GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

[26] KINGMA, D. P. and WELLING, M. (2013). Auto-encoding variational bayes. ArXiv preprint. Available at arXiv:1312.6114.

[27] KNOP, S., SPUREK, P., TABOR, J., PODOLAK, I., MAZUR, M. and JASTRZĘBSKI, S. (2020). Cramer–Wold auto-encoder. *J. Mach. Learn. Res.* **21** Paper No. 164. MR4209450

[28] LEE, J. M. (2013). Smooth manifolds. In *Introduction to Smooth Manifolds* 1–31. Springer.

[29] LEPSKII, O. V. (1991). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

[30] LEVINA, E. and BICKEL, P. (2004). Maximum likelihood estimation of intrinsic dimension. *Adv. Neural Inf. Process. Syst.* **17**.

[31] LI, C.-L., CHANG, W.-C., CHENG, Y., YANG, Y. and PÓCZOS, B. (2017). Mmd gan: Towards deeper understanding of moment matching network. *Adv. Neural Inf. Process. Syst.* **30**.

[32] LI, Y., SWERSKY, K. and ZEMEL, R. (2015). Generative moment matching networks. In *International Conference on Machine Learning* 1718–1727. PMLR.

[33] LIANG, T. (2021). How well generative adversarial networks learn distributions. *J. Mach. Learn. Res.* **22** Paper No. 228. MR4329807

[34] LITTLE, A. V., LEE, J., JUNG, Y.-M. and MAGGIONI, M. (2009). Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD. In 2009 *IEEE/SP* 15*th Workshop on Statistical Signal Processing* 85–88. IEEE Press, New York.

[35] LOADER, C. (2006). *Local Regression and Likelihood*. Springer Science & Business Media.

[36] LUI, Y. M. (2012). Advances in matrix manifolds for computer vision. *Image Vis. Comput.* **30** 380–388.

[37] MROUEH, Y., LI, C.-L., SERCU, T., RAJ, A. and CHENG, Y. (2017). Sobolev GAN. ArXiv preprint. Available at arXiv:1711.04894.

[38] MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938 https://doi.org/10.2307/1428011

[39] NILES-WEED, J. and BERTHET, Q. (2022). Minimax estimation of smooth densities in Wasserstein distance. *Ann. Statist.* **50** 1519–1540. MR4441130 https://doi.org/10.1214/21-aos2161

[40] OZAKIN, A. and GRAY, A. (2009). Submanifold density estimation. In *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta, eds.) **22**. Curran Associates, Red Hook.

[41] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[42] SINGH, S., LI, B., LI, C.-L., ZAHEER, M. and PÓCZOS, B. (2018). Nonparametric density estimation under adversarial losses. *Adv. Neural Inf. Process. Syst.* **31**.

[43] STONE, M. H. (1948). The generalized Weierstrass approximation theorem. *Math. Mag.* **21** 167–184. MR0027121 https://doi.org/10.2307/3029750

[44] TANG, R. and YANG, Y. (2023). Supplement to "Minimax rate of distribution estimation on unknown submanifolds under adversarial losses." https://doi.org/10.1214/23-AOS2291SUPP

[45] TERRADOT, L., DURNELL, N., LI, M., LI, M., ORY, J., LABIGNE, A., LEGRAIN, P., COLLAND, F. and WAKSMAN, G. (2004). Biochemical characterization of protein complexes from the Helicobacter pylori protein interaction map: Strategies for complex formation and evidence for novel interactions within type IV secretion systems. *Mol. Cell. Proteomics* **3** 809–819.

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

[47] TOLSTIKHIN, I., BOUSQUET, O., GELLY, S. and SCHOELKOPF, B. (2017). Wasserstein auto-encoders. ArXiv preprint. Available at arXiv:1711.01558.

[48] TOLSTIKHIN, I., SRIPERUMBUDUR, B. K. and MUANDET, K. (2017). Minimax estimation of kernel mean embeddings. *J. Mach. Learn. Res.* **18** Paper No. 86. MR3714249

[49] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[50] UPPAL, A., SINGH, S. and PÓCZOS, B. (2019). Nonparametric density estimation & convergence rates for gans under Besov ipm losses. *Adv. Neural Inf. Process. Syst.* **32**.

[51] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

[53] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

[54] YANG, Y. and DUNSON, D. B. (2016). Bayesian manifold regression. *Ann. Statist.* **44** 876–905. MR3476620 https://doi.org/10.1214/15-AOS1390

[55] YOU, Z.-H., LEI, Y.-K., GUI, J., HUANG, D.-S. and ZHOU, X. (2010). Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data. *Bioinformatics* **26** 2744–2751.

[56] YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

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

[58] ZHAO, S., SONG, J. and ERMON, S. (2018). InfoVAE: Information maximizing variational autoencoders. ArXiv preprint. Available at arXiv:1706.02262.

INFERENCE FOR LOW-RANK MODELS

BY VICTOR CHERNOZHUKOV[1,a], CHRISTIAN HANSEN[2,b], YUAN LIAO[3,c] AND
YINCHU ZHU[4,d]

[1]*Department of Economics, Massachusetts Institute of Technology,* [a]*vchern@mit.edu*
[2]*Booth School of Business, University of Chicago,* [b]*Christian.Hansen@chicagobooth.edu*
[3]*Department of Economics, Rutgers University,* [c]*yuan.liao@rutgers.edu*
[4]*Department of Economics, Brandeis University,* [d]*yinchuzhu@brandeis.edu*

This paper studies inference in linear models with a high-dimensional parameter matrix that can be well approximated by a "spiked low-rank matrix." A spiked low-rank matrix has rank that grows slowly compared to its dimensions and nonzero singular values that diverge to infinity. We show that this framework covers a broad class of models of latent variables, which can accommodate matrix completion problems, factor models, varying coefficient models and heterogeneous treatment effects. For inference, we apply a procedure that relies on an initial nuclear-norm penalized estimation step followed by two ordinary least squares regressions. We consider the framework of estimating incoherent eigenvectors and use a rotation argument to argue that the eigenspace estimation is asymptotically unbiased. Using this framework, we show that our procedure provides asymptotically normal inference and achieves the semiparametric efficiency bound. We illustrate our framework by providing low-level conditions for its application in a treatment effects context where treatment assignment might be strongly dependent.

REFERENCES

ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2021). Matrix completion methods for causal panel data models. *J. Amer. Statist. Assoc.* **116** 1716–1730. MR4353709 https://doi.org/10.1080/01621459.2021.1891924

BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803 https://doi.org/10.1214/12-AOS1014

CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 https://doi.org/10.1214/13-AOS1178

CANDÈS, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM J. Optim.* **30** 3098–3121. MR4167625 https://doi.org/10.1137/19M1290000

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHERNOZHUKOV, V., HANSEN, C., LIAO, Y. and ZHU, Y. (2018). Inference for heterogeneous effects using low-rank estimations. ArXiv preprint. Available at arXiv:1812.08089.

CHERNOZHUKOV, V., HANSEN, C., LIAO, Y. and ZHU, Y. (2023). Supplement to "Inference for low-rank models." https://doi.org/10.1214/23-AOS2293SUPP

CHERNOZHUKOV, V., HANSEN, C. and SPINDLER, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Ann. Rev. Econ.* **7** 649–688.

DRAY, S. and JOSSE, J. (2015). Principal component analysis with missing values: A comparative survey of methods. *Plant Ecol.* **216** 657–667.

FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* **117** 852–861. MR4436317 https://doi.org/10.1080/01621459.2020.1825448

GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999

HASTIE, T., MAZUMDER, R., LEE, J. D. and ZADEH, R. (2015). Matrix completion and low-rank SVD via fast alternating least squares. *J. Mach. Learn. Res.* **16** 3367–3402. MR3450542

IWAKURA, H. and OKUI, R. (2014). Asymptotic efficiency in factor models and dynamic panel data models. Available at SSRN 2395722.

JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the* 2013 *ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693

JANKOVÁ, J. and VAN DE GEER, S. (2018). Semiparametric efficiency bounds for high-dimensional models. *Ann. Statist.* **46** 2336–2359. MR3845020 https://doi.org/10.1214/17-AOS1622

JANKOVÁ, J. and VAN DE GEER, S. (2021). De-biased sparse PCA: Inference for eigenstructure of large covariance matrices. *IEEE Trans. Inf. Theory* **67** 2507–2527. MR4282369 https://doi.org/10.1109/TIT.2021.3059765

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583 https://doi.org/10.3150/12-BEJ486

KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2020). Efficient estimation of linear functionals of principal components. *Ann. Statist.* **48** 464–490. MR4065170 https://doi.org/10.1214/19-AOS1816

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 https://doi.org/10.1214/11-AOS894

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics*: *The Harald Cramér Volume* (U. Grenander, ed.) 213–234. Almqvist & Wiksell, Stockholm. MR0112201

ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

SUN, R. and LUO, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inf. Theory* **62** 6535–6579. MR3565131 https://doi.org/10.1109/TIT.2016.2598574

SUN, T. and ZHANG, C.-H. (2012). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *Adv. Neural Inf. Process. Syst.* 863–871.

XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 58–77. MR4220984 https://doi.org/10.1111/rssb.12400

ZHU, Z., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional principal component analysis with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2000–2031. MR4515564

EXTREME VALUE INFERENCE FOR HETEROGENEOUS POWER LAW DATA

BY JOHN H.J. EINMAHL[1,a] AND YI HE[2,b]

[1]*Department of Econometrics & OR and CentER, Tilburg University,* [a]*j.h.j.einmahl@uvt.nl*
[2]*Amsterdam School of Economics, University of Amsterdam,* [b]*yi.he@uva.nl*

We extend extreme value statistics to independent data with possibly very different distributions. In particular, we present novel asymptotic normality results for the Hill estimator, which now estimates the extreme value index of the average distribution. Due to the heterogeneity, the asymptotic variance can be substantially smaller than that in the i.i.d. case. As a special case, we consider a heterogeneous scales model where the asymptotic variance can be calculated explicitly. The primary tool for the proofs is the functional central limit theorem for a weighted tail empirical process. We also present asymptotic normality results for the extreme quantile estimator. A simulation study shows the good finite-sample behavior of our limit theorems. We also present applications to assess the tail heaviness of earthquake energies and of cross-sectional stock market losses.

REFERENCES

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer, London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 393–442. MR1086795

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L. and ZHOU, C. (2021). Trends in extreme value indices. *J. Amer. Statist. Assoc.* **116** 1265–1279. MR4309271 https://doi.org/10.1080/01621459.2019.1705307

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855. MR1026315 https://doi.org/10.1214/aos/1176347397

DREES, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875 https://doi.org/10.1214/aoap/1019487617

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886 https://doi.org/10.1214/09-AOS788

EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

EINMAHL, J. H. J. and HE, Y. (2023). Extreme value estimation for heterogeneous data. *J. Bus. Econom. Statist.* **41** 255–269. MR4522166 https://doi.org/10.1080/07350015.2021.2008408

EINMAHL, J. H and HE, Y. (2023). Supplement to "Extreme value inference for heterogeneous power law data." https://doi.org/10.1214/23-AOS2294SUPP

EINMAHL, J. H. J. and SEGERS, J. (2021). Empirical tail copulas for functional data. *Ann. Statist.* **49** 2672–2696. MR4338379 https://doi.org/10.1214/21-aos2050

GABAIX, X., GOPIKRISHNAN, P., PLEROU, V. and STANLEY, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* **423** 267–270. https://doi.org/10.1038/nature01624

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

HSING, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569. MR1126337 https://doi.org/10.1214/aos/1176348261

JESSEN, A. H. and MIKOSCH, T. (2006). Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80** 171–192. MR2281913 https://doi.org/10.2298/PIM0694171J

JI, P., JIN, J., KE, Z. T. and LI, W. (2022). Co-citation and co-authorship networks of statisticians. *J. Bus. Econom. Statist.* **40** 469–485. MR4410871 https://doi.org/10.1080/07350015.2021.1978469

KULIK, R. and SOULIER, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

POTTER, H. SA. (1942). The mean values of certain Dirichlet series, II. *Proc. Lond. Math. Soc.* **2** 1–19.

SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207. MR0902252 https://doi.org/10.1214/aos/1176350499

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VERVAAT, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Gebiete* **23** 245–253. MR0321164 https://doi.org/10.1007/BF00532510

FINITE-SAMPLE COMPLEXITY OF SEQUENTIAL MONTE CARLO ESTIMATORS

BY JOE MARION[a], JOSEPH MATHEWS[b] AND SCOTT C. SCHMIDLER[c]

Department of Statistical Science, Duke University, [a]*joseph@berryconsultants.net,* [b]*joseph.mathews@duke.edu,*
[c]*schmidler@stat.duke.edu*

We present bounds for the finite-sample error of sequential Monte Carlo samplers on static spaces. Our approach explicitly relates the performance of the algorithm to properties of the chosen sequence of distributions and mixing properties of the associated Markov kernels. This allows us to give the first finite-sample comparison to other Monte Carlo schemes. We obtain bounds for the complexity of sequential Monte Carlo approximations for a variety of target distributions such as finite spaces, product measures and log-concave distributions including Bayesian logistic regression. The bounds obtained are within a logarithmic factor of similar bounds obtainable for Markov chain Monte Carlo.

REFERENCES

[1] BESKOS, A., CRISAN, D. and JASRA, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* **24** 1396–1445. MR3211000 https://doi.org/10.1214/13-AAP951

[2] BESKOS, A., JASRA, A., KANTAS, N. and THIERY, A. (2016). On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.* **26** 1111–1146. MR3476634 https://doi.org/10.1214/15-AAP1113

[3] CÉROU, F., DEL MORAL, P., FURON, T. and GUYADER, A. (2012). Sequential Monte Carlo for rare event estimation. *Stat. Comput.* **22** 795–808. MR2909622 https://doi.org/10.1007/s11222-011-9231-6

[4] CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068 https://doi.org/10.1214/10-AIHP358

[5] CHOPIN, N. (2002). A sequential particle filter method for static models. *Biometrika* **89** 539–551. MR1929161 https://doi.org/10.1093/biomet/89.3.539

[6] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989 https://doi.org/10.1214/009053604000000698

[7] CRISAN, D. (2001). Particle filters—a theoretical perspective. In *Sequential Monte Carlo Methods in Practice*. *Stat. Eng. Inf. Sci.* 17–41. Springer, New York. MR1847785

[8] DALALYAN, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[9] DEL MORAL, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

[10] DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333 https://doi.org/10.1111/j.1467-9868.2006.00553.x

[11] DEL MORAL, P. and GUIONNET, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9** 275–297. MR1687359 https://doi.org/10.1214/aoap/1029962742

[12] DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV. Lecture Notes in Math.* **1729** 1–145. Springer, Berlin. MR1768060 https://doi.org/10.1007/BFb0103798

[13] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190 https://doi.org/10.1214/07-AOS514

[14] DURHAM, G. and GEWEKE, J. (2014). Adaptive sequential posterior simulators for massively parallel computing environments. In *Advances in Econometrics* **34** 1–44. Emerald Group Publishing, Limited.

[15] DWIVEDI, R., CHEN, Y., WAINWRIGHT, M. J. and YU, B. (2019). Log-concave sampling: Metropolis–Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183, 42. MR4048994

[16] EBERLE, A. and MARINELLI, C. (2007). Stability of sequential Markov chain Monte Carlo methods. In *Conference Oxford sur les Méthodes de Monte Carlo Séquentielles. ESAIM Proc.* **19** 22–31. EDP Sci., Les Ulis. MR2405646 https://doi.org/10.1051/proc:071905

[17] EBERLE, A. and MARINELLI, C. (2013). Quantitative approximations of evolving probability measures and sequential Markov chain Monte Carlo methods. *Probab. Theory Related Fields* **155** 665–701. MR3034790 https://doi.org/10.1007/s00440-012-0410-y

[18] JERRUM, M. R., VALIANT, L. G. and VAZIRANI, V. V. (1986). Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* **43** 169–188. MR0855970 https://doi.org/10.1016/0304-3975(86)90174-X

[19] KANNAN, R. and LI, G. (1996). Sampling according to the multivariate normal density. In *37th Annual Symposium on Foundations of Computer Science* (*Burlington*, *VT*, 1996) 204–212. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1450618 https://doi.org/10.1109/SFCS.1996.548479

[20] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200 https://doi.org/10.1002/(SICI)1098-2418(199708)11:1<1::AID-RSA1>3.0.CO;2-X

[21] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077 https://doi.org/10.1214/009053605000000426

[22] LE GLAND, F. and OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* **14** 144–187. MR2023019 https://doi.org/10.1214/aoap/1075828050

[23] LEE, A. and WHITELEY, N. (2016). Forest resampling for distributed sequential Monte Carlo. *Stat. Anal. Data Min.* **9** 230–248. MR3529396 https://doi.org/10.1002/sam.11280

[24] LEE, A., YAU, C., GILES, M. B., DOUCET, A. and HOLMES, C. C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *J. Comput. Graph. Statist.* **19** 769–789.

[25] LEE, Y. T., SHEN, R. and TIAN, K. (2020). Logsmooth gradient concentration and tighter runtimes for metropolized Hamiltonian Monte Carlo. In *Conference on Learning Theory, COLT* 2020, 9–12 *July* 2020, *Virtual Event* [*Graz, Austria*] (J. D. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 9–12. PMLR.

[26] LOVÁSZ, L. and VEMPALA, S. (2004). Hit-and-run from a corner. In *Proceedings of the* 36*th Annual ACM Symposium on Theory of Computing* 310–314. ACM, New York. MR2121613 https://doi.org/10.1145/1007352.1007403

[27] LOVÁSZ, L. and VEMPALA, S. (2006). Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *2006* 47*th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS'06*) 57–68.

[28] LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621 https://doi.org/10.1002/rsa.20135

[29] MOTWANI, R. and RAGHAVAN, P. (1995). *Randomized Algorithms*. Cambridge Univ. Press, Cambridge. MR1344451 https://doi.org/10.1017/CBO9780511814075

[30] NEAL, R. M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132 https://doi.org/10.1023/A:1008923215028

[31] PARK, S. and PANDE, V. S. (2007). Choosing weights for simulated tempering. *Phys. Rev. E* **76** 016703.

[32] PAULIN, D., JASRA, A. and THIERY, A. (2019). Error bounds for sequential Monte Carlo samplers for multimodal distributions. *Bernoulli* **25** 310–340. MR3892321 https://doi.org/10.3150/17-bej988

[33] ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565 https://doi.org/10.1214/154957804100000024

[34] SCHWEIZER, N. (2011). Non-asymptotic error bounds for sequential MCMC methods. Ph.D. thesis, Univ. Bonn.

[35] VANDERWERKEN, D. N. and SCHMIDLER, S. C. (2013). Parallel Markov chain Monte Carlo. ArXiv e-prints.

[36] VEMPALA, S. (2005). Geometric random walks: A survey. In *Combinatorial and Computational Geometry. Math. Sci. Res. Inst. Publ.* **52** 577–616. Cambridge Univ. Press, Cambridge. MR2178341

[37] VERGÉ, C., DUBARRY, C., DEL MORAL, P. and MOULINES, E. (2015). On parallel implementation of sequential Monte Carlo methods: The island particle model. *Stat. Comput.* **25** 243–260. MR3306704 https://doi.org/10.1007/s11222-013-9429-x

[38] WHITELEY, N. (2012). Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. *Stoch. Anal. Appl.* **30** 774–798. MR2966098 https://doi.org/10.1080/07362994.2012.684323

[39] WOODARD, D. B., SCHMIDLER, S. C. and HUBER, M. (2009). Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electron. J. Probab.* **14** 780–804. MR2495560 https://doi.org/10.1214/EJP.v14-638

[40] WOODARD, D. B., SCHMIDLER, S. C. and HUBER, M. (2009). Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *Ann. Appl. Probab.* **19** 617–640. MR2521882 https://doi.org/10.1214/08-AAP555

[41] WU, K., SCHMIDLER, S. and CHEN, Y. (2022). Minimax mixing time of the Metropolis-adjusted Langevin algorithm for log-concave sampling. *J. Mach. Learn. Res.* **23** Paper No. [270], 63. MR4577709

[42] ZHOU, Y., JOHANSEN, A. M. and ASTON, J. A. D. (2016). Toward automatic model comparison: An adaptive sequential Monte Carlo approach. *J. Comput. Graph. Statist.* **25** 701–726. MR3533634 https://doi.org/10.1080/10618600.2015.1060885

COVERAGE OF CREDIBLE INTERVALS IN BAYESIAN MULTIVARIATE ISOTONIC REGRESSION

BY KANG WANG[a] AND SUBHASHIS GHOSAL[b]

Department of Statistics, North Carolina State University, [a]*kwang22@ncsu.edu,* [b]*sghosal@stat.ncsu.edu*

We consider the nonparametric multivariate isotonic regression problem, where the regression function is assumed to be nondecreasing with respect to each predictor. Our goal is to construct a Bayesian credible interval for the function value at a given interior point with assured limiting frequentist coverage. A natural prior on the regression function is given by a random step function with a suitable prior on increasing step-heights, but the resulting posterior distribution is hard to analyze theoretically due to the complicated order restriction on the coefficients. We instead put a prior on unrestricted step-functions, but make inference using the induced posterior measure by an "immersion map" from the space of unrestricted functions to that of multivariate monotone functions. This allows for maintaining the natural conjugacy for posterior sampling. A natural immersion map to use is a projection with respect to a distance function, but in the present context, a block isotonization map is found to be more useful. The approach of using the induced "immersion posterior" measure instead of the original posterior to make inference provides a useful extension of the Bayesian paradigm, particularly helpful when the model space is restricted by some complex relations. We establish a key weak convergence result for the posterior distribution of the function at a point in terms of some functional of a multiindexed Gaussian process that leads to an expression for the limiting coverage of the Bayesian credible interval. Analogous to a recent result for univariate monotone functions, we find that the limiting coverage is slightly higher than the credibility, the opposite of a phenomenon observed in smoothing problems. Interestingly, the relation between credibility and limiting coverage does not involve any unknown parameter. Hence, by a recalibration procedure, we can get a predetermined asymptotic coverage by choosing a suitable credibility level smaller than the targeted coverage, and thus also shorten the credible intervals.

REFERENCES

[1] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895 https://doi.org/10.1214/aoms/1177728423

[2] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[3] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743 https://doi.org/10.1214/aos/1015345959

[4] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression. Wiley Series in Probability and Mathematical Statistics.* Wiley, London–Sydney. MR0326887

[5] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

[6] BHAUMIK, P. and GHOSAL, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.* **9** 3124–3154. MR3453972 https://doi.org/10.1214/15-EJS1099

[7] BHAUMIK, P. and GHOSAL, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models. *Bernoulli* **23** 3537–3570. MR3654815 https://doi.org/10.3150/16-BEJ856

[8] BHAUMIK, P. and GHOSAL, S. (2017). Bayesian inference for higher-order ordinary differential equation models. *J. Multivariate Anal.* **157** 103–114. MR3641739 https://doi.org/10.1016/j.jmva.2017.03.003

[9] BHAUMIK, P., SHI, W. and GHOSAL, S. (2022). Two-step Bayesian methods for generalized regression driven by partial differential equations. *Bernoulli* **28** 1625–1647. MR4411505 https://doi.org/10.3150/21-bej1363

[10] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420

[11] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos.*, *Indiana Univ.*, *Bloomington*, *Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

[12] CAI, B. and DUNSON, D. B. (2007). Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies. *J. Amer. Statist. Assoc.* **102** 1158–1171. MR2412540 https://doi.org/10.1198/016214506000000942

[13] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

[14] CHAKRABORTY, M. and GHOSAL, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Ann. Statist.* **49** 1011–1028. MR4255117 https://doi.org/10.1214/20-aos1989

[15] CHAKRABORTY, M. and GHOSAL, S. (2021). Convergence rates for Bayesian estimation and testing in monotone regression. *Electron. J. Stat.* **15** 3478–3503. MR4280172 https://doi.org/10.1214/21-ejs1861

[16] CHAKRABORTY, M. and GHOSAL, S. (2021). Bayesian inference on monotone regression quantile: Coverage and rate acceleration. Preprint.

[17] CHAKRABORTY, M. and GHOSAL, S. (2022). Rates and coverage for monotone densities using projection-posterior. *Bernoulli* **28** 1093–1119. MR4388931 https://doi.org/10.3150/21-bej1379

[18] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324

[19] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865

[20] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525 https://doi.org/10.1214/aos/1176349157

[21] DENG, H., HAN, Q. and ZHANG, C.-H. (2021). Confidence intervals for multiple isotonic regression and other monotone models. *Ann. Statist.* **49** 2021–2052. MR4319240 https://doi.org/10.1214/20-aos2025

[22] DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann. Statist.* **48** 3672–3698. MR4185824 https://doi.org/10.1214/20-AOS1947

[23] DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491 https://doi.org/10.3150/bj/1065444812

[24] DÜMBGEN, L. and JOHNS, R. B. (2004). Confidence bands for isotonic median curves using sign tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998 https://doi.org/10.1198/1061860043506

[25] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961 https://doi.org/10.1214/aos/996986504

[26] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497

[27] DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036 https://doi.org/10.1214/12-AOS1015

[28] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2020). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Trans. Inf. Theory* **66** 6389–6402. MR4173546 https://doi.org/10.1109/TIT.2020.3013390

[29] GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082. MR1810919 https://doi.org/10.1214/aos/1015956707

[30] GIJBELS, I., HALL, P., JONES, M. C. and KOCH, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** 663–673. MR1789816 https://doi.org/10.1093/biomet/87.3.663

[31] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

[32] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

[33] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

[34] GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875 https://doi.org/10.1214/15-AOS1335

[35] HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902 https://doi.org/10.1214/aos/1016120363

[36] HAN, Q. (2021). Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* **49** 2642–2671. MR4338378 https://doi.org/10.1214/21-aos2049

[37] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753

[38] HAN, Q. and ZHANG, C.-H. (2020). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist.* **48** 3251–3282. MR4185808 https://doi.org/10.1214/19-AOS1928

[39] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

[40] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[41] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

[42] KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. MR2211085 https://doi.org/10.1214/009053605000000462

[43] LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317. MR3215349 https://doi.org/10.1093/biomet/ast063

[44] NEELON, B. and DUNSON, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60** 398–406. MR2066274 https://doi.org/10.1111/j.0006-341X.2004.00184.x

[45] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

[46] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[47] SAARELA, O. and ARJAS, E. (2011). A method for Bayesian monotonic multiple regression. *Scand. J. Stat.* **38** 499–513. MR2833843 https://doi.org/10.1111/j.1467-9469.2010.00716.x

[48] SALOMOND, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8** 1380–1404. MR3263126 https://doi.org/10.1214/14-EJS929

[49] SALOMOND, J.-B. (2014). Adaptive Bayes test for monotonicity. In *The Contribution of Young Researchers to Bayesian Statistics*. *Springer Proc. Math. Stat.* **63** 29–33. Springer, Cham. MR3133254 https://doi.org/10.1007/978-3-319-02084-6_7

[50] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812 https://doi.org/10.1214/13-AOS1089

[51] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777

[52] SHIVELY, T. S., SAGER, T. W. and WALKER, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 159–175. MR2655528 https://doi.org/10.1111/j.1467-9868.2008.00677.x

[53] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[54] WANG, K. and GHOSAL, S. (2023). Supplement to "Coverage of credible intervals in Bayesian multivariate isotonic regression." https://doi.org/10.1214/23-AOS2298SUPP

[55] WANG, X. (2008). Bayesian free-knot monotone cubic spline regression. *J. Comput. Graph. Statist.* **17** 373–387. MR2421700 https://doi.org/10.1198/106186008X321077

[56] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

[57] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

UNIVERSAL REGRESSION WITH ADVERSARIAL RESPONSES

BY MOÏSE BLANCHARD[1,a] AND PATRICK JAILLET[2,b]

[1]*Operations Research Center, Massachusetts Institute of Technology,* [a]*moiseb@mit.edu*

[2]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,* [b]*jaillet@mit.edu*

We provide algorithms for regression with adversarial responses under large classes of non-i.i.d. instance sequences, on general separable metric spaces, with *provably minimal* assumptions. We also give characterizations of learnability in this regression context. We consider *universal consistency*, which asks for strong consistency of a learner without restrictions on the value responses. Our analysis shows that such an objective is achievable for a significantly larger class of instance sequences than stationary processes, and unveils a fundamental dichotomy between value spaces: whether finite-horizon mean estimation is achievable or not. We further provide *optimistically universal* learning rules, that is, such that if they fail to achieve universal consistency, any other algorithms will fail as well. For unbounded losses, we propose a mild integrability condition under which there exist algorithms for adversarial regression under large classes of non-i.i.d. instance sequences. In addition, our analysis also provides a learning rule for mean estimation in general metric spaces that is consistent under adversarial responses without any moment conditions on the sequence, a result of independent interest.

REFERENCES

[1] BEN-DAVID, S., PÁL, D. and SHALEV-SHWARTZ, S. (2009). Agnostic online learning. In *Conference on Learning Theory* **3** 1.

[2] BLANCHARD, M. (2022). Universal online learning: An optimistically universal learning rule. In *Conference on Learning Theory* 1077–1125. PMLR.

[3] BLANCHARD, M. and COSSON, R. (2022). Universal online learning with bounded loss: Reduction to binary classification. In *Conference on Learning Theory* 479–495. PMLR.

[4] BLANCHARD, M., COSSON, R. and HANNEKE, S. (2022). Universal online learning with unbounded losses: Memory is all you need. In *Algorithmic Learning Theory* 2022. *Proc. Mach. Learn. Res.* (*PMLR*) **167** 107–127. Proceedings of Machine Learning Research PMLR. MR4461151

[5] BLANCHARD, M. and JAILLET, P. (2023). Supplement to "Universal regression with adversarial responses." https://doi.org/10.1214/23-AOS2299SUPP

[6] BUBECK, S. and CESA-BIANCHI, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Mach. Learn.* **5** 1–122.

[7] CESA-BIANCHI, N., FREUND, Y., HAUSSLER, D., HELMBOLD, D. P., SCHAPIRE, R. E. and WARMUTH, M. K. (1997). How to use expert advice. *J. ACM* **44** 427–485. MR1470152 https://doi.org/10.1145/258128.258179

[8] CESA-BIANCHI, N. and LUGOSI, G. (2006). *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge. MR2409394 https://doi.org/10.1017/CBO9780511546921

[9] CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306. MR0981451 https://doi.org/10.1214/aos/1176347017

[10] COHEN, D. T. and KONTOROVICH, A. (2022). Learning with metric losses. In *Conference on Learning Theory* 662–700. PMLR.

[11] DAVIS, B. C., FLETCHER, P. T., BULLITT, E. and JOSHI, S. (2010). Population shape regression from random design data. *Int. J. Comput. Vis.* **90** 255–266.

[12] DEVROYE, L., GYÖRFI, L., KRZYŻAK, A. and LUGOSI, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Statist.* **22** 1371–1385. MR1311980 https://doi.org/10.1214/aos/1176325633

[13] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

[14] EVANS, S. N. and JAFFE, A. Q. (2020). Strong laws of large numbers for Fréchet means. arXiv preprint arXiv:2012.12859.

[15] FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

[16] 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 https://doi.org/10.1006/jcss.1997.1504

[17] GRAY, R. M. (2009). *Probability, Random Processes, and Ergodic Properties*, 2nd ed. Springer, Dordrecht. MR2840299 https://doi.org/10.1007/978-1-4419-1090-5

[18] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848

[19] GYÖRFI, L., LUGOSI, G. and MORVAI, G. (1999). A simple randomized algorithm for sequential prediction of ergodic time series. *IEEE Trans. Inf. Theory* **45** 2642–2650. MR1725166 https://doi.org/10.1109/18.796420

[20] GYÖRFI, L. and OTTUCSÁK, G. (2007). Sequential prediction of unbounded stationary time series. *IEEE Trans. Inf. Theory* **53** 1866–1872. MR2317147 https://doi.org/10.1109/TIT.2007.894660

[21] GYÖRFI, L. and WEISS, R. (2021). Universal consistency and rates of convergence of multiclass prototype algorithms in metric spaces. *J. Mach. Learn. Res.* **22** Paper No. 151, 25. MR4318507 https://doi.org/10.22405/2226-8383-2021-22-1-133-151

[22] HANNEKE, S. (2021). Open problem: Is there an online learning algorithm that learns whenever online learning is possible? In *Conference on Learning Theory* 4642–4646. PMLR.

[23] HANNEKE, S. (2021). Learning whenever learning is possible: Universal learning under general stochastic processes. *J. Mach. Learn. Res.* **22** Paper No. 130, 116. MR4318486 https://doi.org/10.1093/biostatistics/kxz020

[24] HANNEKE, S. (2022). Universally consistent online learning with arbitrarily dependent responses. In *Algorithmic Learning Theory* 2022. *Proc. Mach. Learn. Res.* (*PMLR*) **167** 488–497. Proceedings of Machine Learning Research PMLR. MR4461166

[25] HANNEKE, S., KONTOROVICH, A., SABATO, S. and WEISS, R. (2021). Universal Bayes consistency in metric spaces. *Ann. Statist.* **49** 2129–2150. MR4319244 https://doi.org/10.1214/20-aos2029

[26] JAFFE, A. Q. (2022). Strong consistency for a class of adaptive clustering procedures. arXiv preprint arXiv:2202.13423.

[27] LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. MR4352539 https://doi.org/10.1214/21-aos2095

[28] LITTLESTONE, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Mach. Learn.* **2** 285–318.

[29] LITTLESTONE, N. and WARMUTH, M. K. (1994). The weighted majority algorithm. *Inform. and Comput.* **108** 212–261. MR1265851 https://doi.org/10.1006/inco.1994.1009

[30] MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. Revised reprint of ıt Statistics of directional data by Mardia [MR0336854 (49 #1627)]. MR1828667

[31] MORVAI, G., KULKARNI, S. R. and NOBEL, A. B. (1999). Regression estimation from an individual stable sequence. *Statistics* **33** 99–118. MR1744748 https://doi.org/10.1080/02331889908802686

[32] MORVAI, G., YAKOWITZ, S. and GYÖRFI, L. (1996). Nonparametric inference for ergodic, stationary time series. *Ann. Statist.* **24** 370–379. MR1389896 https://doi.org/10.1214/aos/1033066215

[33] RAKHLIN, A., SRIDHARAN, K. and TEWARI, A. (2015). Online learning via sequential complexities. *J. Mach. Learn. Res.* **16** 155–186. MR3333006

[34] SCHÖTZ, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics* **56** 34–52. MR4392146 https://doi.org/10.1080/02331888.2022.2032063

[35] SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. and ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 192–199. Springer.

[36] SLIVKINS, A. (2019). Introduction to multi-armed bandits. arXiv preprint arXiv:1904.07272.

[37] STEINWART, I., HUSH, D. and SCOVEL, C. (2009). Learning from dependent observations. *J. Multivariate Anal.* **100** 175–194. MR2460486 https://doi.org/10.1016/j.jmva.2008.04.001

[38] STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. With discussion and a reply by the author. MR0443204

[39] ZAICHYK, H., BIESS, A., KONTOROVICH, A. and MAKARYCHEV, Y. (2019). Efficient Kirszbraun extension with applications to regression. arXiv preprint arXiv:1905.11930.

ASYMPTOTIC NORMALITY FOR EIGENVALUE STATISTICS OF A GENERAL SAMPLE COVARIANCE MATRIX WHEN $p/n \to \infty$ AND APPLICATIONS

BY JIAXIN QIU[1,a] , ZENG LI[2,b] AND JIANFENG YAO[3,c]

[1]*Department of Statistics and Actuarial Science, The University of Hong Kong,* [a]*qiujx@connect.hku.hk*

[2]*Department of Statistics and Data Science, Southern University of Science and Technology,* [b]*liz9@sustech.edu.cn*

[3]*School of Data Science, The Chinese University of Hong Kong (Shenzhen),* [c]*jeffyao@cuhk.edu.cn*

The asymptotic normality for a large family of eigenvalue statistics of a general sample covariance matrix is derived under the ultrahigh-dimensional setting, that is, when the dimension to sample size ratio $p/n \to \infty$. Based on this CLT result, we extend the covariance matrix test problem to the new ultra-high-dimensional context, and apply it to test a matrix-valued white noise. Simulation experiments are conducted for the investigation of finite-sample properties of the general asymptotic normality of eigenvalue statistics, as well as the two developed tests.

REFERENCES

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. MR2189081 https://doi.org/10.3150/bj/1137421640

BAI, Z. D. and YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

BAO, Z. (2015). On asymptotic expansion and central limit theorem of linear eigenvalue statistics for sample covariance matrices when $N/M \to 0$. *Theory Probab. Appl.* **59** 185–207. MR3416046 https://doi.org/10.1137/S0040585X97T987089

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

CHEN, B. and PAN, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. MR3338658 https://doi.org/10.3150/14-BEJ599

CHEN, B. B. and PAN, G. M. (2012). Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli* **18** 1405–1420. MR2995802 https://doi.org/10.3150/11-BEJ381

CHEN, R., XIAO, H. and YANG, D. (2021). Autoregressive models for matrix-valued time series. *J. Econometrics* **222** 539–560. MR4234832 https://doi.org/10.1016/j.jeconom.2020.07.015

COUILLET, R. and DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge. MR2884783 https://doi.org/10.1017/CBO9780511994746

KHORUNZHY, A. M., KHORUZHENKO, B. A. and PASTUR, L. A. (1996). Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37** 5033–5060. MR1411619 https://doi.org/10.1063/1.531589

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

LI, Z. and YAO, J. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electron. J. Stat.* **10** 2973–3010. MR3567239 https://doi.org/10.1214/16-EJS1199

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

QIU, J., LI, Z. and YAO, J. (2023). Supplement to "Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when $p/n \to \infty$ and applications." https://doi.org/10.1214/23-AOS2300SUPP

WANG, L. and PAUL, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. MR3173080 https://doi.org/10.1016/j.jmva.2013.12.015

WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916 https://doi.org/10.1214/13-EJS842

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

ims

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS

ims Textbooks

Exponential
Families in
Theory and
Practice

Bradley Efron

CAMBRIDGE
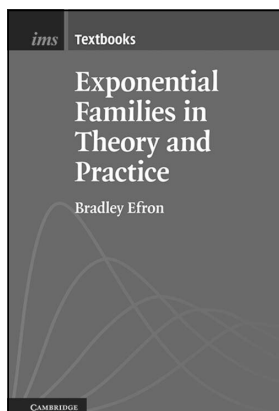
Exponential Families in Theory and Practice

Bradley Efron, Stanford University

During the past half-century, exponential families have attained a position at the center of parametric statistical inference. Theoretical advances have been matched, and more than matched, in the world of applications, where logistic regression by itself has become the go-to methodology in medical statistics, computer-based prediction algorithms, and the social sciences. This book is based on a one-semester graduate course for first year Ph.D. and advanced master's students. After presenting the basic structure of univariate and multivariate exponential families, their application to generalized linear models including logistic and Poisson regression is described in detail, emphasizing geometrical ideas, computational practice, and the analogy with ordinary linear regression. Connections are made with a variety of current statistical methodologies: missing data, survival analysis and proportional hazards, false discovery rates, bootstrapping, and empirical Bayes analysis. The book connects exponential family theory with its applications in a way that doesn't require advanced mathematical preparation.

www.imstat.org/cup/

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS
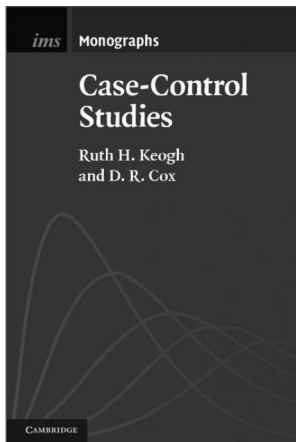
Case-Control Studies

Ruth H. Keogh
and D. R. Cox

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

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