ISSN 0090-5364 (print)
ISSN 2168-8966 (online)

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 48, No. 6—December 2020

ISSN 0090-5364 (print)
ISSN 2168-8966 (online)

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Vol. 48, No. 6—December 2020

THE ANNALS OF STATISTICS Vol. 48, No. 6, pp. 3113–3720 December 2020

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.

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 48, Number 6, December 2020. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Rd, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2020, Vol. 48, No. 6, 3113–3137
https://doi.org/10.1214/19-AOS1845

EMPIRICAL BAYES ORACLE UNCERTAINTY QUANTIFICATION FOR REGRESSION

BY EDUARD BELITSER[1] AND SUBHASHIS GHOSAL[2]

[1]*Department of Mathematics, Vrije Universiteit Amsterdam, e.n.belitser@vu.nl*
[2]*Department of Statistics, North Carolina State University, sghosal@stat.ncsu.edu*

We propose an empirical Bayes method for high-dimensional linear regression models. Following an oracle approach that quantifies the error locally for each possible value of the parameter, we show that an empirical Bayes posterior contracts at the optimal rate at all parameters and leads to uniform size-optimal credible balls with guaranteed coverage under an "excessive bias restriction" condition. This condition gives rise to a new slicing of the entire space that is suitable for ensuring uniformity in uncertainty quantification. The obtained results immediately lead to optimal contraction and coverage properties for many conceivable classes simultaneously. The results are also extended to high-dimensional additive nonparametric regression models.

REFERENCES

[1] BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168 https://doi.org/10.1214/009053604000000085

[2] BELITSER, E. (2017). On coverage and local radial rates of credible sets. *Ann. Statist.* **45** 1124–1151. MR3662450 https://doi.org/10.1214/16-AOS1477

[3] BELITSER, E. and GHOSAL, S. (2020). Supplement to "Empirical Bayes oracle uncertainty quantification for regression." https://doi.org/10.1214/19-AOS1845SUPP

[4] BELITSER, E. and NURUSHEV, N. (2020). Needles and straw in a haystack: Robust empirical Bayes confidence for possibly sparse sequences. *Bernoulli* **26** 191–225. MR403632 https://doi.org/10.3150/19-BEJ1122

[5] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

[6] BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. MR3102549 https://doi.org/10.3150/12-BEJSP11

[7] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

[8] CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049

[9] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856 https://doi.org/10.1214/13-AOS1133

[10] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246

[11] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[12] CASTILLO, I. and SZABÓ, B. (2020). Spike and slab empirical Bayes sparse credible sets. *Bernoulli* **26** 127–158. MR4036030 https://doi.org/10.3150/19-BEJ1119

[13] 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

[14] CURTIS, S. M., BANERJEE, S. and GHOSAL, S. (2014). Fast Bayesian model assessment for nonparametric additive regression. *Comput. Statist. Data Anal.* **71** 347–358. MR3131975 https://doi.org/10.1016/j.csda.2013.05.012

[15] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

MSC2020 subject classifications. Primary 62G15, 62C05, 62H35; secondary 62G99.
Key words and phrases. Credible ball, coverage, empirical Bayes, excessive bias restriction, oracle rate.

[16] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119 https://doi.org/10.1214/aos/1017938917

[17] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[18] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

[19] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 https://doi.org/10.1214/11-AOS920

[20] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135 https://doi.org/10.1214/aos/1176347253

[21] LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722

[22] 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

[23] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. MR2572443 https://doi.org/10.1214/09-AOS692

[24] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450 https://doi.org/10.1214/13-AOS1170

[25] PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913 https://doi.org/10.1214/aos/1016120374

[26] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704

[27] RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

[28] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 https://doi.org/10.1214/16-AOS1533

[29] SNIEKERS, S. and VAN DER VAART, A. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166** 78–86. MR3390135 https://doi.org/10.1016/j.jspi.2014.07.008

[30] SNIEKERS, S. and VAN DER VAART, A. (2015). Adaptive Bayesian credible sets in regression with a Gaussian process prior. *Electron. J. Stat.* **9** 2475–2527. MR3425364 https://doi.org/10.1214/15-EJS1078

[31] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

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

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

[34] 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

[35] YOO, W. W. and GHOSAL, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.* **44** 1069–1102. MR3485954 https://doi.org/10.1214/15-AOS1398

The Annals of Statistics
2020, Vol. 48, No. 6, 3138–3160
https://doi.org/10.1214/19-AOS1882

ASYMPTOTIC JOINT DISTRIBUTION OF EXTREME EIGENVALUES AND TRACE OF LARGE SAMPLE COVARIANCE MATRIX IN A GENERALIZED SPIKED POPULATION MODEL

BY ZENG LI[1], FANG HAN[2] AND JIANFENG YAO[3]

[1]*Department of Statistics and Data Science, Southern University of Science and Technology, liz9@sustech.edu.cn*
[2]*Department of Statistics, University of Washington, fanghan@uw.edu*
[3]*Department of Statistics and Actuarial Science, University of Hong Kong, jeffyao@hku.hk*

This paper studies the joint limiting behavior of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model, where the asymptotic regime is such that the dimension and sample size grow proportionally. The form of the joint limiting distribution is applied to conduct Johnson–Graybill-type tests, a family of approaches testing for signals in a statistical model. For this, higher order correction is further made, helping alleviate the impact of finite-sample bias. The proof rests on determining the joint asymptotic behavior of two classes of spectral processes, corresponding to the extreme and linear spectral statistics, respectively.

REFERENCES

[1] BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl*. **1** 1150011, 21. MR2934717 https://doi.org/10.1142/S2010326311500110
[2] 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
[3] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal*. **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009
[4] BAI, Z. and SILVERSTEIN, J. (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
[5] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233
[6] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal*. **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003
[7] BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist*. **44** 598–628. MR3476611 https://doi.org/10.1214/15-AOS1378
[8] BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inf. Theory* **57** 2400–2419. MR2809098 https://doi.org/10.1109/TIT.2011.2111710
[9] 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
[10] CHOI, Y., TAYLOR, J. and TIBSHIRANI, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *Ann. Statist*. **45** 2590–2617. MR3737903 https://doi.org/10.1214/16-AOS1536
[11] CHOW, T. L. and TEUGELS, J. L. (1979). The sum and the maximum of i.i.d. random variables. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (*Hradec Králové*, 1978) 81–92. North-Holland, Amsterdam. MR0571177
[12] DAVIS, R. A., HEINY, J., MIKOSCH, T. and XIE, X. (2016). Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes* **19** 517–547. MR3535965 https://doi.org/10.1007/s10687-016-0251-7

MSC2020 subject classifications. Primary 62E20, 62H15; secondary 15B52.
Key words and phrases. Generalized spiked model, asymptotic distribution, extreme eigenvalues, trace, large sample covariance matrix, random matrix theory.

[13] DEO, R. S. (2016). On the Tracy–Widom approximation of studentized extreme eigenvalues of Wishart matrices. *J. Multivariate Anal.* **147** 265–272. MR3484182 https://doi.org/10.1016/j.jmva.2016.01.010

[14] HSING, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab.* **23** 938–947. MR1334178

[15] JOHNSON, D. E. and GRAYBILL, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* **67** 862–868. MR0400566

[16] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[17] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

[18] KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.

[19] LI, Z., HAN, F. and YAO, J. (2020). Supplement to "Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model." https://doi.org/10.1214/19-AOS1882SUPP

[20] MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359. MR2888709 https://doi.org/10.3150/10-BEJ334

[21] NADLER, B. (2011). On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. *J. Multivariate Anal.* **102** 363–371. MR2739121 https://doi.org/10.1016/j.jmva.2010.10.005

[22] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

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

[24] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

[25] SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541 https://doi.org/10.1006/jmva.1995.1058

[26] WANG, Q., SILVERSTEIN, J. W. and YAO, J. (2014). A note on the CLT of the LSS for sample covariance matrix from a spiked population model. *J. Multivariate Anal.* **130** 194–207. MR3229532 https://doi.org/10.1016/j.jmva.2014.04.021

[27] WANG, Q., SU, Z. and YAO, J. (2014). Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models. *Electron. J. Probab.* **19** no. 103, 28. MR3275855 https://doi.org/10.1214/ejp.v19-3339

[28] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

[29] 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

[30] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 https://doi.org/10.1214/11-AIHP414

[31] ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

[32] ZHENG, S., BAI, Z., YAO, J. and ZHU, H. (2016). CLT for linear spectral statistics of large dimensional sample covariance matrices with dependent data. Preprint. Available at arXiv:1708.03749.

The Annals of Statistics
2020, Vol. 48, No. 6, 3161–3182
https://doi.org/10.1214/19-AOS1924

SINGULARITY, MISSPECIFICATION AND THE CONVERGENCE RATE OF EM

BY RAAZ DWIVEDI[1,*], NHAT HO[1,†], KOULIK KHAMARU[2], MARTIN J. WAINWRIGHT[1,2,3,‡], MICHAEL I. JORDAN[1,2,§] AND BIN YU[1,2,¶]

[1]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley,* [*]*raaz.rsk@berkeley.edu;* [†]*minhnhat@berkeley.edu;* [‡]*wainwrig@berkeley.edu;* [§]*jordan@cs.berkeley.edu;* [¶]*binyu@berkeley.edu*

[2]*Department of Statistics, University of California, Berkeley, koulik@berkeley.edu*

[3]*Voleon Group*

A line of recent work has analyzed the behavior of the Expectation-Maximization (EM) algorithm in the well-specified setting, in which the population likelihood is locally strongly concave around its maximizing argument. Examples include suitably separated Gaussian mixture models and mixtures of linear regressions. We consider over-specified settings in which the number of fitted components is larger than the number of components in the true distribution. Such mis-specified settings can lead to singularity in the Fisher information matrix, and moreover, the maximum likelihood estimator based on n i.i.d. samples in d dimensions can have a nonstandard $\mathcal{O}((d/n)^{\frac{1}{4}})$ rate of convergence. Focusing on the simple setting of two-component mixtures fit to a d-dimensional Gaussian distribution, we study the behavior of the EM algorithm both when the mixture weights are different (unbalanced case), and are equal (balanced case). Our analysis reveals a sharp distinction between these two cases: in the former, the EM algorithm converges geometrically to a point at Euclidean distance of $\mathcal{O}((d/n)^{\frac{1}{2}})$ from the true parameter, whereas in the latter case, the convergence rate is exponentially slower, and the fixed point has a much lower $\mathcal{O}((d/n)^{\frac{1}{4}})$ accuracy. Analysis of this singular case requires the introduction of some novel techniques: in particular, we make use of a careful form of localization in the associated empirical process, and develop a recursive argument to progressively sharpen the statistical rate.

REFERENCES

[1] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. MR3611487 https://doi.org/10.1214/16-AOS1435

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

[3] CAI, T. T., MA, J. and ZHANG, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Statist.* **47** 1234–1267. MR3911111 https://doi.org/10.1214/18-AOS1711

[4] CHEN, J. and LI, P. (2009). Hypothesis test for normal mixture models: The EM approach. *Ann. Statist.* **37** 2523–2542. MR2543701 https://doi.org/10.1214/08-AOS651

[5] CHEN, J. H. (1995). Optimal rate of convergence for finite mixture models. *Ann. Statist.* **23** 221–233. MR1331665 https://doi.org/10.1214/aos/1176324464

[6] CHEN, Y. C. (2018). Statistical inference with local optima. Preprint. Available at arXiv:1807.04431.

[7] DASKALAKIS, C., TZAMOS, C. and ZAMPETAKIS, M. (2017). Ten steps of EM suffice for mixtures of two Gaussians. In *Proceedings of the* 2017 *Conference on Learning Theory*.

[8] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

MSC2020 subject classifications. Primary 62F15, 62G05; secondary 62G20.
Key words and phrases. Mixture models, expectation-maximization, Fisher information matrix, empirical process, nonasymptotic convergence guarantees, localization argument.

[9] DWIVEDI, R., HO, N., KHAMARU, K., WAINWRIGHT, M. J. and JORDAN, M. I. (2018). Theoretical guarantees for EM under misspecified Gaussian mixture models. In *NeurIPS* 31.

[10] DWIVEDI, R., HO, N., KHAMARU, K., WAINWRIGHT, M. J., JORDAN, M. I. and YU, B. (2020). Supplement to "Singularity, misspecification and the convergence rate of EM." https://doi.org/10.1214/19-AOS1924SUPP

[11] DWIVEDI, R., HO, N., KHAMARU, K., WAINWRIGHT, M. J., JORDAN, M. I. and YU, B. (2019). Challenges with EM in application to weakly identifiable mixture models. Preprint. Available at arXiv:1902.00194.

[12] GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329 https://doi.org/10.1214/aos/1013203453

[13] HAO, B., SUN, W. W., LIU, Y. and CHENG, G. (2018). Simultaneous clustering and estimation of heterogeneous graphical models. *J. Mach. Learn. Res.* **18** Art. ID 217. MR3827105

[14] HEINRICH, P. and KAHN, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870. MR3851757 https://doi.org/10.1214/17-AOS1641

[15] ISHWARAN, H., JAMES, L. F. and SUN, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc.* **96** 1316–1332. MR1946579 https://doi.org/10.1198/016214501753382255

[16] KLUSOWSKI, J. M., YANG, D. and BRINDA, W. D. (2019). Estimating the coefficients of a mixture of two linear regressions by expectation maximization. *IEEE Trans. Inform. Theory* **65** 3515–3524. MR3959002 https://doi.org/10.1109/TIT.2019.2891628

[17] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442 https://doi.org/10.1214/009053606000001019

[18] KUMAR, R. and SCHMIDT, M. (2017). Convergence rate of expectation-maximization. In 10*th NIPS Workshop on Optimization for Machine Learning*.

[19] LI, P., CHEN, J. and MARRIOTT, P. (2009). Non-finite Fisher information and homogeneity: An EM approach. *Biometrika* **96** 411–426. MR2507152 https://doi.org/10.1093/biomet/asp011

[20] MA, J., XU, L. and JORDAN, M. I. (2000). Asymptotic convergence rate of the EM algorithm for Gaussian mixtures. *Neural Comput.* **12** 2881–2907.

[21] NGUYEN, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

[22] REDNER, R. A. and WALKER, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** 195–239. MR0738930 https://doi.org/10.1137/1026034

[23] RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

[24] ROUSSEAU, J. and MENGERSEN, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 689–710. MR2867454 https://doi.org/10.1111/j.1467-9868.2011.00781.x

[25] STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

[26] VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

[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] WANG, Z., GU, Q., NING, Y. and LIU, H. (2015). High-dimensional expectation-maximization algorithm: Statistical optimization and asymptotic normality. In *Advances in Neural Information Processing Systems* 28.

[29] WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103. MR0684867 https://doi.org/10.1214/aos/1176346060

[30] XU, J., HSU, D. and MALEKI, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. In *Advances in Neural Information Processing Systems* 29.

[31] XU, L. and JORDAN, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Comput.* **8** 129–151. https://doi.org/10.1162/neco.1996.8.1.129

[32] YAN, B., YIN, M. and SARKAR, P. (2017). Convergence of gradient EM on multi-component mixture of Gaussians. In *Advances in Neural Information Processing Systems* 30.

[33] YI, X. and CARAMANIS, C. (2015). Regularized EM algorithms: A unified framework and statistical guarantees. In *Advances in Neural Information Processing Systems* 28.

The Annals of Statistics
2020, Vol. 48, No. 6, 3183–3205
https://doi.org/10.1214/19-AOS1925

TOWARDS OPTIMAL ESTIMATION OF BIVARIATE ISOTONIC MATRICES WITH UNKNOWN PERMUTATIONS

BY CHENG MAO[1], ASHWIN PANANJADY[2] AND MARTIN J. WAINWRIGHT[2,3]

[1]*School of Mathematics, Georgia Institute of Technology, cheng.mao@math.gatech.edu*
[2]*Department of EECS, University of California, Berkeley, ashwinpm@berkeley.edu*
[3]*Department of Statistics, University of California, Berkeley, wainwrig@stat.berkeley.edu*

Many applications, including rank aggregation, crowd-labeling and graphon estimation, can be modeled in terms of a bivariate isotonic matrix with unknown permutations acting on its rows and/or columns. We consider the problem of estimating an unknown matrix in this class, based on noisy observations of (possibly, a subset of) its entries. We design and analyze polynomial-time algorithms that improve upon the state of the art in two distinct metrics, showing, in particular, that minimax optimal, computationally efficient estimation is achievable in certain settings. Along the way, we prove matching upper and lower bounds on the minimax radii of certain cone testing problems, which may be of independent interest.[1]

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 177. MR3827065

[2] AGARWAL, A., AGARWAL, S., ASSADI, S. and KHANNA, S. (2017). Learning with limited rounds of adaptivity: Coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In *Conference on Learning Theory* 39–75.

[3] BALLINGER, T. P. and WILCOX, N. T. (1997). Decisions, error and heterogeneity. *Econ. J.* **107** 1090–1105.

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

[5] BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2015). Consistent nonparametric estimation for heavy-tailed sparse graphs. ArXiv preprint. Available at arXiv:1508.06675.

[6] 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

[7] 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

[8] BRIL, G., DYKSTRA, R., PILLERS, C. and ROBERTSON, T. (1984). Algorithm AS 206: Isotonic regression in two independent variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **33** 352–357.

[9] CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216.

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

[11] 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

[12] CHATTERJEE, S. and MUKHERJEE, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans. Inform. Theory* **65** 3525–3539. MR3959003 https://doi.org/10.1109/TIT.2019.2893911

MSC2020 subject classifications. Primary 62J15, 62F07.

Key words and phrases. Permutation-based models, ranking, pairwise comparisons, crowd-labeling, statistical-computational gap, shape-constrained estimation.

[1]A corollary of Theorem 3.5 of this paper was presented at the Conference on Learning Theory (COLT) 2018, and a statement of this result appears in the abstract (In *Proceedings of the 31st Conference On Learning Theory* (2018) 2037–2042 PMLR).

[13] CHEN, Y. and SUH, C. (2015). Spectral MLE: Top-k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning* 371–380.

[14] COLLIER, O. and DALALYAN, A. S. (2016). Minimax rates in permutation estimation for feature matching. *J. Mach. Learn. Res.* **17** 6. MR3482926

[15] DALVI, N., DASGUPTA, A., KUMAR, R. and RASTOGI, V. (2013). Aggregating crowdsourced binary ratings. In *Proceedings of the* 22*nd International Conference on World Wide Web* 285–294. ACM, New York.

[16] DAWID, A. P. and SKENE, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *Appl. Stat.* 20–28.

[17] FALAHATGAR, M., ORLITSKY, A., PICHAPATI, V. and SURESH, A. T. (2017). Maximum selection and ranking under noisy comparisons. ArXiv preprint. Available at arXiv:1705.05366.

[18] FISHBURN, P. C. (1973). Binary choice probabilities: On the varieties of stochastic transitivity. *J. Math. Psych.* **10** 327–352. MR0327330 https://doi.org/10.1016/0022-2496(73)90021-7

[19] 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

[20] FOGEL, F., JENATTON, R., BACH, F. and D'ASPREMONT, A. (2013). Convex relaxations for permutation problems. In *Adv. in Neural Information Processing Systems* 26 1016–1024. Curran Associates.

[21] GAO, C. (2017). Phase transitions in approximate ranking. ArXiv preprint. Available at arXiv:1711.11189.

[22] GAO, C., LU, Y. and ZHOU, D. (2016). Exact exponent in optimal rates for crowdsourcing. In *International Conference on Machine Learning* 603–611.

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

[24] GAO, C. and ZHOU, D. (2013). Minimax optimal convergence rates for estimating ground truth from crowdsourced labels. ArXiv preprint. Available at arXiv:1310.5764.

[25] GAO, F. and WELLNER, J. A. (2007). Entropy estimate for high-dimensional monotonic functions. *J. Multivariate Anal.* **98** 1751–1764. MR2392431 https://doi.org/10.1016/j.jmva.2006.09.003

[26] GHOSH, A., KALE, S. and MCAFEE, P. (2011). Who moderates the moderators?: Crowdsourcing abuse detection in user-generated content. In *Proceedings of the* 12*th ACM Conference on Electronic Commerce* 167–176. ACM, New York.

[27] HAJEK, B., OH, S. and XU, J. (2014). Minimax-optimal inference from partial rankings. In *Adv. in Neural Information Processing Sys* 1475–1483.

[28] 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

[29] HECKEL, R., SHAH, N. B., RAMCHANDRAN, K. and WAINWRIGHT, M. J. (2019). Active ranking from pairwise comparisons and when parametric assumptions do not help. *Ann. Statist.* **47** 3099–3126. MR4025736 https://doi.org/10.1214/18-AOS1772

[30] KAHN, A. B. (1962). Topological sorting of large networks. *Commun. ACM* **5** 558–562.

[31] KARGER, D. R., OH, S. and SHAH, D. (2011). Budget-optimal crowdsourcing using low-rank matrix approximations. In *Communication, Control, and Computing* (*Allerton*), 2011 49*th Annual Allerton Conference on* 284–291. IEEE.

[32] KARGER, D. R., OH, S. and SHAH, D. (2011). Iterative learning for reliable crowdsourcing systems. In *Adv. in Neural Information Processing Systems* 1953–1961.

[33] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

[34] 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

[35] KYNG, R., RAO, A. and SACHDEVA, S. (2015). Fast, provable algorithms for isotonic regression in all l_p-norms. In *Adv. in Neural Information Processing Sys* 2719–2727.

[36] LE, J., EDMONDS, A., HESTER, V. and BIEWALD, L. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution.

[37] LEE, C. and SHAH, D. (2018). Reducing crowdsourcing to graphon estimation, statistically. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research* **84** 1741–1750.

[38] LIU, Q., PENG, J. and IHLER, A. T. (2012). Variational inference for crowdsourcing. In *Adv. in Neural Information Processing Systems* 692–700.

[39] LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York. MR0108411

[40] MAO, C., PANANJADY, A. and WAINWRIGHT, M. J. (2018). Breaking the $1/\sqrt{n}$ barrier: Faster rates for permutation-based models in polynomial time. In *Proceedings of the* 31*st Conference on Learning Theory. Proceedings of Machine Learning Research* **75** 2037–2042. PMLR.

[41] MAO, C., PANANJADY, A. and WAINWRIGHT, M. J. (2020). Supplement to "Towards Optimal Estimation of Bivariate Isotonic Matrices with Unknown Permutations." https://doi.org/10.1214/19-AOS1925SUPP.

[42] 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

[43] MCLAUGHLIN, D. H. and LUCE, R. D. (1965). Stochastic transitivity and cancellation of preferences between bitter-sweet solutions. *Psychol. Sci.* **2** 1–12. 89–90.

[44] MONTANARI, A. (2015). Computational implications of reducing data to sufficient statistics. *Electron. J. Stat.* **9** 2370–2390. MR3417186 https://doi.org/10.1214/15-EJS1059

[45] NEGAHBAN, S., OH, S. and SHAH, D. (2017). Rank centrality: Ranking from pairwise comparisons. *Oper. Res.* **65** 266–287. MR3613103 https://doi.org/10.1287/opre.2016.1534

[46] NEGAHBAN, S., OH, S., THEKUMPARAMPIL, K. K. and XU, J. (2018). Learning from comparisons and choices. *J. Mach. Learn. Res.* **19** 40. MR3862447

[47] NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

[48] NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1985). The rate of convergence of nonparametric estimates of maximum likelihood type. *Problemy Peredachi Informatsii* **21** 17–33. MR0820705

[49] OLESON, D., SOROKIN, A., LAUGHLIN, G., HESTER, V., LE, J. and BIEWALD, L. (2011). Programmatic gold: Targeted and scalable quality assurance in crowdsourcing. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

[50] OYAMA, S., BABA, Y., SAKURAI, Y. and KASHIMA, H. (2013). Accurate integration of crowdsourced labels using workers' self-reported confidence scores. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

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

[52] PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2017). Denoising linear models with permuted data. In *Information Theory* (*ISIT*), 2017 *IEEE International Symposium on* 446–450. IEEE.

[53] PARK, D., NEEMAN, J., ZHANG, J., SANGHAVI, S. and DHILLON, I. (2015). Preference completion: Large-scale collaborative ranking from pairwise comparisons. In *International Conference on Machine Learning* 1907–1916.

[54] RAJKUMAR, A. and AGARWAL, S. (2014). A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *International Conference on Machine Learning* 118–126.

[55] RAJKUMAR, A. and AGARWAL, S. (2016). When can we rank well from comparisons of O $(n \log(n))$ non-actively chosen pairs? In *Conf. on Learning Theory* 1376–1401.

[56] RZESZOTARSKI, J. M. and KITTUR, A. (2011). Instrumenting the crowd: Using implicit behavioral measures to predict task performance. In *Proceedings of the* 24*th Annual ACM Symposium on User Interface Software and Technology* 13–22. ACM, New York.

[57] SHAH, N. B., BALAKRISHNAN, S., BRADLEY, J., PAREKH, A., RAMCHANDRAN, K. and WAINWRIGHT, M. J. (2016). Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *J. Mach. Learn. Res.* **17** 58. MR3504618

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

[59] SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2016). A permutation-based model for crowd labeling: Optimal estimation and robustness. ArXiv preprint. Available at arXiv:1606.09632v2.

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

[61] SHAH, N. B. and WAINWRIGHT, M. J. (2017). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **18** 199. MR3827087

[62] SONG, D. and SHAH, D. (2018). Learning mixture model with missing values and its application to rankings. ArXiv preprint. Available at arXiv:1812.11917.

[63] THURSTONE, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* **34** 273.

[64] WEI, Y., WAINWRIGHT, M. J. and GUNTUBOYINA, A. (2019). The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii. *Ann. Statist.* **47** 994–1024. MR3909958 https://doi.org/10.1214/18-AOS1701

[65] WELINDER, P., BRANSON, S., PERONA, P. and BELONGIE, S. J. (2010). The multidimensional wisdom of crowds. In *Adv. in Neural Information Processing Sys* 2424–2432.

[66] WHITEHILL, J., WU, T.-F., BERGSMA, J., MOVELLAN, J. R. and RUVOLO, P. L. (2009). Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems* 2035–2043.

[67] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. ArXiv preprint. Available at arXiv:1309.5936.

[68] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

[69] ZHANG, Y., CHEN, X., ZHOU, D. and JORDAN, M. I. (2016). Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. *J. Mach. Learn. Res.* **17** 102. MR3543508

The Annals of Statistics
2020, Vol. 48, No. 6, 3206–3227
https://doi.org/10.1214/19-AOS1926

HIGH-DIMENSIONAL CONSISTENT INDEPENDENCE TESTING WITH MAXIMA OF RANK CORRELATIONS

BY MATHIAS DRTON[1], FANG HAN[2,*] AND HONGJIAN SHI[2,†]

[1]*Department of Mathematics, Technical University of Munich, mathias.drton@tum.de*
[2]*Department of Statistics, University of Washington, [*]fanghan@uw.edu; [†]hongshi@uw.edu*

Testing mutual independence for high-dimensional observations is a fundamental statistical challenge. Popular tests based on linear and simple rank correlations are known to be incapable of detecting nonlinear, nonmonotone relationships, calling for methods that can account for such dependences. To address this challenge, we propose a family of tests that are constructed using maxima of pairwise rank correlations that permit consistent assessment of pairwise independence. Built upon a newly developed Cramér-type moderate deviation theorem for degenerate U-statistics, our results cover a variety of rank correlations including Hoeffding's D, Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^*. The proposed tests are distribution-free in the class of multivariate distributions with continuous margins, implementable without the need for permutation, and are shown to be rate-optimal against sparse alternatives under the Gaussian copula model. As a by-product of the study, we reveal an identity between the aforementioned three rank correlation statistics, and hence make a step towards proving a conjecture of Bergsma and Dassios.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR1990662

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542. MR1235426

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

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

BAO, Z. (2019). Tracy–Widom limit for Kendall's tau. *Ann. Statist.* **47** 3504–3532. MR4025750 https://doi.org/10.1214/18-AOS1786

BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 1–32. MR2988403 https://doi.org/10.1214/EJP.v17-1962

BENTKUS, V. and GÖTZE, F. (1997). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Related Fields* **109** 367–416. MR1481126 https://doi.org/10.1007/s004400050136

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

MSC2020 subject classifications. 62G10.
Key words and phrases. Degenerate U-statistics, extreme value distribution, independence test, maximum-type test, rank statistics, rate-optimality.

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

CHATTERJEE, S. K. (2020). A new coefficient of correlation. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2020.1758115

DRTON, M., HAN, F. and SHI, H. (2020). Supplement to "High-dimensional consistent independence testing with maxima of rank correlations." https://doi.org/10.1214/19-AOS1926SUPP

GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

GÖTZE, F. and ZAITSEV, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *Ann. Probab.* **42** 354–397. MR3161488 https://doi.org/10.1214/13-AOP839

GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 20 (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) 984–991. Curran Associates, Red Hook, NY.

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

HAN, F., XU, S. and ZHOU, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli* **24** 1787–1833. MR3757515 https://doi.org/10.3150/16-BEJ912

HELLER, R., HELLER, Y., KAUFMAN, S., BRILL, B. and GORFINE, M. (2016). Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17** Paper No. 29, 54. MR3491123

HELLER, Y. and HELLER, R. (2016). Computing the Bergsma Dassios sign-covariance. Available at arXiv:1605.08732.

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

HOLLANDER, M., WOLFE, D. A. and CHICKEN, E. (2014). *Nonparametric Statistical Methods*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3221959

HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. MR3556612 https://doi.org/10.1080/00401706.2015.1054435

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857 https://doi.org/10.1214/13-AOS1134

KENDALL, M. and STUART, A. (1979). *The Advanced Theory of Statistics. Vol.* 2: *Inference and Relationship*, 4th ed. Charles Griffin & Co. Ltd., London.

KINNEY, J. B. and ATWAL, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **111** 3354–3359. MR3200177 https://doi.org/10.1073/pnas.1309933111

KNIGHT, W. R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *J. Amer. Statist. Assoc.* **61** 436–439.

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

LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

NANDY, P., WEIHS, L. and DRTON, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Stat.* **10** 2287–2311. MR3541972 https://doi.org/10.1214/16-EJS1166

PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York. MR0092296

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197 https://doi.org/10.1093/biomet/92.4.951

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

WEIHS, L., DRTON, M. and LEUNG, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Comput*. *Statist*. **31** 315–328. MR3481807 https://doi.org/10.1007/s00180-015-0639-x

WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

YANAGIMOTO, T. (1970). On measures of association and a related problem. *Ann*. *Inst*. *Statist*. *Math*. **22** 57–63.

YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans*. *Amer*. *Math*. *Soc*. **359** 5345–5363. MR2327033 https://doi.org/10.1090/S0002-9947-07-04192-X

The Annals of Statistics
2020, Vol. 48, No. 6, 3228–3250
https://doi.org/10.1214/19-AOS1927

OPTIMAL RATES OF ENTROPY ESTIMATION OVER LIPSCHITZ BALLS

BY YANJUN HAN[1,*], JIANTAO JIAO[2], TSACHY WEISSMAN[1,†] AND YIHONG WU[3]

[1]*Department of Electrical Engineering, Stanford University, [*]yjhan@stanford.edu; [†]tsachy@stanford.edu*
[2]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, jiantao@eecs.berkeley.edu*
[3]*Department of Statistics and Data Science, Yale University, yihong.wu@yale.edu*

We consider the problem of minimax estimation of the entropy of a density over Lipschitz balls. Dropping the usual assumption that the density is bounded away from zero, we obtain the minimax rates $(n \ln n)^{-s/(s+d)} + n^{-1/2}$ for $0 < s \le 2$ for densities supported on $[0, 1]^d$, where s is the smoothness parameter and n is the number of independent samples. We generalize the results to densities with unbounded support: given an Orlicz functions Ψ of rapid growth (such as the subexponential and sub-Gaussian classes), the minimax rates for densities with bounded Ψ-Orlicz norm increase to $(n \ln n)^{-s/(s+d)}(\Psi^{-1}(n))^{d(1-d/p(s+d))} + n^{-1/2}$, where p is the norm parameter in the Lipschitz ball. We also show that the integral-form plug-in estimators with kernel density estimates fail to achieve the minimax rates, and characterize their worst case performances over the Lipschitz ball.

One of the key steps in analyzing the bias relies on a novel application of the Hardy–Littlewood maximal inequality, which also leads to a new inequality on the Fisher information that may be of independent interest.

REFERENCES

[1] BEIRLANT, J., DUDEWICZ, E. J., GYÖRFI, L. and VAN DER MEULEN, E. C. (1997). Nonparametric entropy estimation: An overview. *Int. J. Math. Stat. Sci.* **6** 17–39. MR1471870

[2] BERRETT, T. B., SAMWORTH, R. J. and YUAN, M. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist.* **47** 288–318. MR3909934 https://doi.org/10.1214/18-AOS1688

[3] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[4] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya*, *Ser. A* **50** 381–393. MR1065550

[5] BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

[6] BU, Y., ZOU, S., LIANG, Y. and VEERAVALLI, V. V. (2018). Estimation of KL divergence: Optimal minimax rate. *IEEE Trans. Inform. Theory* **64** 2648–2674. MR3782280 https://doi.org/10.1109/TIT.2018.2805844

[7] CAI, T. T. and LOW, M. G. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.* **31** 1140–1153. MR2001645 https://doi.org/10.1214/aos/1059655908

[8] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147

[9] CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041. MR2816346 https://doi.org/10.1214/10-AOS849

[10] COSTA, J. A. and HERO, A. O. (2004). Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. Signal Process.* **52** 2210–2221. MR2085582 https://doi.org/10.1109/TSP.2004.831130

[11] DELATTRE, S. and FOURNIER, N. (2017). On the Kozachenko–Leonenko entropy estimator. *J. Statist. Plann. Inference* **185** 69–93. MR3612672 https://doi.org/10.1016/j.jspi.2017.01.004

MSC2020 subject classifications. Primary 62G05; secondary 62C20.
Key words and phrases. Nonparametrics, nonsmooth functional estimation, minimax rate, polynomial approximation.

[12] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Springer, Berlin. MR1261635 https://doi.org/10.1007/978-3-662-02888-9

[13] Donoho, D. L. (1997). Renormalizing experiments for nonlinear functionals. In *Festschrift for Lucien Le Cam* 167–181. Springer, New York. MR1462945 https://doi.org/10.1007/978-1-4612-1880-7_11

[14] Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9

[15] El Haje Hussein, F. and Golubev, Y. (2009). On entropy estimation by m-spacing method. *J. Math. Sci.* **163** 290–309.

[16] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294. MR1126325 https://doi.org/10.1214/aos/1176348249

[17] Gao, W., Oh, S. and Viswanath, P. (2018). Breaking the bandwidth barrier: Geometrical adaptive entropy estimation. *IEEE Trans. Inform. Theory* **64** 3313–3330. MR3798379 https://doi.org/10.1109/TIT.2018.2810313

[18] Gao, W., Oh, S. and Viswanath, P. (2018). Demystifying fixed k-nearest neighbor information estimators. *IEEE Trans. Inform. Theory* **64** 5629–5661. MR3832327 https://doi.org/10.1109/TIT.2018.2807481

[19] Goldstein, L. and Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.* **20** 1306–1328. MR1186251 https://doi.org/10.1214/aos/1176348770

[20] Györfi, L. and van der Meulen, E. C. (1991). On the nonparametric estimation of the entropy functional. In *Nonparametric Functional Estimation and Related Topics* (*Spetses*, 1990). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **335** 81–95. Kluwer Academic, Dordrecht. MR1154321 https://doi.org/10.1007/978-94-011-3222-0_6

[21] Hall, P. (1984). Limit theorems for sums of general functions of m-spacings. *Math. Proc. Cambridge Philos. Soc.* **96** 517–532. MR0757846 https://doi.org/10.1017/S0305004100062459

[22] Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115. MR0907270 https://doi.org/10.1016/0167-7152(87)90083-6

[23] Hall, P. and Morton, S. C. (1993). On the estimation of entropy. *Ann. Inst. Statist. Math.* **45** 69–88. MR1220291 https://doi.org/10.1007/BF00773669

[24] Han, Y., Jiao, J., Mukherjee, R. and Weissman, T. (2017). On estimation of L_r-norms in Gaussian white noise models. Preprint. Available at arXiv:1710.03863.

[25] Han, Y., Jiao, J. and Weissman, T. (2016). Minimax rate-optimal estimation of divergences between discrete distributions. Preprint. Available at arXiv:1605.09124.

[26] Han, Y., Jiao, J., Weissman, T. and Wu, Y. (2020). Supplement to "Optimal rates of entropy estimation over Lipschitz balls." https://doi.org/10.1214/19-AOS1927SUPP.

[27] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (2012). *Wavelets*, *Approximation*, *and Statistical Applications*. *Lecture Notes in Statistics* **129**. Springer, New York. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

[28] Has'minskiĭ, R. Z. and Ibragimov, I. A. (1980). Some estimation problems for stochastic differential equations. In *Stochastic Differential Systems* (*Proc. IFIP-WG* 7/1 *Working Conf.*, *Vilnius*, 1978). *Lecture Notes in Control and Information Sci.* **25** 1–12. Springer, Berlin. MR0609167

[29] Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M. and Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **441** 1–46.

[30] Ibragimov, I. A., Nemirovskiĭ, A. S. and Khas'minskiĭ, R. Z. (1987). Some problems of nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

[31] Jawerth, B. (1977). Some observations on Besov and Lizorkin–Triebel spaces. *Math. Scand.* **40** 94–104. MR0454618 https://doi.org/10.7146/math.scand.a-11678

[32] Jiao, J., Han, Y. and Weissman, T. (2018). Minimax estimation of the L_1 distance. *IEEE Trans. Inform. Theory* **64** 6672–6706. MR3860754 https://doi.org/10.1109/TIT.2018.2846245

[33] Jiao, J., Venkat, K., Han, Y. and Weissman, T. (2015). Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inform. Theory* **61** 2835–2885. MR3342309 https://doi.org/10.1109/TIT.2015.2412945

[34] Joe, H. (1989). Estimation of entropy and other functionals of a multivariate density. *Ann. Inst. Statist. Math.* **41** 683–697. MR1039399 https://doi.org/10.1007/BF00057735

[35] Kandasamy, K., Krishnamurthy, A., Poczos, B. and Wasserman, L. (2015). Nonparametric von Mises estimators for entropies, divergences and mutual informations. In *Advances in Neural Information Processing Systems* 397–405.

[36] Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24** 485–507. MR1394973 https://doi.org/10.1214/aos/1032894450

[37] Korostelëv, A. P. and Tsybakov, A. B. (2012). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450 https://doi.org/10.1007/978-1-4612-2712-0

[38] Kraskov, A., Stögbauer, H. and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* (3) **69** 066138, 16. MR2096503 https://doi.org/10.1103/PhysRevE.69.066138

[39] Krishnamurthy, A., Kandasamy, K., Poczos, B. and Wasserman, L. (2014). Nonparametric estimation of Rényi divergence and friends. In *International Conference on Machine Learning* 919–927.

[40] Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

[41] Lepski, O., Nemirovski, A. and Spokoiny, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

[42] Levit, B. Y. (1978). Asymptotically efficient estimation of nonlinear functionals. *Probl. Inf. Transm.* **14** 65–72. MR0533450

[43] Moon, K. R., Sricharan, K., Greenewald, K. and Hero, A. O. III (2016). Nonparametric ensemble estimation of distributional functionals. Preprint. Available at arXiv:1601.06884.

[44] Mukherjee, R., Newey, W. K. and Robins, J. M. (2017). Semiparametric efficient empirical higher order influence function estimators. Preprint. Available at arXiv:1705.07577.

[45] Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

[46] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959 https://doi.org/10.1214/aos/1032181160

[47] Paninski, L. (2004). Estimating entropy on m bins given fewer than m samples. *IEEE Trans. Inform. Theory* **50** 2200–2203. MR2097210 https://doi.org/10.1109/TIT.2004.833360

[48] Paninski, L. and Yajima, M. (2008). Undersmoothed kernel entropy estimators. *IEEE Trans. Inform. Theory* **54** 4384–4388. MR2451978 https://doi.org/10.1109/TIT.2008.928251

[49] Patschkowski, T. and Rohde, A. (2016). Adaptation to lowest density regions with application to support recovery. *Ann. Statist.* **44** 255–287. MR3449768 https://doi.org/10.1214/15-AOS1366

[50] Ray, K. and Schmidt-Hieber, J. (2019). Asymptotic nonequivalence of density estimation and Gaussian white noise for small densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 2195–2208. MR4029152 https://doi.org/10.1214/18-AIHP946

[51] Rényi, A. (1959). On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hung.* **10** 193–215. MR0107575 https://doi.org/10.1007/BF02063299

[52] 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*. *Inst. Math. Stat.* (*IMS*) *Collect.* **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

[53] Robins, J. M., Li, L., Mukherjee, R., Tchetgen, E. T. and van der Vaart, A. (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist.* **45** 1951–1987. MR3718158 https://doi.org/10.1214/16-AOS1515

[54] Singh, S. and Póczos, B. (2016). Finite-sample analysis of fixed-k nearest neighbor density functional estimators. In *Advances in Neural Information Processing Systems* 1217–1225.

[55] Sricharan, K., Raich, R. and Hero, A. O. III (2012). Estimation of nonlinear functionals of densities with confidence. *IEEE Trans. Inform. Theory* **58** 4135–4159. MR2943080 https://doi.org/10.1109/TIT.2012.2195549

[56] Steele, J. M. (1986). An Efron–Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14** 753–758. MR0840528 https://doi.org/10.1214/aos/1176349952

[57] Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. *Princeton Mathematical Series* **30**. Princeton Univ. Press, Princeton, NJ. MR0290095

[58] Tchetgen, E., Li, L., Robins, J. and van der Vaart, A. (2008). Minimax estimation of the integral of a power of a density. *Statist. Probab. Lett.* **78** 3307–3311. MR2479495 https://doi.org/10.1016/j.spl.2008.07.001

[59] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[60] Tsybakov, A. B. and van der Meulen, E. C. (1996). Root-n consistent estimators of entropy for densities with unbounded support. *Scand. J. Stat.* **23** 75–83. MR1380483

[61] Valiant, G. and Valiant, P. (2011). The power of linear estimators. In 2011 *IEEE* 52*nd Annual Symposium on Foundations of Computer Science—FOCS* 2011 403–412. IEEE Computer Soc., Los Alamitos, CA. MR2932716 https://doi.org/10.1109/FOCS.2011.81

[62] van Es, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Stat.* **19** 61–72. MR1172967

[63] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[64] WANG, Q., KULKARNI, S. R. and VERDÚ, S. (2009). Universal estimation of information measures for analog sources. *Found*. *Trends Commun*. *Inf*. *Theory* **5** 265–353.

[65] WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans*. *Inform*. *Theory* **62** 3702–3720. MR3506758 https://doi.org/10.1109/TIT.2016.2548468

[66] WU, Y. and YANG, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann*. *Statist*. **47** 857–883. MR3909953 https://doi.org/10.1214/17-AOS1665

The Annals of Statistics
2020, Vol. 48, No. 6, 3251–3282
https://doi.org/10.1214/19-AOS1928

LIMIT DISTRIBUTION THEORY FOR BLOCK ESTIMATORS IN MULTIPLE ISOTONIC REGRESSION

BY QIYANG HAN[1,*] AND CUN-HUI ZHANG[1,†]

[1]*Department of Statistics, Rutgers University,* [*]*qh85@stat.rutgers.edu;* [†]*czhang@stat.rutgers.edu*

We study limit distributions for the tuning-free max–min block estimator originally proposed in (Fokianos, Leucht and Neumann (2017)) in the problem of multiple isotonic regression, under both fixed lattice design and random design settings. We show that, if the regression function f_0 admits vanishing derivatives up to order α_k along the kth dimension ($k = 1, \ldots, d$) at a fixed point $x_0 \in (0, 1)^d$, and the errors have variance σ^2, then the max–min block estimator $\hat{f}_n$ satisfies

$$(n_*/\sigma^2)^{\frac{1}{2+\sum_{k\in\mathcal{D}_*}\alpha_k^{-1}}}\big(\hat{f}_n(x_0) - f_0(x_0)\big) \rightsquigarrow \mathbb{C}(f_0, x_0).$$

Here, $\mathcal{D}_*, n_*$, depending on $\{\alpha_k\}$ and the design points, are the set of all "effective dimensions" and the size of "effective samples" that drive the asymptotic limiting distribution, respectively. If furthermore either $\{\alpha_k\}$ are relative primes to each other or all mixed derivatives of f_0 of certain critical order vanish at x_0, then the limiting distribution can be represented as $\mathbb{C}(f_0, x_0) =_d K(f_0, x_0) \cdot \mathbb{D}_{\boldsymbol{\alpha}}$, where $K(f_0, x_0)$ is a constant depending on the local structure of the regression function f_0 at x_0, and $\mathbb{D}_{\boldsymbol{\alpha}}$ is a nonstandard limiting distribution generalizing the well-known Chernoff distribution in univariate problems. The above limit theorem is also shown to be optimal both in terms of the local rate of convergence and the dependence on the unknown regression function whenever such dependence is explicit (i.e., $K(f_0, x_0)$), for the full range of $\{\alpha_k\}$ in a local asymptotic minimax sense.

There are two interesting features in our local theory. First, the max–min block estimator automatically adapts to the local smoothness and the intrinsic dimension of the isotonic regression function at the optimal rate. Second, the optimally adaptive local rates are in general not the same in fixed lattice and random designs. In fact, the local rate in the fixed lattice design case is no slower than that in the random design case, and can be much faster when the local smoothness levels of the isotonic regression function or the sizes of the lattice differ substantially along different dimensions.

REFERENCES

[1] ALEXANDER, K. S. (1987). Central limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Related Fields* **75** 351–378. MR0890284 https://doi.org/10.1007/BF00318707

[2] ANEVSKI, D. and HÖSSJER, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. MR2283721 https://doi.org/10.1214/009053606000000443

[3] ANEVSKI, D. and PASTUKHOV, V. (2018). The asymptotic distribution of the isotonic regression estimator over a general countable pre-ordered set. *Electron. J. Stat.* **12** 4180–4208. MR3890765 https://doi.org/10.1214/18-EJS1507

[4] ANEVSKI, D. and SOULIER, P. (2011). Monotone spectral density estimation. *Ann. Statist.* **39** 418–438. MR2797852 https://doi.org/10.1214/10-AOS804

[5] BAGCHI, P., BANERJEE, M. and STOEV, S. A. (2016). Inference for monotone functions under short- and long-range dependence: Confidence intervals and new universal limits. *J. Amer. Statist. Assoc.* **111** 1634–1647. MR3601723 https://doi.org/10.1080/01621459.2015.1100622

MSC2020 subject classifications. Primary 60E15; secondary 62G05.
Key words and phrases. Limit distribution theory, multiple isotonic regression, Gaussian process, shape constraints, adaptation.

[6] BALABDAOUI, F., DUROT, C. and KOLADJO, F. (2017). On asymptotics of the discrete convex LSE of a p.m.f. *Bernoulli* **23** 1449–1480. MR3624867 https://doi.org/10.3150/15-BEJ754

[7] BALABDAOUI, F. and JANKOWSKI, H. (2016). Maximum likelihood estimation of a unimodal probability mass function. *Statist*. *Sinica* **26** 1061–1086. MR3559943

[8] BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **75** 769–790. MR3091658 https://doi.org/10.1111/rssb.12011

[9] BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann*. *Statist*. **37** 1299–1331. MR2509075 https://doi.org/10.1214/08-AOS609

[10] BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann*. *Statist*. **35** 2536–2564. MR2382657 https://doi.org/10.1214/009053607000000262

[11] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann*. *Statist*. **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[12] BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann*. *Statist*. **35** 543–574. MR2336859 https://doi.org/10.1214/009053606000001415

[13] 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

[14] 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

[15] BIAU, G. and DEVROYE, L. (2003). On the risk of estimates for block decreasing densities. *J*. *Multivariate Anal*. **86** 143–165. MR1994726 https://doi.org/10.1016/S0047-259X(02)00028-3

[16] 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

[17] BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann*. *Statist*. **30** 927–961. MR1926165 https://doi.org/10.1214/aos/1031689014

[18] CAROLAN, C. and DYKSTRA, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad*. *J*. *Statist*. **27** 557–566. MR1745821 https://doi.org/10.2307/3316111

[19] 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

[20] 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

[21] CHEN, Y. and WELLNER, J. A. (2016). On convex least squares estimation when the truth is linear. *Electron*. *J*. *Stat*. **10** 171–209. MR3466180 https://doi.org/10.1214/15-EJS1098

[22] DENG, H. and ZHANG, C.-H. (2018). Isotonic regression in multi-dimensional spaces and graphs. Preprint. Available at arXiv:1812.08944.

[23] DOSS, C. R. and WELLNER, J. A. (2019). Inference for the mode of a log-concave density. *Ann*. *Statist*. **47** 2950–2976. MR3988778 https://doi.org/10.1214/18-AOS1770

[24] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. MR3445285

[25] 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

[26] 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

[27] DUROT, C. and LOPUHAÄ, H. P. (2018). Limit theory in monotone function estimation. *Statist*. *Sci*. **33** 547–567. MR3881208 https://doi.org/10.1214/18-STS664

[28] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2017). On integrated L_1 convergence rate of an isotonic regression estimator for multivariate observations. Preprint. Available at arXiv:1710.04813.

[29] GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann*. *Probab*. **11** 1016–1027. MR0714964

[30] 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

[31] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab*. *Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[32] GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math*. **1648** 67–164. Springer, Berlin. MR1600884 https://doi.org/10.1007/BFb0095675

[33] GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann*. *Statist*. **27** 1316–1347. MR1740109 https://doi.org/10.1214/aos/1017938928

[34] GROENEBOOM, P. and JONGBLOED, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann*. *Statist*. **23** 1518–1542. MR1370294 https://doi.org/10.1214/aos/1176324310

[35] 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

[36] 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

[37] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann*. *Statist*. **29** 1620–1652. MR1891741 https://doi.org/10.1214/aos/1015345957

[38] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann*. *Statist*. **29** 1653–1698. MR1891742 https://doi.org/10.1214/aos/1015345958

[39] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[40] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab*. *Theory Related Fields* **163** 379–411. MR3405621 https://doi.org/10.1007/s00440-014-0595-3

[41] HAN, Q. (2019). Global empirical risk minimizers with "shape constraints" are rate optimal in general dimensions. Preprint. Available at arXiv:1905.12823.

[42] 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

[43] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann*. *Statist*. **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408

[44] HAN, Q. and ZHANG, C.-H. (2020). Supplement to "Limit distribution theory for multiple isotonic regression." https://doi.org/10.1214/19-AOS1928SUPP

[45] 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

[46] 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

[47] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann*. *Statist*. **42** 625–653. MR3210981 https://doi.org/10.1214/13-AOS1196

[48] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electron*. *J*. *Stat*. **3** 1567–1605. MR2578839 https://doi.org/10.1214/09-EJS526

[49] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli* **15** 1010–1035. MR2597581 https://doi.org/10.3150/09-BEJ202

[50] JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist*. *Probab*. *Lett*. **50** 279–284. MR1792307 https://doi.org/10.1016/S0167-7152(00)00104-8

[51] KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Adaptation in log-concave density estimation. *Ann*. *Statist*. **46** 2279–2306. MR3845018 https://doi.org/10.1214/17-AOS1619

[52] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann*. *Statist*. **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[53] 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

[54] 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

[55] PAVLIDES, M. G. (2012). Local asymptotic minimax theory for block-decreasing densities. *J*. *Statist*. *Plann*. *Inference* **142** 2322–2329. MR2911847 https://doi.org/10.1016/j.jspi.2012.03.003

[56] POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann*. *Statist*. **26** 1857–1877. MR1673281 https://doi.org/10.1214/aos/1024691360

[57] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā*, *Ser*. *A* **31** 23–36. MR0267677

[58] PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann*. *Math*. *Stat*. **41** 507–519. MR0260133 https://doi.org/10.1214/aoms/1177697091

[59] 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

[60] 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

[61] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[62] VAN EEDEN, C. (1958). *Testing and Estimating Ordered Parameters of Probability Distributions*. Mathematical Centre, Amsterdam. MR0102874

[63] 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

[64] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann*. *Statist*. **9** 443–448. MR0606630

[65] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

The Annals of Statistics
2020, Vol. 48, No. 6, 3283–3311
https://doi.org/10.1214/19-AOS1929

ASSESSMENT OF THE EXTENT OF CORROBORATION OF AN ELABORATE THEORY OF A CAUSAL HYPOTHESIS USING PARTIAL CONJUNCTIONS OF EVIDENCE FACTORS

BY BIKRAM KARMAKAR* AND DYLAN S. SMALL†

Department of Statistics, The Wharton School, University of Pennsylvania, *bikramk@wharton.upenn.edu; †dsmall@wharton.upenn.edu

An elaborate theory of predictions of a causal hypothesis consists of several falsifiable statements derived from the causal hypothesis. Statistical tests for the various pieces of the elaborate theory help to clarify how much the causal hypothesis is corroborated. In practice, the degree of corroboration of the causal hypothesis has been assessed by a verbal description of which of the several tests provides evidence for which of the several predictions. This verbal approach can miss quantitative patterns. In this paper, we develop a quantitative approach. We first decompose these various tests of the predictions into independent factors with different sources of potential biases. Support for the causal hypothesis is enhanced when many of these evidence factors support the predictions. A sensitivity analysis is used to assess the potential bias that could make the finding of the tests spurious. Along with this multiparameter sensitivity analysis, we consider the partial conjunctions of the tests. These partial conjunctions quantify the evidence supporting various fractions of the collection of predictions. A partial conjunction test involves combining tests of the components in the partial conjunction. We find the asymptotically optimal combination of tests in the context of a sensitivity analysis. Our analysis of an elaborate theory of a causal hypothesis controls for the familywise error rate.

REFERENCES

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Stat.* **38** 303–324. MR0207085 https://doi.org/10.1214/aoms/1177698949

BECKER, B. J. (1994). Combining significance levels. In *A Handbook of Research Synthesis* (H. Cooper and L. V. Hedges, eds.) 215–230. Russell Sage, New York.

BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x

BENJAMINI, Y., HELLER, R. and YEKUTIELI, D. (2009). Selective inference in complex research. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4255–4271. MR2546387 https://doi.org/10.1098/rsta.2009.0127

CENTERWALL, B. S. (1989). Exposure to television as a risk factor for violence. *Am. J. Epidemiol.* **129** 643–652.

CHEN, Z. (2011). Is the weighted z-test the best method for combining probabilities from independent tests? *J. Evol. Biol.* **24** 926–930.

COCHRAN, W. G. (1965). The planning of observational studies in human population (with discussion). *J. Roy. Statist. Soc. Ser. A* 234–266.

CORNFIELD, J., HAENSZEL, W., HAMMOND, E. C., LILIENFELD, A. M., SHIMKIN, M. B. and WYNDER, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* **22** 173–203.

CRUPI, V., CHATER, N. and TENTORI, K. (2013). New axioms for probability and likelihood ratio measures. *British J. Philos. Sci.* **64** 189–204. MR3024180 https://doi.org/10.1093/bjps/axs018

DING, P. and VANDERWEELE, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology* **27** 368–377.

EDGINGTON, E. S. (1972). An additive method for combining probability values from independent experiments. *J. Psychol.* **80** 351–363.

MSC2020 subject classifications. Primary 62G10; secondary 62K15, 03A10.

Key words and phrases. Causal inference, degree of corroboration, elaborate theory, evidence factors, observational studies.

EGLESTON, B. L., SCHARFSTEIN, D. O. and MACKENZIE, E. (2009). On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. *Biometrics* **65** 497–504. MR2751473 https://doi.org/10.1111/j.1541-0420.2008.01111.x

ERTEFAIE, A., SMALL, D. S. and ROSENBAUM, P. R. (2018). Quantitative evaluation of the trade-off of strengthened instruments and sample size in observational studies. *J. Amer. Statist. Assoc.* **113** 1122–1134. MR3862344 https://doi.org/10.1080/01621459.2017.1305275

FISHER, R. A. (1932). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.

FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.

FOGARTY, C. B. and HASEGAWA, R. B. (2019). Extended sensitivity analysis for heterogeneous unmeasured confounding with an application to sibling studies of returns to education. *Ann. Appl. Stat.* **13** 767–796. MR3963552 https://doi.org/10.1214/18-AOAS1215

FOGARTY, C. B. and SMALL, D. S. (2016). Sensitivity analysis for multiple comparisons in matched observational studies through quadratically constrained linear programming. *J. Amer. Statist. Assoc.* **111** 1820–1830. MR3601738 https://doi.org/10.1080/01621459.2015.1120675

GILBERT, P. B., BOSCH, R. J. and HUDGENS, M. G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59** 531–541. MR2004258 https://doi.org/10.1111/1541-0420.00063

HANSEN, B. B. (2004). Full matching in an observational study of coaching for the SAT. *J. Amer. Statist. Assoc.* **99** 609–618. MR2086387 https://doi.org/10.1198/016214504000000647

HANSEN, B. B., ROSENBAUM, P. R. and SMALL, D. S. (2014). Clustered treatment assignments and sensitivity to unmeasured biases in observational studies. *J. Amer. Statist. Assoc.* **109** 133–144. MR3180552 https://doi.org/10.1080/01621459.2013.863157

HELLAND, I. S. (1995). Simple counterexamples against the conditionality principle. *Amer. Statist.* **49** 351–356. MR1368487 https://doi.org/10.2307/2684572

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

HOSMAN, C. A., HANSEN, B. B. and HOLLAND, P. W. (2010). The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *Ann. Appl. Stat.* **4** 849–870. MR2758424 https://doi.org/10.1214/09-AOAS315

HSU, J. Y., SMALL, D. S. and ROSENBAUM, P. R. (2013). Effect modification and design sensitivity in observational studies. *J. Amer. Statist. Assoc.* **108** 135–148. MR3174608 https://doi.org/10.1080/01621459.2012.742018

KALBFLEISCH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62** 251–268. MR0386075 https://doi.org/10.1093/biomet/62.2.251

KEELE, L. and MINOZZI, W. (2013). How much is Minnesota like Wisconsin? Assumptions and conterfactuals in causal inference with observational data. *Polit. Anal.* **21** 193–216.

LANCASTER, H. O. (1961). The combination of probabilities: An application of orthonormal functions. *Aust. J. Stat.* **3** 20–33. MR0130742 https://doi.org/10.1111/j.1467-842x.1961.tb00058.x

LIPTÁK, T. (1958). On the combination of independent tests. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **3** 171–196.

LIU, W., KURAMOTO, J. and STUART, E. (2013). Sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prev. Sci.* **14** 570–580.

MORTON, D., SAAH, A., SILBERG, S., OWENS, W., ROBERTS, M. and SAAH, M. (1982). Lead absorption in children of emplyees in a lead-related factory. *Am. J. Epidemiol.* **115** 549–555.

PIMENTEL, S. D., KELZ, R. R., SILBER, J. H. and ROSENBAUM, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *J. Amer. Statist. Assoc.* **110** 515–527. MR3367244 https://doi.org/10.1080/01621459.2014.997879

POPPER, K. R. (1934). *Logik der Forschung*. Julius Springer Verlag, Vienna. MR1290323

POPPER, K. R. (1954). Degree of confirmation. *British J. Philos. Sci.* **5** 143–149. MR0065041 https://doi.org/10.1093/bjps/V.18.143

POPPER, K. R. (1963). *Conjectures and Refutations*: *The Growth of Scientific Knowledge*. Routledge and Kegan Paul, New York. MR0152413

POPPER, K. R. (1972). *The Logic of Scientific Discovery*, 6th impression revised. Hutchinson and Co., Ltd., London. MR0107593

REYNOLDS, K. D. and WEST, S. G. (1987). A multiplist strategy for strengthening nonequivalent control group designs. *Eval. Rev.* **11** 691–714.

ROSENBAUM, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74** 13–26. MR0885915 https://doi.org/10.1093/biomet/74.1.13

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

ROSENBAUM, P. R. (2004). Design sensitivity in observational studies. *Biometrika* **91** 153–164. MR2050466 https://doi.org/10.1093/biomet/91.1.153

ROSENBAUM, P. R. (2005). Observational study. In *Encyclopedia of Statistics in Behavioral Science* (B. S. Everitt and D. C. Howell, eds.) **3** 1451–1462.
ROSENBAUM, P. R. (2010). Design sensitivity and efficiency in observational studies. *J. Amer. Statist. Assoc.* **105** 692–702. MR2724853 https://doi.org/10.1198/jasa.2010.tm09570
ROSENBAUM, P. R. (2011). Some approximate evidence factors in observational studies. *J. Amer. Statist. Assoc.* **106** 285–295. MR2816721 https://doi.org/10.1198/jasa.2011.tm10422
ROSENBAUM, P. R. (2015). Bahadur efficiency of sensitivity analyses in observational studies. *J. Amer. Statist. Assoc.* **110** 205–217. MR3338497 https://doi.org/10.1080/01621459.2014.960968
ROSENBAUM, P. R. (2017). The general structure of evidence factors in observational studies. *Statist. Sci.* **32** 514–530. MR3730520 https://doi.org/10.1214/17-STS621
ROSENBAUM, P. R. (2018). Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. *Ann. Appl. Stat.* **12** 2312–2334. MR3875702 https://doi.org/10.1214/18-AOAS1153
ROSENBAUM, P. R. and KRIEGER, A. M. (1990). Sensitivity of two-sample permutation inferences in observational studies. *J. Amer. Statist. Assoc.* **85** 493–498.
ROWBOTTOM, D. P. (2013). Popper's measure of corroboration and $P(h|b)$. *British J. Philos. Sci.* **64** 739–745. MR3138871 https://doi.org/10.1093/bjps/axs029
RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.
SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5
SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 https://doi.org/10.1093/biomet/73.3.751
SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. MR1092986
SPRENGER, J. (2018). Two impossibility results for measures of corroboration. *British J. Philos. Sci.* **69** 139–159. MR3771377 https://doi.org/10.1093/bjps/axw016
STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, S. A. and WILLIAMS, R. M. JR. (1949). *The American Soldier, Vol.* 1: *Adjustment During Army Life*. Princeton Univ. Press, Princeton.
STUART, E. A., DUGOFF, E., ABRAMS, M., SALKEVER, D. and STEINWACHS, D. (2013). Estimating causal effects in observational studies using electronic health data: Challenges and (some) solutions. *EGEMs* **1** Article 4.
WANG, J. and OWEN, A. B. (2019). Admissibility in partial conjunction testing. *J. Amer. Statist. Assoc.* **114** 158–168. MR3941245 https://doi.org/10.1080/01621459.2017.1385465
WHITLOCK, M. C. (2005). Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **18** 1368–1373.
WONG, M., COOK, T. D. and STEINER, P. M. (2015). Adding design elements to improve time series designs: No child left behind as an example of causal pattern-matching. *J. Res. Educ. Eff.* **8** 245–279.
YU, B. B. and GASTWIRTH, J. L. (2005). Sensitivity analysis for trend tests: Application to the risk of radiation exposure. *Biostatistics* **6** 201–209.
ZAYKIN, D. V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24** 1836–1841.
ZAYKIN, D. V., ZHIVOTOVSKY, L. A., WESTFALL, P. H. and WEIR, B. S. (2002). Truncated product method for combining P-values. *Genet. Epidemiol.* **22** 170–185.
ZHAO, Q. (2019). On sensitivity value of pair-matched observational studies. *J. Amer. Statist. Assoc.* **114** 713–722. MR3963174 https://doi.org/10.1080/01621459.2018.1429277
ZUBIZARRETA, J. R., PAREDES, R. D. and ROSENBAUM, P. R. (2014). Matching for balance, pairing for heterogeneity in an observational study of the effectiveness of for-profit and not-for-profit high schools in Chile. *Ann. Appl. Stat.* **8** 204–231. MR3191988 https://doi.org/10.1214/13-AOAS713
ZUBIZARRETA, J. R., NEUMAN, M., SILBER, J. H. and ROSENBAUM, P. R. (2012). Contrasting evidence within and between institutions that provide treatment in an observational study of alternative forms of anesthesia. *J. Amer. Statist. Assoc.* **107** 901–915. MR3010879 https://doi.org/10.1080/01621459.2012.682533

The Annals of Statistics
2020, Vol. 48, No. 6, 3312–3335
https://doi.org/10.1214/19-AOS1930

FRÉCHET CHANGE-POINT DETECTION

BY PAROMITA DUBEY[*] AND HANS-GEORG MÜLLER[†]

Department of Statistics, University of California, Davis, [*]*pdubey@ucdavis.edu;* [†]*hgmueller@ucdavis.edu*

We propose a method to infer the presence and location of change-points in the distribution of a sequence of independent data taking values in a general metric space, where change-points are viewed as locations at which the distribution of the data sequence changes abruptly in terms of either its Fréchet mean, Fréchet variance or both. The proposed method is based on comparisons of Fréchet variances before and after putative change-point locations and does not require a tuning parameter, except for the specification of cut-off intervals near the endpoints where change-points are assumed not to occur. Our results include theoretical guarantees for consistency of the test under contiguous alternatives when a change-point exists and also for consistency of the estimated location of the change-point, if it exists, where, under the null hypothesis of no change-point, the limit distribution of the proposed scan function is the square of a standardized Brownian bridge. These consistency results are applicable for a broad class of metric spaces under mild entropy conditions. Examples include the space of univariate probability distributions and the space of graph Laplacians for networks. Simulation studies demonstrate the effectiveness of the proposed methods, both for inferring the presence of a change-point and estimating its location. We also develop theory that justifies bootstrap-based inference and illustrate the new approach with sequences of maternal fertility distributions and communication networks.

REFERENCES

AHIDAR-COUTRIX, A., GOUIC, T. L. and PARIS, Q. (2018). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendible geodesics. arXiv preprint arXiv:1806.02740.

ARCONES, M. A. (1998). A remark on approximate M-estimators. *Statist. Probab. Lett.* **38** 311–321. MR1631213 https://doi.org/10.1016/S0167-7152(98)00038-8

ARLOT, S., CELISSE, A. and HARCHAOUI, Z. (2012). Kernel change-point detection. arXiv preprint arXiv:1202.3878.

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

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11–16, 1992. MR1477909

CAZELLES, E., SEGUY, V., BIGOT, J., CUTURI, M. and PAPADAKIS, N. (2018). Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM J. Sci. Comput.* **40** B429–B456. MR3780753 https://doi.org/10.1137/17M1143459

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. MR3646580 https://doi.org/10.1080/01621459.2016.1147356

CHEN, J. and GUPTA, A. K. (2012). *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*, 2nd ed. Birkhäuser/Springer, New York. MR3025631 https://doi.org/10.1007/978-0-8176-4801-5

CHEN, H. and ZHANG, N. (2015). Graph-based change-point detection. *Ann. Statist.* **43** 139–176. MR3285603 https://doi.org/10.1214/14-AOS1269

MSC2020 subject classifications. Primary 60K35.

Key words and phrases. Bootstrap, Brownian bridge, dynamics of networks, empirical processes, graph Laplacians, metric space, object data, random densities, random objects.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Stat.* **35** 999–1018. MR0179874 https://doi.org/10.1214/aoms/1177700517

CHU, L. and CHEN, H. (2019). Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *Ann. Statist.* **47** 382–414. MR3910545 https://doi.org/10.1214/18-AOS1691

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis. Wiley Series in Probability and Statistics*. Wiley, Chichester. With a foreword by David Kendall. MR2743035

DE RIDDER, S., VANDERMARLIERE, B. and RYCKEBUSCH, J. (2016). Detection and localization of change points in temporal networks with the aid of stochastic block models. *J. Stat. Mech. Theory Exp.* **2016** 113302.

DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200 https://doi.org/10.1093/biomet/asz052

DUBEY, P. and MÜLLER, H.-G. (2020). Supplement to "Fréchet change-point detection." https://doi.org/10.1214/19-AOS1930SUPP

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

GARREAU, D. and ARLOT, S. (2018). Consistent change-point detection with kernels. *Electron. J. Stat.* **12** 4440–4486. MR3892345 https://doi.org/10.1214/18-ejs1513

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inf. Theory* **59** 1957–1965. MR3043776 https://doi.org/10.1109/TIT.2012.2235172

HELLSTRAND, J., MYRSKYLÄ, M., NISÉN, J. et al. (2019). All-time low period fertility in Finland: Drivers, tempo effects, and cohort implications. Technical Report, Max Planck Institute for Demographic Research, Rostock, Germany.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 71–83. MR0885920 https://doi.org/10.1093/biomet/74.1.71

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statist. Sinica* **2** 69–90. MR1152298

JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

KOHLER, H.-P., BILLARI, F. C. and ORTEGA, J. A. (2002). The emergence of lowest-low fertility in Europe during the 1990s. *Popul. Dev. Rev.* **28** 641–680.

KOHLER, H.-P., BILLARI, F. C. and ORTEGA, J. A. (2006). Low fertility in Europe: Causes, implications and policy options. In *The Baby Bust: Who Will do the Work* 48–109.

LUNG-YUT-FONG, A., LÉVY-LEDUC, C. and CAPPÉ, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *J. SFdS* **156** 133–162. MR3436651

MACNEILL, I. B. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion. *Ann. Statist.* **2** 950–962. MR0426253

MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345. MR3180567 https://doi.org/10.1080/01621459.2013.849605

MONTGOMERY-SMITH, S. J. and PRUSS, A. R. (2001). A comparison inequality for sums of independent random variables. *J. Math. Anal. Appl.* **254** 35–42. MR1807885 https://doi.org/10.1006/jmaa.2000.7200

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919. MR0893905

PEEL, L. and CLAUSET, A. (2015). Detecting change points in the large scale structure of evolving networks. In *AAAI* **15** 1–11.

PETERSEN, A. and MÜLLER, H.-G. (2019a). Wasserstein covariance for multiple random densities. *Biometrika* **106** 339–351. MR3949307 https://doi.org/10.1093/biomet/asz005

PETERSEN, A. and MÜLLER, H.-G. (2019b). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Int. Stat. Rev.* **56** 31–48. MR0963139 https://doi.org/10.2307/1403360

SRIVASTAVA, M. S. and WORSLEY, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.* **81** 199–204. MR0830581

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds*, *Graphs*, *and Metric Spaces* (*Paris*, 2002). *Contemp*. *Math*. **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 https://doi.org/10.1090/conm/338/06080

STURM, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math*. **196** 65–131. MR2237206 https://doi.org/10.1007/s11511-006-0002-8

TAVAKOLI, S., PIGOLI, D., ASTON, J. A. D. and COLEMAN, J. S. (2019). A spatial modeling approach for linguistic object data: Analyzing dialect sound variations across Great Britain. *J. Amer. Statist. Assoc*. **114** 1081–1096. MR4011760 https://doi.org/10.1080/01621459.2019.1607357

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

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WANG, D., YU, Y. and RINALDO, A. (2018). Optimal change point detection and localization in sparse dynamic networks. arXiv preprint arXiv:1809.09602.

The Annals of Statistics
2020, Vol. 48, No. 6, 3336–3365
https://doi.org/10.1214/19-AOS1933

NONPARAMETRIC DRIFT ESTIMATION FOR I.I.D. PATHS OF STOCHASTIC DIFFERENTIAL EQUATIONS

BY FABIENNE COMTE[*] AND VALENTINE GENON-CATALOT[†]

MAP5 UMR8145, Université de Paris, [*]*fabienne.comte@parisdescartes.fr;* [†]*valentine.genon-catalot@parisdescartes.fr*

We consider N independent stochastic processes $(X_i(t), t \in [0, T])$, $i = 1, \dots, N$, defined by a one-dimensional stochastic differential equation, which are continuously observed throughout a time interval $[0, T]$ where T is fixed. We study nonparametric estimation of the drift function on a given subset A of $\mathbb{R}$. Projection estimators are defined on finite dimensional subsets of $\mathbb{L}^2(A, dx)$. We stress that the set A may be compact or not and the diffusion coefficient may be bounded or not. A data-driven procedure to select the dimension of the projection space is proposed where the dimension is chosen within a random collection of models. Upper bounds of risks are obtained, the assumptions are discussed and simulation experiments are reported.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas*, *Graphs*, *and Mathematical Tables*. *National Bureau of Standards Applied Mathematics Series* **55**. U.S. Government Printing Office, Washington, D.C. MR0167642

BARAUD, Y., COMTE, F. and VIENNET, G. (2001). Model selection for (auto-)regression with dependent data. *ESAIM Probab*. *Stat*. **5** 33–49. MR1845321 https://doi.org/10.1051/ps:2001101

BIRGÉ, L. and MASSART, P. (2007). Minimal penalties for Gaussian model selection. *Probab*. *Theory Related Fields* **138** 33–73. MR2288064 https://doi.org/10.1007/s00440-006-0011-8

CHALEYAT-MAUREL, M. and GENON-CATALOT, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process*. *Appl*. **116** 1447–1467. MR2260743 https://doi.org/10.1016/j.spa.2006.03.004

COHEN, A., DAVENPORT, M. A. and LEVIATAN, D. (2013). On the stability and accuracy of least squares approximations. *Found*. *Comput*. *Math*. **13** 819–834. MR3105946 https://doi.org/10.1007/s10208-013-9142-3

COMTE, F. and GENON-CATALOT, V. (2018). Laguerre and Hermite bases for inverse problems. *J*. *Korean Statist*. *Soc*. **47** 273–296. MR3840862 https://doi.org/10.1016/j.jkss.2018.03.001

COMTE, F. and GENON-CATALOT, V. (2019). Regression function estimation on non compact support as a partly inverse problem. *Ann*. *Inst*. *Statist*. *Math*. To appear. Available at https://doi.org/10.1007/s10463-019-00718-2.

COMTE, F. and GENON-CATALOT, V. (2020). Regression function estimation on non compact support in an heteroscesdastic model. *Metrika* **83** 93–128. MR4054254 https://doi.org/10.1007/s00184-019-00727-4

COMTE, F., GENON-CATALOT, V. and ROZENHOLC, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262 https://doi.org/10.3150/07-BEJ5173

COMTE, F., GENON-CATALOT, V. and SAMSON, A. (2013). Nonparametric estimation for stochastic differential equations with random effects. *Stochastic Process*. *Appl*. **123** 2522–2551. MR3054535 https://doi.org/10.1016/j.spa.2013.04.009

DALALYAN, A. S. and KUTOYANTS, YU. A. (2003). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math*. *Methods Statist*. **11** 402–427. MR1979742

DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab*. *Theory Related Fields* **134** 248–282. MR2222384 https://doi.org/10.1007/s00440-004-0416-1

DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab*. *Theory Related Fields* **137** 25–47. MR2278451 https://doi.org/10.1007/s00440-006-0502-7

MSC2020 subject classifications. 62G07, 62M05.

Key words and phrases. Diffusion process, Hermite basis, Laguerre basis, model selection, nonparametric drift estimation, projection estimators.

DELATTRE, M., GENON-CATALOT, V. and LARÉDO, C. (2018). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic Process. Appl.* **128** 1929–1957. MR3797649 https://doi.org/10.1016/j.spa.2017.08.016

DELATTRE, M., GENON-CATALOT, V. and SAMSON, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scand. J. Stat.* **40** 322–343. MR3066417 https://doi.org/10.1111/j.1467-9469.2012.00813.x

DELATTRE, M. and LAVIELLE, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat. Interface* **6** 519–532. MR3164656 https://doi.org/10.4310/SII.2013.v6.n4.a10

DENIS, C., DION, C. and MARTINEZ, M. (2018). Consistent procedures for multiclass classification of discrete diffusion paths. *Scand. J. Stat.*. To appear. https://doi.org/10.1111/sjos.12415

DION, C. and GENON-CATALOT, V. (2016). Bidimensional random effect estimation in mixed stochastic differential model. *Stat. Inference Stoch. Process.* **19** 131–158. MR3506629 https://doi.org/10.1007/s11203-015-9122-0

DITLEVSEN, S. and DE GAETANO, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT* **3** 137–153. MR2259358

HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** 135–163. MR1670522 https://doi.org/10.1016/S0304-4149(98)00074-X

HSIAO, C. (2003). *Analysis of Panel Data*, 2nd ed. *Econometric Society Monographs* **34**. Cambridge Univ. Press, Cambridge. MR1962511 https://doi.org/10.1017/CBO9780511754203

IACUS, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations*: *With R Examples*. *Springer Series in Statistics*. Springer, New York. MR2410254 https://doi.org/10.1007/978-0-387-75839-8

INDRITZ, J. (1961). An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.* **12** 981–983. MR0132852 https://doi.org/10.2307/2034406

KESSLER, M., LINDNER, A. and SØRENSEN, M., eds. (2012). *Statistical Methods for Stochastic Differential Equations*. *Monographs on Statistics and Applied Probability* **124**. CRC Press, Boca Raton, FL. MR2975799

KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes*. *Research and Exposition in Mathematics* **6**. Heldermann, Berlin. MR0777685

KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. Springer, London. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

OVERGAARD, R., JONSSON, N., TORNØE, C. and MADSEN, H. (2005). Non-linear mixed effects models with stochastic differential equations: Implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.* **32** 85–107.

PICCHINI, U., DE GAETANO, A. and DITLEVSEN, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Stat.* **37** 67–90. MR2675940 https://doi.org/10.1111/j.1467-9469.2009.00665.x

PICCHINI, U. and DITLEVSEN, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55** 1426–1444. MR2741425 https://doi.org/10.1016/j.csda.2010.10.003

RAMSAY, J. O. and SILVERMAN, B. W. (2007). *Applied Functional Data Analysis*: *Methods and Case Studies* Springer, New York.

REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

ROGERS, L. C. G. (1985). Smooth transition densities for one-dimensional diffusions. *Bull. Lond. Math. Soc.* **17** 157–161. MR0806242 https://doi.org/10.1112/blms/17.2.157

ROGERS, L. C. G. and WILLIAMS, D. (1990). *Diffusions*, *Markov Processes*, *and Martingales*. *Vol.* 2: *Ito Calculus*. Wiley, Chichester.

SZEGŐ, G. (1975). *Orthogonal Polynomials*: *Colloquium Publications*, *Vol. XXIII*, 4th ed. Amer. Math. Soc., Providence, R.I. MR0372517

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

WANG, J.-L., CHIOU, J.-M. and MUELLER, H.-G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

The Annals of Statistics
2020, Vol. 48, No. 6, 3366–3394
https://doi.org/10.1214/19-AOS1934

DISTANCE-BASED AND RKHS-BASED DEPENDENCE METRICS IN HIGH DIMENSION

BY CHANGBO ZHU[1,*], XIANYANG ZHANG[2], SHUN YAO[3] AND XIAOFENG SHAO[1,†]

[1]*Department of Statistics, University of Illinois at Urbana-Champaign,* [*]*changbo2@illinois.edu;* [†]*xshao@illinois.edu*

[2]*Department of Statistics, Texas A&M University, zhangxiany@stat.tamu.edu*

[3]*Quantitative Researcher, Point72 Asset Management, cnshunyao@gmail.com*

In this paper, we study distance covariance, Hilbert–Schmidt covariance (aka Hilbert–Schmidt independence criterion [In *Advances in Neural Information Processing Systems* (2008) 585–592]) and related independence tests under the high dimensional scenario. We show that the sample distance/Hilbert–Schmidt covariance between two random vectors can be approximated by the sum of squared componentwise sample cross-covariances up to an asymptotically constant factor, which indicates that the standard distance/Hilbert–Schmidt covariance based test can only capture linear dependence in high dimension. Under the assumption that the components within each high dimensional vector are weakly dependent, the distance correlation based t test developed by Székely and Rizzo (*J. Multivariate Anal.* **117** (2013) 193–213) for independence is shown to have trivial limiting power when the two random vectors are nonlinearly dependent but componentwisely uncorrelated. This new and surprising phenomenon, which seems to be discovered and carefully studied for the first time, is further confirmed in our simulation study. As a remedy, we propose tests based on an aggregation of marginal sample distance/Hilbert–Schmidt covariances and show their superior power behavior against their joint counterparts in simulations. We further extend the distance correlation based t test to those based on Hilbert–Schmidt covariance and marginal distance/Hilbert–Schmidt covariance. A novel unified approach is developed to analyze the studentized sample distance/Hilbert–Schmidt covariance as well as the studentized sample marginal distance covariance under both null and alternative hypothesis. Our theoretical and simulation results shed light on the limitation of distance/Hilbert–Schmidt covariance when used jointly in the high dimensional setting and suggest the aggregation of marginal distance/Hilbert–Schmidt covariance as a useful alternative.

REFERENCES

AHN, J., MARRON, J. S., MULLER, K. M. and CHI, Y.-Y. (2007). The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94** 760–766. MR2410023 https://doi.org/10.1093/biomet/asm050

AOSHIMA, M., SHEN, D., SHEN, H., YATA, K., ZHOU, Y.-H. and MARRON, J. S. (2018). A survey of high dimension low sample size asymptotics. *Aust. N. Z. J. Stat.* **60** 4–19. MR3780619 https://doi.org/10.1111/anzs.12212

BEDO, J. (2008). Microarray design using the Hilbert–Schmidt independence criterion. In *IAPR International Conference on Pattern Recognition in Bioinformatics* 288–298. Springer, Berlin.

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

MSC2020 subject classifications. Primary 62G10, 60K35; secondary 62G20.
Key words and phrases. Distance covariance, high dimensionality, Hilbert–Schmidt independence criterion, independence test, $\mathcal{U}$-statistics.

BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions. Vol.* 1. Kendrick Press, Heber City, UT. MR2325295

CHAKRABORTY, S. and ZHANG, X. (2019a). Distance metrics for measuring joint dependence with application to causal inference. *J. Amer. Statist. Assoc.* **114** 1638–1650. MR4047289 https://doi.org/10.1080/01621459.2018.1513364

CHAKRABORTY, S. and ZHANG, X. (2019b). A new framework for distance and kernel-based metrics in high dimensions. Available at https://arxiv.org/abs/1909.13469.

CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290–299. MR0793494 https://doi.org/10.1016/0047-259X(85)90022-3

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics*. Springer, New York. MR2664452

DAU, H. A., KEOGH, E., KAMGAR, K., YEH, C.-C. M., ZHU, Y., GHARGHABI, S., RATANAMAHATANA, C. A., YANPING, HU, B. et al. (2018). The UCR time series classification archive. Available at https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

DAVIDSON, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians. Advanced Texts in Econometrics*. Oxford Univ. Press, New York. MR1430804 https://doi.org/10.1093/0198774036.001.0001

DE WET, T. (1980). Cramér–von Mises tests for independence. *J. Multivariate Anal.* **10** 38–50. MR0569795 https://doi.org/10.1016/0047-259X(80)90080-9

DEHEUVELS, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* **11** 102–113. MR0612295 https://doi.org/10.1016/0047-259X(81)90136-6

DOUKHAN, P. and NEUMANN, M. H. (2008). The notion of ψ-weak dependence and its applications to bootstrapping time series. *Probab. Surv.* **5** 146–168. MR2426177 https://doi.org/10.1214/06-PS086

DUECK, J., EDELMANN, D., GNEITING, T. and RICHARDS, D. (2014). The affinely invariant distance correlation. *Bernoulli* **20** 2305–2330. MR3263106 https://doi.org/10.3150/13-BEJ558

EDELMANN, D., RICHARDS, D. and VOGEL, D. (2017). The distance standard deviation. Preprint. Available at arXiv:1705.05777.

GALLANT, A. R. and WHITE, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Blackwell, Oxford.

GIESER, P. W. and RANDLES, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.* **92** 561–567. MR1467849 https://doi.org/10.2307/2965704

GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 585–592.

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

HALL, P., MARRON, J. S. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 427–444. MR2155347 https://doi.org/10.1111/j.1467-9868.2005.00510.x

HETTMANSPERGER, T. P. and OJA, H. (1994). Affine invariant multivariate multisample sign tests. *J. Roy. Statist. Soc. Ser. B* **56** 235–249. MR1257810

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

HUA, W.-Y. and GHOSH, D. (2015). Equivalence of kernel machine regression and kernel distance covariance for multidimensional phenotype association studies. *Biometrics* **71** 812–820. MR3402617 https://doi.org/10.1111/biom.12314

JUNG, S. and MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

KONG, J., KLEIN, B. E., KLEIN, R., LEE, K. E. and WAHBA, G. (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci. USA* **109** 20352–20357.

KROUPI, E., YAZDANI, A., VESIN, J.-M. and EBRAHIMI, T. (2012). Multivariate spectral analysis for identifying the brain activations during olfactory perception. In 2012 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 6172–6175. IEEE, New York.

KROUPI, E., YAZDANI, A., VESIN, J.-M. and EBRAHIMI, T. (2014). EEG correlates of pleasant and unpleasant odor perception. *ACM Trans. Multimed. Comput. Commun. Appl.* **11** 13.

LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900 https://doi.org/10.1080/01621459.2012.695654

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* **112** 623–637. MR3671757 https://doi.org/10.1080/01621459.2016.1150851

MIKALSEN, K. Ø., SOGUERO-RUIZ, C., BIANCHI, F. M. and JENSSEN, R. (2019). Noisy multi-label semi-supervised dimensionality reduction. *Pattern Recognit.* **90** 257–270.

PAN, G., GAO, J. and YANG, Y. (2014). Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.* **109** 600–612. MR3223736 https://doi.org/10.1080/01621459.2013.872037

PARK, T., SHAO, X. and YAO, S. (2015). Partial martingale difference correlation. *Electron. J. Stat.* **9** 1492–1517. MR3367668 https://doi.org/10.1214/15-EJS1047

RAMDAS, A., REDDI, S. J., PÓCZOS, B., SINGH, A. and WASSERMAN, L. (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. MR3265698 https://doi.org/10.1080/01621459.2014.887012

SINHA, B. K. and WIEAND, H. S. (1977). Multivariate nonparametric tests for independence. *J. Multivariate Anal.* **7** 572–583. MR0471173 https://doi.org/10.1016/0047-259X(77)90068-9

SZEKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *InterStat* **5** 1–6.

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 https://doi.org/10.1214/14-AOS1255

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 https://doi.org/10.1016/S0167-7152(02)00399-1

WEI, S., LEE, C., WICHERS, L. and MARRON, J. S. (2016). Direction-projection-permutation for high-dimensional hypothesis tests. *J. Comput. Graph. Statist.* **25** 549–569. MR3499694 https://doi.org/10.1080/10618600.2015.1027773

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

XU, J., LIU, J., YIN, J. and SUN, C. (2016). A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously. *Knowl.-Based Syst.* **98** 172–184.

YANG, Y. (2017). Source-space analyses in MEG/EEG and applications to explore spatio-temporal neural dynamics in human vision. Ph.D. thesis.

YANG, Y. and PAN, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187 https://doi.org/10.1214/14-AOS1284

YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

ZHANG, X., YAO, S. and SHAO, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Ann. Statist.* **46** 219–246. MR3766951 https://doi.org/10.1214/17-AOS1548

ZHANG, Y. and ZHOU, Z.-H. (2010). Multilabel dimensionality reduction via dependence maximization. *ACM Trans. Knowl. Discov. Data* **4** 14.

ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. MR2915095 https://doi.org/10.1111/j.1467-9892.2011.00780.x

ZHU, C. and SHAO, X. (2019). Interpoint distance based two sample tests in high dimension. Available at https://arxiv.org/pdf/1902.07279.pdf.

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

ZHU, C., ZHANG, X., YAO, S. and SHAO, X. (2020). Supplement to "Distance-based and RKHS-based dependence metrics in high dimension." https://doi.org/10.1214/19-AOS1934SUPP.

The Annals of Statistics
2020, Vol. 48, No. 6, 3395–3416
https://doi.org/10.1214/19-AOS1935

THE DISTANCE STANDARD DEVIATION

BY DOMINIC EDELMANN,[1] DONALD RICHARDS[2] AND DANIEL VOGEL[3]

[1]*German Cancer Research Center, dominic.edelmann@dkfz-heidelberg.de*
[2]*Department of Statistics, Pennsylvania State University, richards@stat.psu.edu*
[3]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, daniel.vogel@tu-dortmund.de*

The distance standard deviation, which arises in distance correlation analysis of multivariate data, is studied as a measure of spread. The asymptotic distribution of the empirical distance standard deviation is derived under the assumption of finite second moments. Applications are provided to hypothesis testing on a data set from materials science and to multivariate statistical quality control. The distance standard deviation is compared to classical scale measures for inference on the spread of heavy-tailed distributions. Inequalities for the distance variance are derived, proving that the distance standard deviation is bounded above by the classical standard deviation and by Gini's mean difference. New expressions for the distance standard deviation are obtained in terms of Gini's mean difference and the moments of spacings of order statistics. It is also shown that the distance standard deviation satisfies the axiomatic properties of a measure of spread.

REFERENCES

[1] APARISI, F., JABALOYES, J. and CARRIÓN, A. (1999). Statistical properties of the $|S|$ multivariate control chart. *Comm. Statist. Theory Methods* **28** 2671–2686. MR1729896 https://doi.org/10.1080/03610929908832445

[2] ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Stat.* **40** 2076–2100. MR0264805 https://doi.org/10.1214/aoms/1177697287

[3] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

[4] BICKEL, P. J. and LEHMANN, E. L. (2012). Descriptive statistics for nonparametric models, III. Dispersion. In *Selected Works of E. L. Lehmann* (J. Rojo, ed.) 499–518. *Selected Works in Probability and Statistics*. Springer, New York. MR2920510

[5] BICKEL, P. J. and LEHMANN, E. L. (2012). Descriptive statistics for nonparametric models, IV. Spread. In *Selected Works of E. L. Lehmann*. *Selected Works in Probability and Statistics* (J. Rojo, ed.) 519–526. Springer, New York. MR2920510

[6] BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2019). Distance multivariance: New dependence measures for random vectors. *Ann. Statist.* **47** 2757–2789. MR3988772 https://doi.org/10.1214/18-AOS1764

[7] CHANDRA, M. J. (2001). *Statistical Quality Control*. CRC Press, Melbourne, FL.

[8] DUECK, J., EDELMANN, D., GNEITING, T. and RICHARDS, D. (2014). The affinely invariant distance correlation. *Bernoulli* **20** 2305–2330. MR3263106 https://doi.org/10.3150/13-BEJ558

[9] DUECK, J., EDELMANN, D. and RICHARDS, D. (2015). A generalization of an integral arising in the theory of distance correlation. *Statist. Probab. Lett.* **97** 116–119. MR3299759 https://doi.org/10.1016/j.spl.2014.11.012

[10] DUECK, J., EDELMANN, D. and RICHARDS, D. (2017). Distance correlation coefficients for Lancaster distributions. *J. Multivariate Anal.* **154** 19–39. MR3588555 https://doi.org/10.1016/j.jmva.2016.10.012

[11] EDELMANN, D. and FIEDLER, J. (2019). dcortools: Providing fast and flexible functions for distance correlation analysis. R package version 0.0.0.9000.

MSC2020 subject classifications. Primary 60E15, 62H20; secondary 60E05, 60E10.

Key words and phrases. Asymptotic efficiency, characteristic function, dispersive ordering, distance correlation coefficient, distance variance, Gini's mean difference, measure of spread, order statistic, sample spacing, stochastic ordering, U-statistic.

[12] EDELMANN, D., RICHARDS, D. and VOGEL, D. (2020). Supplement to "The distance standard deviation." https://doi.org/10.1214/19-AOS1935SUPP.

[13] FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433.

[14] FIEDLER, J. (2016). Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields. Doctoral dissertation, Unive. Heidelberg.

[15] FOKIANOS, K. and PITSILLOU, M. (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59** 262–270. MR3635048 https://doi.org/10.1080/00401706.2016.1156024

[16] GENE ONTOLOGY CONSORTIUM (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32** D258–D261. https://doi.org/10.1093/nar/gkh036

[17] GERSTENBERGER, C. and VOGEL, D. (2015). On the efficiency of Gini's mean difference. *Stat. Methods Appl.* **24** 569–596. MR3421674 https://doi.org/10.1007/s10260-015-0315-x

[18] GOEMAN, J. J. and BÜHLMANN, P. (2007). Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics* **23** 980–987.

[19] GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Measuring statistical dependence with Hilbert–Schmidt norms. *Adv. Neural Inf. Process. Syst.* 585–592.

[20] HUANG, C. and HUO, X. (2017). A statistically and numerically efficient independence test based on random projections and distance covariance. Preprint. Available at arXiv:1701.06054.

[21] HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. MR3556612 https://doi.org/10.1080/00401706.2015.1054435

[22] JENTSCH, C., LEUCHT, A., MEYER, M. and BEERING, C. (2020). Empirical characteristic functions-based estimation and distance correlation for locally stationary processes. *J. Time Series Anal.* **41** 110–133. MR4048684

[23] KONG, J., KLEIN, B. E. K., KLEIN, R. and WAHBA, G. (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci. USA* **109** 20352–20357.

[24] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

[25] MARTÍNEZ-GÓMEZ, E., RICHARDS, M. T. and RICHARDS, D. ST. P. (2014). Distance correlation methods for discovering associations in large astrophysical databases. *Astrophys. J.* **781** Art. ID 39.

[26] MONTGOMERY, D. C. (2009). *Introduction to Statistical Quality Control*. Wiley, New York.

[27] PAN, W., WANG, X., ZHANG, H., ZHU, H. and ZHU, J. (2020). Ball covariance: A generic measure of dependence in Banach space. *J. Amer. Statist. Assoc.* **115** 307–317. MR4078465 https://doi.org/01621459.2018.1543600

[28] PEREZ-RECHE, F. J., TRIGUERO, C., ZANZOTTO, G. and TRUSKINOVSKY, L. (2016). Origin of scale-free intermittency in structural first-order phase transitions. *Phys. Rev. B* **94** 144102.

[29] RAMDAS, A., REDDI, S. J., PÓCZOS, B., SINGH, A. and WASSERMAN, L. (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[30] RICHARDS, M. T., RICHARDS, D. ST. P. and MARTÍNEZ-GÓMEZ, E. (2014). Interpreting the distance correlation results for the COMBO-17 survey. *Astrophys. J. Lett.* **784** Art. ID L34.

[31] RIZZO, M. L. and SZÉKELY, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Stat.* **4** 1034–1055. MR2758432 https://doi.org/10.1214/09-AOAS245

[32] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

[33] SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

[34] SIMON, N. and TIBSHIRANI, R. (2014). Comment on "Detecting novel associations in large data sets" by Reshef et al Science Dec 16, 2011. Preprint. Available at arXiv:1401.7645.

[35] SZÉKELY, G. J. and RIZZO, M. (2019). energy: E-Statistics: Multivariate inference via the energy of data. R package version 1.7-6.

[36] SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

[37] SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

[38] SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 https://doi.org/10.1214/14-AOS1255

[39] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

[40] TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* 448–485. Stanford Univ. Press, Stanford, CA. MR0120720

[41] 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

[42] YITZHAKI, S. (2003). Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron* **61** 285–316. MR2025523

[43] YITZHAKI, S. and SCHECHTMAN, E. (2013). *The Gini Methodology*: *A Primer on a Statistical Methodology*. *Springer Series in Statistics*. Springer, New York. MR3012052 https://doi.org/10.1007/978-1-4614-4720-7

[44] ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J*. *Time Series Anal*. **33** 438–457. MR2915095 https://doi.org/10.1111/j.1467-9892.2011.00780.x

The Annals of Statistics
2020, Vol. 48, No. 6, 3417–3441
https://doi.org/10.1214/19-AOS1936

ROBUST MULTIVARIATE NONPARAMETRIC TESTS VIA PROJECTION AVERAGING

BY ILMUN KIM[*], SIVARAMAN BALAKRISHNAN[†] AND LARRY WASSERMAN[‡]

Department of Statistics & Data Science, Carnegie Mellon University, [*]*ilmunk@stat.cmu.edu;* [†]*siva@stat.cmu.edu;* [‡]*larry@stat.cmu.edu*

In this work, we generalize the Cramér–von Mises statistic via projection averaging to obtain a robust test for the multivariate two-sample problem. The proposed test is consistent against all fixed alternatives, robust to heavy-tailed data and minimax rate optimal against a certain class of alternatives. Our test statistic is completely free of tuning parameters and is computationally efficient even in high dimensions. When the dimension tends to infinity, the proposed test is shown to have comparable power to the existing high-dimensional mean tests under certain location models. As a by-product of our approach, we introduce a new metric called *the angular distance* which can be thought of as a robust alternative to the Euclidean distance. Using the angular distance, we connect the proposed method to the reproducing kernel Hilbert space approach. In addition to the Cramér–von Mises statistic, we demonstrate that the projection-averaging technique can be used to define robust multivariate tests in many other problems.

REFERENCES

ANDERSON, T. W. (1962). On the distribution of the two-sample Cramér–von Mises criterion. *Ann. Math. Stat.* **33** 1148–1159. MR0145607 https://doi.org/10.1214/aoms/1177704477

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR1990662

ANDERSON, N. H., HALL, P. and TITTERINGTON, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50** 41–54. MR1292607 https://doi.org/10.1006/jmva.1994.1033

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

BARINGHAUS, L. and HENZE, N. (2017). Cramér–von Mises distance: Probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *J. Nonparametr. Stat.* **29** 167–188. MR3635009 https://doi.org/10.1080/10485252.2017.1285029

BERA, A. K., GHOSH, A. and XIAO, Z. (2013). A smooth test for the equality of distributions. *Econometric Theory* **29** 419–446. MR3042761 https://doi.org/10.1017/S0266466612000370

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BHAT, B. V. (1995). Theory of U-statistics and its applications. Ph.D. thesis, Karnatak Univ.

BHATTACHARYA, B. B. (2018). Two-sample tests based on geometric graphs: Asymptotic distribution and detection thresholds. Preprint. Available at arXiv:1512.00384v3.

BHATTACHARYA, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 575–602. MR3961499

BISWAS, M. and GHOSH, A. K. (2014). A nonparametric two-sample test applicable to high dimensional data. *J. Multivariate Anal.* **123** 160–171. MR3130427 https://doi.org/10.1016/j.jmva.2013.09.004

BISWAS, M., MUKHOPADHYAY, M. and GHOSH, A. K. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **101** 913–926. MR3286925 https://doi.org/10.1093/biomet/asu045

MSC2020 subject classifications. Primary 62G10, 62H15, 62H20; secondary 62G35.

Key words and phrases. Energy statistic, high dimension and low sample size, independence testing, maximum mean discrepancy, permutation tests, U-statistic.

CHAKRABORTY, A. and CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.* **45** 771–799. MR3650400 https://doi.org/10.1214/16-AOS1467

CHEN, H., CHEN, X. and SU, Y. (2018). A weighted edge-count two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **113** 1146–1155. MR3862346 https://doi.org/10.1080/01621459.2017.1307757

CHEN, L., DOU, W. W. and QIAO, Z. (2013). Ensemble subsampling for imbalanced multivariate two-sample tests. *J. Amer. Statist. Assoc.* **108** 1308–1323. MR3174710 https://doi.org/10.1080/01621459.2013.800763

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. MR3646580 https://doi.org/10.1080/01621459.2016.1147356

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHIKKAGOUDAR, M. S. and BHAT, B. V. (2014). Limiting distribution of two-sample degenerate U-statistic under contiguous alternatives and applications. *J. Appl. Statist. Sci.* **22** 127–139. MR3616873

CHUNG, E. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. MR3099111 https://doi.org/10.1214/13-AOS1090

CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. Aktuarietidskr.* **11** 141–180.

CUI, H. (2002). Average projection type weighted Cramér–von Mises statistics for testing some distributions. *Sci. China Ser. A* **45** 562–577. MR1911172

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527 https://doi.org/10.1017/S0266466606060506

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

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

HALL, P., MARRON, J. S. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 427–444. MR2155347 https://doi.org/10.1111/j.1467-9868.2005.00510.x

HARCHAOUI, Z., BACH, F., CAPPE, O. and MOULINES, E. (2013). Kernel-based methods for hypothesis testing: A unified view. *IEEE Signal Process. Mag.* **30** 87–97.

HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. MR0947577 https://doi.org/10.1214/aos/1176350835

HETTMANSPERGER, T. P., MÖTTÖNEN, J. and OJA, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica* **8** 785–800. MR1651508

HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Stat.* **23** 169–192. MR0057521 https://doi.org/10.1214/aoms/1177729436

HU, J. and BAI, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Sci. China Math.* **59** 2281–2300. MR3578957 https://doi.org/10.1007/s11425-016-0131-0

KANAMORI, T., SUZUKI, T. and SUGIYAMA, M. (2012). f-divergence estimation and two-sample homogeneity test under semiparametric density-ratio models. *IEEE Trans. Inform. Theory* **58** 708–720. MR2917977 https://doi.org/10.1109/TIT.2011.2163380

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Supplement to "Robust multivariate nonparametric tests via projection averaging." https://doi.org/10.1214/19-AOS1936SUPP.

KRUSKAL, J. B. JR. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* **7** 48–50. MR0078686 https://doi.org/10.2307/2033241

LEE, A. J. (1990). *U-Statistics: Theory and Practice. Statistics: Textbooks and Monographs* **110**. Dekker, New York. MR1075417

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

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938 https://doi.org/10.1214/12-AOS993

LIU, R. Y. (2006). *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications* **72**. Amer. Math. Soc., Providence.

LIU, Z. and MODARRES, R. (2011). A triangle test for equality of distribution functions in high dimensions. *J. Nonparametr. Stat.* **23** 605–615. MR2836279 https://doi.org/10.1080/10485252.2010.485644

LOPEZ-PAZ, D. and OQUAB, M. (2016). Revisiting classifier two-sample tests. Preprint. Available at arXiv:1610.06545.

MONDAL, P. K., BISWAS, M. and GHOSH, A. K. (2015). On high dimensional two-sample tests based on nearest neighbors. *J. Multivariate Anal.* **141** 168–178. MR3390065 https://doi.org/10.1016/j.jmva.2015.07.002

MUKHOPADHYAY, S. and WANG, K. (2018). A nonparametric approach to high-dimensional K-sample comparison problem. Preprint. Available at arXiv:1810.01724.

OJA, H. (2010). *Multivariate Nonparametric Methods with R*: *An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

OJA, H. and RANDLES, R. H. (2004). Multivariate nonparametric tests. *Statist*. *Sci*. **19** 598–605. MR2185581 https://doi.org/10.1214/088342304000000558

PAN, W., TIAN, Y., WANG, X. and ZHANG, H. (2018). Ball divergence: Nonparametric two sample test. *Ann*. *Statist*. **46** 1109–1137. MR3797998 https://doi.org/10.1214/17-AOS1579

PESARIN, F. (2001). *Multivariate Permutation Tests*: *With Applications in Biostatistics*. Wiley, Chichester. MR1855501

RAMDAS, A., REDDI, S. J., POCZOS, B., SINGH, A. and WASSERMAN, L. (2015). Adaptivity and computation-statistics tradeoffs for kernel and distance based high dimensional two sample testing. Preprint. Available at arXiv:1508.00655.

ROSENBAUM, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **67** 515–530. MR2168202 https://doi.org/10.1111/j.1467-9868.2005.00513.x

SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J*. *Amer*. *Statist*. *Assoc*. **81** 799–806. MR0860514

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann*. *Statist*. **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *Interstate* **5**.

SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J*. *Statist*. *Plann*. *Inference* **143** 1249–1272. MR3055745 https://doi.org/10.1016/j.jspi.2013.03.018

THAS, O. (2010). *Comparing Distributions*. *Springer Series in Statistics*. Springer, New York. MR2547894 https://doi.org/10.1007/b99044

WALD, A. and WOLFOWITZ, J. (1940). On a test whether two samples are from the same population. *Ann*. *Math*. *Stat*. **11** 147–162. MR0002083 https://doi.org/10.1214/aoms/1177731909

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J*. *Amer*. *Statist*. *Assoc*. **110** 1658–1669. MR3449062 https://doi.org/10.1080/01621459.2014.988215

ZHOU, W.-X., ZHENG, C. and ZHANG, Z. (2017). Two-sample smooth tests for the equality of distributions. *Bernoulli* **23** 951–989. MR3606756 https://doi.org/10.3150/15-BEJ766

ZHU, L.-X., FANG, K.-T. and BHATTI, M. I. (1997). On estimated projection pursuit-type Crámer–von Mises statistics. *J*. *Multivariate Anal*. **63** 1–14. MR1491563 https://doi.org/10.1006/jmva.1997.1673

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

The Annals of Statistics
2020, Vol. 48, No. 6, 3442–3464
https://doi.org/10.1214/19-AOS1937

INFERENCE FOR CONDITIONAL VALUE-AT-RISK OF A PREDICTIVE REGRESSION

BY YI HE[1], YANXI HOU[2], LIANG PENG[3] AND HAIPENG SHEN[3]

[1]*Amsterdam School of Economics, University of Amsterdam,* y.he2@uva.nl

[2]*School of Data Science, Fudan University,* yxhou@fudan.edu.cn

[3]*Department of Risk Management and Insurance, Georgia State University,* lpeng@gsu.edu

[4]*Faculty of Business and Economics, The University of Hong Kong,* haipeng@hku.hk

Conditional value-at-risk is a popular risk measure in risk management. We study the inference problem of conditional value-at-risk under a linear predictive regression model. We derive the asymptotic distribution of the least squares estimator for the conditional value-at-risk. Our results relax the model assumptions made in (*Oper. Res.* **60** (2012) 739–756) and correct their mistake in the asymptotic variance expression. We show that the asymptotic variance depends on the quantile density function of the unobserved error and whether the model has a predictor with infinite variance, which makes it challenging to actually quantify the uncertainty of the conditional risk measure. To make the inference feasible, we then propose a smooth empirical likelihood based method for constructing a confidence interval for the conditional value-at-risk based on either independent errors or GARCH errors. Our approach not only bypasses the challenge of directly estimating the asymptotic variance but also does not need to know whether there exists an infinite variance predictor in the predictive model. Furthermore, we apply the same idea to the quantile regression method, which allows infinite variance predictors and generalizes the parameter estimation in (*Econometric Theory* **22** (2006) 173–205) to conditional value-at-risk in the Supplementary Material. We demonstrate the finite sample performance of the derived confidence intervals through numerical studies before applying them to real data.

REFERENCES

ADRIAN, T., BOYARCHENKO, N. and GIANNONE, D. (2019). Vulnerable growth. *Am. Econ. Rev.* **109** 1263–1289.

ADRIAN, T. and BRUNNERMEIER, M. (2016). CoVaR. *Am. Econ. Rev.* **106** 1705–1741.

BARTOLUCCI, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statist. Probab. Lett.* **77** 104–110. MR2339024 https://doi.org/10.1016/j.spl.2006.05.016

BAYSAL, R. E. and STAUM, J. (2008). Empirical likelihood for value-at-risk and expected shortfall. *J. Risk* **11** 3–32.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

CHEN, S. X. and HALL, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21** 1166–1181. MR1241263 https://doi.org/10.1214/aos/1176349256

CHEN, J., VARIYATH, A. M. and ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17** 426–443. MR2439967 https://doi.org/10.1198/106186008X321068

CHERNOZHUKOV, V. (2005). Extremal quantile regression. *Ann. Statist.* **33** 806–839. MR2163160 https://doi.org/10.1214/009053604000001165

CHRISTOFFERSEN, P. and GONÇALVES, S. (2005). Estimation risk in financial risk management. *J. Risk* **7** 1–28.

CHUN, S. Y., SHAPIRO, A. and URYASEV, S. (2012). Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Oper. Res.* **60** 739–756. MR2979419 https://doi.org/10.1287/opre.1120.1072

COCHRANE, J. H. (2009). *Asset Pricing*, rev. ed. Princeton Univ. Press, Princeton, NJ.

MSC2020 subject classifications. 62G20, 62M10, 62P20.

Key words and phrases. Conditional risk measures, empirical likelihood, least squares estimation, interval estimation, quantile regression, risk management, value-at-risk.

DAVIS, R. A. and HSING, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23** 879–917. MR1334176

DAVIS, R. A. and WU, W. (1997). M-estimation for linear regression with infinite variance. *Probab. Math. Statist.* **17** 1–20. MR1455606

DUFFIE, D. and PAN, J. (1997). An overview of value at risk. *J. Deriv.* **4** 7–49.

DUFFIE, D. and PAN, J. (2001). Analytical value-at-risk with jumps and credit risk. *Finance Stoch.* **5** 155–180. MR1841715 https://doi.org/10.1007/PL00013531

EMERSON, S. C. and OWEN, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Stat.* **3** 1161–1192. MR2566185 https://doi.org/10.1214/09-EJS518

ENGLE, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *J. Econ. Perspect.* **15** 157–168.

ENGLE, R. (2004). Risk and volatility: Econometric models and financial practice. *Am. Econ. Rev.* **94** 405–420.

ENGLE, R. F. and MANGANELLI, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22** 367–381. MR2091566 https://doi.org/10.1198/073500104000000370

FAMA, E. F. and ROLL, R. (1971). Parameter estimates for symmetric stable distributions. *J. Amer. Statist. Assoc.* **66** 331–338.

FAN, Y., HÄRDLE, W. K., WANG, W. and ZHU, L. (2018). Single-index-based CoVaR with very high-dimensional covariates. *J. Bus. Econom. Statist.* **36** 212–226. MR3790209 https://doi.org/10.1080/07350015.2016.1180990

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II. Second Edition*. Wiley, New York. MR0270403

GAO, F. and SONG, F. (2008). Estimation risk in Garch VaR and ES estimates. *Econometric Theory* **24** 1404–1424. MR2526553 https://doi.org/10.1017/S0266466608080559

GRANGER, C. W. J. and ORR, D. (1972). Infinite variance and research strategy in time series analysis. *J. Amer. Statist. Assoc.* **67** 275–285.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application: Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

HALL, P. and YAO, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71** 285–317. MR1956860 https://doi.org/10.1111/1468-0262.00396

HE, Y., HOU, Y., PENG, L. and SHEN, H. (2020). Supplement to "Inference for conditional value-at-risk of a predictive regression." https://doi.org/10.1214/19-AOS1937SUPP.

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

HOROWITZ, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66** 1327–1351. MR1654307 https://doi.org/10.2307/2999619

JORION, P. (1996). Risk^2: Measuring the risk in value at risk. *Financ. Anal. J.* **52** 47–56.

JORION, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw-Hill, Boston, MA.

KITAMURA, Y. (2007). Empirical likelihood methods in econometrics: Theory and practice. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*. Cambridge Univ. Press, Cambridge.

KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657 https://doi.org/10.1017/CBO9780511754098

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

KOU, S. and PENG, X. (2016). On the measurement of economic tail risk. *Oper. Res.* **64** 1056–1072. MR3558435 https://doi.org/10.1287/opre.2016.1539

KOU, S., PENG, X. and HEYDE, C. C. (2013). External risk measures and Basel accords. *Math. Oper. Res.* **38** 393–417. MR3092538 https://doi.org/10.1287/moor.1120.0577

KOUL, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture Notes in Statistics* **166**. Springer, New York. Second edition of *Weighted empiricals and linear models* [Inst. Math. Statist., Hayward, CA, 1992; MR1218395 (95c:62061)]. MR1911855 https://doi.org/10.1007/978-1-4613-0055-7

LAHIRI, S. N. and MUKHOPADHYAY, S. (2012). A penalized empirical likelihood method in high dimensions. *Ann. Statist.* **40** 2511–2540. MR3097611 https://doi.org/10.1214/12-AOS1040

LAI, T. L. and XING, H. (2013). Stochastic change-point ARX-GARCH models and their applications to econometric time series. *Statist. Sinica* **23** 1573–1594. MR3222811

LEE, S. and NOH, J. (2013). Quantile regression estimator for GARCH models. *Scand. J. Stat.* **40** 2–20. MR3024029 https://doi.org/10.1111/j.1467-9469.2011.00759.x

LING, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J. Econometrics* **140** 849–873. MR2408929 https://doi.org/10.1016/j.jeconom.2006.07.016

MAMMEN, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24** 307–335. MR1389892 https://doi.org/10.1214/aos/1033066211

MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Bus.* **36** 394–419.

MCNEIL, A. J., FREY, R. and EMBRECHTS, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised ed. *Princeton Series in Finance*. Princeton Univ. Press, Princeton, NJ. MR3445371

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *J. Statist. Plann. Inference* **142** 552–566. MR2843057 https://doi.org/10.1016/j.jspi.2011.08.013

OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. MR1041387 https://doi.org/10.1214/aos/1176347494

OWEN, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, New York.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085 https://doi.org/10.1214/aos/1176325370

RACHEV, S. T., MENN, C. and FABOZZI, F. J. (2005). *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Wiley, Hoboken, NJ.

RACHEV, S. T. and MITTNIK, S. (2000). *Stable Paretian Models in Finance*. Wiley, New York.

RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2271424

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705 https://doi.org/10.2307/2951780

SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics*. *Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. Reprint of the 1986 original [MR0838963]. MR3396731 https://doi.org/10.1137/1.9780898719017.ch1

SPOKOINY, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Ann. Statist.* **37** 1405–1436. MR2509078 https://doi.org/10.1214/08-AOS612

TSAO, M. and WU, F. (2013). Empirical likelihood on the full parameter space. *Ann. Statist.* **41** 2176–2196. MR3127862 https://doi.org/10.1214/13-AOS1143

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

WHANG, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* **22** 173–205. MR2230386 https://doi.org/10.1017/S0266466606060087

XU, K.-L. (2016). Model-free inference for tail risk measures. *Econometric Theory* **32** 122–153. MR3442504 https://doi.org/10.1017/S0266466614000802

The Annals of Statistics
2020, Vol. 48, No. 6, 3465–3487
https://doi.org/10.1214/19-AOS1938

SIMULTANEOUS HIGH-PROBABILITY BOUNDS ON THE FALSE DISCOVERY PROPORTION IN STRUCTURED, REGRESSION AND ONLINE SETTINGS

BY EUGENE KATSEVICH[1] AND AADITYA RAMDAS[2]

[1]*Department of Statistics and Data Science, Carnegie Mellon University, ekatsevi@andrew.cmu.edu*
[2]*Department of Statistics and Data Science, Carnegie Mellon University, aramdas@cmu.edu*

While traditional multiple testing procedures prohibit adaptive analysis choices made by users, Goeman and Solari (*Statist. Sci.* **26** (2011) 584–597) proposed a simultaneous inference framework that allows users such flexibility while preserving high-probability bounds on the false discovery proportion (FDP) of the chosen set. In this paper, we propose a new class of such simultaneous FDP bounds, tailored for nested sequences of rejection sets. While most existing simultaneous FDP bounds are based on closed testing using global null tests based on sorted p-values, we additionally consider the setting where side information can be leveraged to boost power, the variable selection setting where knockoff statistics can be used to order variables, and the online setting where decisions about rejections must be made as data arrives. Our finite-sample, closed form bounds are based on repurposing the FDP estimates from false discovery rate (FDR) controlling procedures designed for each of the above settings. These results establish a novel connection between the parallel literatures of simultaneous FDP bounds and FDR control methods, and use proof techniques employing martingales and filtrations that are new to both these literatures. We demonstrate the utility of our results by augmenting a recent knockoffs analysis of the UK Biobank dataset.

REFERENCES

AHARONI, E. and ROSSET, S. (2014). Generalized α-investing: Definitions, optimality results and application to public databases. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 771–794. MR3248676 https://doi.org/10.1111/rssb.12048

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S. et al. (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25** 25.

BACHOC, F., PREINERSTORFER, D. and STEINBERGER, L. (2016). Uniformly valid confidence intervals post-model-selection. Available at arXiv:1611.01043.

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

BARBER, R. F. and RAMDAS, A. (2017). The p-filter: Multilayer false discovery rate control for grouped hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1247–1268. MR3689317 https://doi.org/10.1111/rssb.12218

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BENJAMINI, Y. and LIU, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Inference* **82** 163–170. MR1736441 https://doi.org/10.1016/S0378-3758(99)00040-3

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

MSC2020 subject classifications. Primary 62J15, 62F03; secondary 60G42, 62P10.
Key words and phrases. False discovery rate, FDR, multiple testing, post hoc confidence bounds, false discovery exceedance, FDX, uniform martingale concentration.

BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc confidence bounds on false positives using reference families. *Ann. Statist.* **48** 1281–1303. MR4124323 https://doi.org/10.1214/19-AOS1847

BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562** 203.

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

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642–669. MR0083864 https://doi.org/10.1214/aoms/1177728174

DWORK, C., SU, W. J. and ZHANG, L. (2018). Differentially private false discovery rate control. Available at arXiv:1807.04209v1.

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

FOSTER, D. P. and STINE, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 429–444. MR2424761 https://doi.org/10.1111/j.1467-9868.2007.00643.x

G'SELL, M. G., WAGER, S., CHOULDECHOVA, A. and TIBSHIRANI, R. (2016). Sequential selection procedures and false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 423–444. MR3454203 https://doi.org/10.1111/rssb.12122

GAVRILOV, Y., BENJAMINI, Y. and SARKAR, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.* **37** 619–629. MR2502645 https://doi.org/10.1214/07-AOS586

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101** 1408–1417. MR2279468 https://doi.org/10.1198/016214506000000339

GOEMAN, J., HEMERIK, J. and SOLARI, A. (2019). Only closed testing procedures are admissible for controlling false discovery proportions. Available at arXiv:1901.04885.

GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statist. Sci.* **26** 584–597. MR2951390 https://doi.org/10.1214/11-STS356

GOEMAN, J. J., MEIJER, R. J., KREBS, T. J. P. and SOLARI, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika* **106** 841–856. MR4046036 https://doi.org/10.1093/biomet/asz041

HEMERIK, J., SOLARI, A. and GOEMAN, J. J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika* **106** 635–649. MR3992394 https://doi.org/10.1093/biomet/asz021

JAVANMARD, A. and MONTANARI, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *Ann. Statist.* **46** 526–554. MR3782376 https://doi.org/10.1214/17-AOS1559

KATSEVICH, E. and RAMDAS, A. (2020). Supplement to "Simultaneous high-probability bounds on the false discovery proportion in structured, regression and online settings." https://doi.org/10.1214/19-AOS1938SUPP.

KATSEVICH, E. and SABATTI, C. (2019). Multilayer knockoff filter: Controlled variable selection at multiple resolutions. *Ann. Appl. Stat.* **13** 1–33. MR3937419 https://doi.org/10.1214/18-AOAS1185

KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., CAI, J., GEORGE, E. I. and ZHAO, L. (2020). Valid post-selection inference in model-free linear regression. *Ann. Statist.* To appear.

LEHMANN, E. L. and ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631 https://doi.org/10.1214/009053605000000084

LEI, L. and FITHIAN, W. (2016). Power of ordered hypothesis testing. In *International Conference on Machine Learning* 2924–2932.

LEI, L. and FITHIAN, W. (2018). AdaPT: An interactive procedure for multiple testing with side information. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 649–679. MR3849338 https://doi.org/10.1111/rssb.12253

LEI, L., RAMDAS, A. and FITHIAN, W. (2017). STAR: A general interactive framework for FDR control under structural constraints. Available at arXiv:1710.02776.

LI, A. and BARBER, R. F. (2017). Accumulation tests for FDR control in ordered hypothesis testing. *J. Amer. Statist. Assoc.* **112** 837–849. MR3671774 https://doi.org/10.1080/01621459.2016.1180989

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. and BEJERANO, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28** 495–501. https://doi.org/10.1038/nbt.1630

MEINSHAUSEN, N. (2006). False discovery control for multiple tests of association under general dependence. *Scand. J. Stat.* **33** 227–237. MR2279639 https://doi.org/10.1111/j.1467-9469.2005.00488.x

MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246 https://doi.org/10.1214/009053605000000741

OWEN, A. B. (2005). Variance of the number of false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 411–426. MR2155346 https://doi.org/10.1111/j.1467-9868.2005.00509.x

RAMDAS, A., YANG, F., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). Online control of the false discovery rate with decaying memory. In *Advances in Neural Information Processing Systems*.

RAMDAS, A., ZRNIC, T., WAINWRIGHT, M. and JORDAN, M. (2018). SAFFRON: An adaptive algorithm for online control of the false discovery rate. In *Proceedings of the* 35*th International Conference on Machine Learning* 4286–4294.

RAMDAS, A. K., BARBER, R. F., WAINWRIGHT, M. J. and JORDAN, M. I. (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *Ann. Statist.* **47** 2790–2821. MR3988773 https://doi.org/10.1214/18-AOS1765

ROBBINS, H. (1954). A one-sided confidence interval for an unknown distribution function. *Ann. Math. Stat.* **25** 409.

SESIA, M., SABATTI, C. and CANDÈS, E. J. (2019). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18. MR3912377 https://doi.org/10.1093/biomet/asy033

SESIA, M., KATSEVICH, E., BATES, S., CANDÈS, E. and SABATTI, C. (2019). Multi-resolution localization of causal variants across the genome. *BioRxiv*. https://doi.org/10.1101/631390

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766 https://doi.org/10.1111/j.1467-9868.2004.00439.x

VAN DER LAAN, M. J., DUDOIT, S. and POLLARD, K. S. (2004). Multiple testing. Part III. Procedures for control of the generalized family-wise error rate and proportion of false positives.

VILLE, J. (1939). *Etude critique de la notion de collectif*. Gauthier-Villars, Paris. MR3533075

WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88. MR0651392 https://doi.org/10.1007/BF00635964

The Annals of Statistics
2020, Vol. 48, No. 6, 3488–3512
https://doi.org/10.1214/19-AOS1939

CLUSTERING IN BLOCK MARKOV CHAINS

BY JARON SANDERS[1,2,*], ALEXANDRE PROUTIÈRE[1,†] AND SE-YOUNG YUN[3]

[1]*Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology,* [*]*jaron.sanders@tue.nl;* [†]*alepro@kth.se*

[2]*Department of Quantum & Computer Engineering, Faculty of Electrical Engineering, Mathematics & Computer Science, Delft University of Technology, j.sanders@tudelft.nl*

[3]*Department of Industrial & Systems Engineering, Graduate School of AI, Korea Advanced Institute of Science and Technology, yunseyoung@kaist.ac.kr*

This paper considers cluster detection in Block Markov Chains (BMCs). These Markov chains are characterized by a block structure in their transition matrix. More precisely, the n possible states are divided into a finite number of K groups or clusters, such that states in the same cluster exhibit the same transition rates to other states. One observes a trajectory of the Markov chain, and the objective is to recover, from this observation only, the (initially unknown) clusters. In this paper, we devise a clustering procedure that accurately, efficiently and provably detects the clusters. We first derive a fundamental information-theoretical lower bound on the detection error rate satisfied under any clustering algorithm. This bound identifies the parameters of the BMC, and trajectory lengths, for which it is possible to accurately detect the clusters. We next develop two clustering algorithms that can together accurately recover the cluster structure from the shortest possible trajectories, whenever the parameters allow detection. These algorithms thus reach the fundamental detectability limit, and are optimal in that sense.

REFERENCES

[1] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

[2] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334 https://doi.org/10.1109/FOCS.2015.47

[3] ABBE, E. and SANDON, C. (2015). Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems* 28 676–684.

[4] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2010). Spectrum of large random reversible Markov chains: Two examples. *ALEA Lat. Am. J. Probab. Math. Stat.* **7** 41–64. MR2644041

[5] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2011). Spectrum of large random reversible Markov chains: Heavy-tailed weights on the complete graph. *Ann. Probab.* **39** 1544–1590. MR2857250 https://doi.org/10.1214/10-AOP587

[6] BRÉMAUD, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. *Texts in Applied Mathematics* **31**. Springer, New York. MR1689633 https://doi.org/10.1007/978-1-4757-3124-8

[7] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.* **107** Art. ID 065701. https://doi.org/10.1103/PhysRevLett.107.065701

[8] FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709 https://doi.org/10.1002/rsa.20089

[9] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Art. ID 60. MR3687603

[10] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inf. Theory* **62** 2788–2797. MR3493879 https://doi.org/10.1109/TIT.2016.2546280

MSC2020 subject classifications. 62H30, 60J10, 60J20.

Key words and phrases. Clustering, Markov chains, mixing times, community detection, change of measure, asymptotic analysis, information theory.

[11] HALKO, N., MARTINSSON, P. G. and TROPP, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53** 217–288. MR2806637 https://doi.org/10.1137/090771806

[12] HOCHSTÄTTLER, W., KIRSCH, W. and WARZEL, S. (2016). Semicircle law for a matrix ensemble with dependent entries. *J. Theoret. Probab.* **29** 1047–1068. MR3540489 https://doi.org/10.1007/s10959-015-0602-3

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

[14] JOG, V. and LOH, P.-L. (2015). Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence. Preprint. Available at arXiv:1509.06418.

[15] KIRSCH, W. and KRIECHERBAUER, T. (2018). Sixty years of moments for random matrices. In *Non-linear Partial Differential Equations, Mathematical Physics, and Stochastic Analysis. EMS Ser. Congr. Rep.* 349–379. Eur. Math. Soc., Zürich. MR3823851

[16] KIRSCH, W. and KRIECHERBAUER, T. (2018). Semicircle law for generalized Curie–Weiss matrix ensembles at subcritical temperature. *J. Theoret. Probab.* **31** 2446–2458. MR3866621 https://doi.org/10.1007/s10959-017-0768-y

[17] LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826 https://doi.org/10.1016/0196-8858(85)90002-8

[18] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[19] LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

[20] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

[21] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6

[22] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* 69–75. ACM, New York.

[23] ORTNER, R. (2013). Adaptive aggregation for reinforcement learning in average reward Markov decision processes. *Ann. Oper. Res.* **208** 321–336. MR3100636 https://doi.org/10.1007/s10479-012-1064-y

[24] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** Art. ID 79. MR3383563 https://doi.org/10.1214/EJP.v20-4039

[25] SANDERS, J., PROUTIÈRE, A. and YUN, S.-Y. (2020). Supplement to "Clustering in Block Markov Chains." https://doi.org/10.1214/19-AOS1939SUPP

[26] SINGH, S. P., JAAKKOLA, T. and JORDAN, M. I. (1995). Reinforcement learning with soft state aggregation. In *Advances in Neural Information Processing Systems* 7 (G. Tesauro, D. S. Touretzky and T. K. Leen, eds.) 361–368. MIT Press, Cambridge, MA.

[27] SUTTON, R. S. and BARTO, A. G. (1998). *Introduction to Reinforcement Learning*, 1st ed. MIT Press, Cambridge, MA. MR3889951

[28] TAO, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. MR2906465 https://doi.org/10.1090/gsm/132

[29] TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

[30] WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2) **67** 325–327. MR0095527 https://doi.org/10.2307/1970008

[31] YUN, S.-Y. and PROUTIERE, A. (2014). Community detection via random and adaptive sampling. In *COLT* 138–175.

[32] YUN, S.-Y. and PROUTIERE, A. (2014). Accurate community detection in the stochastic block model via spectral algorithms. Preprint. Available at arXiv:1412.7335.

[33] YUN, S.-Y. and PROUTIERE, A. (2016). Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems* 29 965–973.

The Annals of Statistics
2020, Vol. 48, No. 6, 3513–3544
https://doi.org/10.1214/19-AOS1940

MODEL SELECTION AND LOCAL GEOMETRY

BY ROBIN J. EVANS

Department of Statistics, University of Oxford, evans@stats.ox.ac.uk

We consider problems in model selection caused by the geometry of models close to their points of intersection. In some cases—including common classes of causal or graphical models, as well as time series models—distinct models may nevertheless have identical tangent spaces. This has two immediate consequences: first, in order to obtain constant power to reject one model in favour of another we need local alternative hypotheses that decrease to the null at a slower rate than the usual parametric $n^{-1/2}$ (typically we will require $n^{-1/4}$ or slower); in other words, to distinguish between the models we need large effect sizes or very large sample sizes. Second, we show that under even weaker conditions on their tangent cones, models in these classes cannot be made simultaneously convex by a reparameterization.

This shows that Bayesian network models, amongst others, cannot be learned directly with a convex method similar to the graphical lasso. However, we are able to use our results to suggest methods for model selection that learn the tangent space directly, rather than the model itself. In particular, we give a generic algorithm for learning Bayesian network models.

REFERENCES

ARAGAM, B. and ZHOU, Q. (2015). Concave penalized estimation of sparse Gaussian Bayesian networks. *J. Mach. Learn. Res.* **16** 2273–2328. MR3450507

BEINLICH, I. A., SUERMONDT, H. J., CHAVEZ, R. M. and COOPER, G. F. (1989). The ALARM monitoring system. In *Proceedings of AIME* 89: *Second European Conference on Artificial Intelligence in Medicine* (J. Hunter, J. Cookson and J. Wyatt, eds.) 247–256. Springer, Berlin Heidelberg.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BERGSMA, W. P. and RUDAS, T. (2002). Marginal models for categorical data. *Ann. Statist.* **30** 140–159. MR1892659 https://doi.org/10.1214/aos/1015362188

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

BOCHNAK, J., COSTE, M. and ROY, M.-F. (2013). *Real algebraic geometry*. Springer, Berlin.

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

BÜHLMANN, P., PETERS, J. and ERNEST, J. (2014). CAM: Causal additive models, high-dimensional order search and penalized regression. *Ann. Statist.* **42** 2526–2556. MR3277670 https://doi.org/10.1214/14-AOS1260

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

CHICKERING, D. M. (1996). Learning Bayesian networks is NP-complete. In *Learning from Data* (*Fort Lauderdale*, *FL*, 1995). *Lect. Notes Stat.* **112** 121–130. Springer, New York. MR1473013 https://doi.org/10.1007/978-1-4612-2404-4_12

CONLON, L. (2008). *Differentiable Manifolds*, 2nd ed. *Modern Birkhäuser Classics*. Birkhäuser, Boston, MA. MR2413709

COX, D. R. (1990). Role of models in statistical analysis. *Statist. Sci.* **5** 169–174. MR1062575

MSC2020 subject classifications. 62F03, 62H15.

Key words and phrases. Asymptotic theory, causal inference, convex optimization, geometry, hypothesis test, model selection.

COX, D., LITTLE, J. and O'SHEA, D. (2007). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd ed. *Undergraduate Texts in Mathematics*. Springer, New York. MR2290010 https://doi.org/10.1007/978-0-387-35651-8

DRTON, M. (2009a). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. MR2502658 https://doi.org/10.1214/07-AOS571

DRTON, M. (2009b). Discrete chain graph models. *Bernoulli* **15** 736–753. MR2555197 https://doi.org/10.3150/08-BEJ172

DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Birkhäuser, Basel. MR2723140 https://doi.org/10.1007/978-3-7643-8905-5

DRTON, M. and SULLIVANT, S. (2007). Algebraic statistical models. *Statist. Sinica* **17** 1273–1297. MR2398596

EVANS, R. J. (2015). Smoothness of marginal log-linear parameterizations. *Electron. J. Stat.* **9** 475–491. MR3326132 https://doi.org/10.1214/15-EJS1009

FERRAROTTI, M., FORTUNA, E. and WILSON, L. (2002). Local approximation of semialgebraic sets. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **1** 1–11. MR1994799

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

FU, F. and ZHOU, Q. (2013). Learning sparse causal Gaussian networks with experimental intervention: Regularization and coordinate descent. *J. Amer. Statist. Assoc.* **108** 288–300. MR3174620 https://doi.org/10.1080/01621459.2012.754359

GU, J., FU, F. and ZHOU, Q. (2019). Penalized estimation of directed acyclic graphs from discrete data. *Stat. Comput.* **29** 161–176. MR3905546 https://doi.org/10.1007/s11222-018-9801-y

KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227

LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. MR0981437 https://doi.org/10.1214/aos/1176347003

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

NARDI, Y. and RINALDO, A. (2012). The log-linear group-lasso estimator and its asymptotic properties. *Bernoulli* **18** 945–974. MR2948908 https://doi.org/10.3150/11-BEJ364

NI, Y., STINGO, F. C. and BALADANDAYUTHAPANI, V. (2015). Bayesian nonlinear model selection for gene regulatory networks. *Biometrics* **71** 585–595. MR3402594 https://doi.org/10.1111/biom.12309

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

RICHARDSON, T. EVANS, J. ROBINS, M. and SHPITSER, I. (2017). Nested Markov properties for acyclic directed mixed graphs. Available at arXiv:1701.06686.

RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015

ROCHA, G. V., WANG, X. and YU, B. (2009). Asymptotic distribution and sparsistency for ℓ_1-penalized parametric M-estimators with applications to linear SVM and logistic regression. Preprint. Available at arXiv:0908.1940.

RUDAS, T., BERGSMA, W. P. and NÉMETH, R. (2010). Marginal log-linear parameterization of conditional independence models. *Biometrika* **97** 1006–1012. MR2746171 https://doi.org/10.1093/biomet/asq037

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

SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. MR2274431

SHOJAIE, A. and MICHAILIDIS, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* **97** 519–538. MR2672481 https://doi.org/10.1093/biomet/asq038

SHPITSER, I., EVANS, R., RICHARDSON, T. and ROBINS, J. (2013). Sparse nested Markov models with log-linear parameters. In *UAI*-2013 576–585.

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. *Interdisciplinary Statistics*. CRC Press/CRC, Boca Raton, FL. MR2059021 https://doi.org/10.1201/9780203489437

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

TCHETGEN TCHETGEN, E. J. and SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *Ann. Statist.* **40** 1816–1845. MR3015045 https://doi.org/10.1214/12-AOS990

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

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

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

WITTEN, D. M., FRIEDMAN, J. H. and SIMON, N. (2011). New insights and faster computations for the graphical lasso. *J. Comput. Graph. Statist.* **20** 892–900. MR2878953 https://doi.org/10.1198/jcgs.2011.11051a

ZHANG, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence* **172** 1873–1896. MR2459793 https://doi.org/10.1016/j.artint.2008.08.001

ZWIERNIK, P., UHLER, C. and RICHARDS, D. (2017). Maximum likelihood estimation for linear Gaussian covariance models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1269–1292. MR3689318 https://doi.org/10.1111/rssb.12217

The Annals of Statistics
2020, Vol. 48, No. 6, 3545–3564
https://doi.org/10.1214/19-AOS1941

IRREDUCIBILITY AND GEOMETRIC ERGODICITY OF HAMILTONIAN MONTE CARLO

BY ALAIN DURMUS[1], ÉRIC MOULINES[2] AND EERO SAKSMAN[3]

[1]*CNRS, ENS, Université Paris-Saclay, alain.durmus@ens-paris-saclay.fr*
[2]*Faculty of Computer Sciences, HSE University, eric.moulines@hse.ru*
[3]*Department of Mathematics and Statistics, University of Helsinki, eero.saksman@helsinki.fi*

Hamiltonian Monte Carlo (HMC) is currently one of the most popular Markov Chain Monte Carlo algorithms to sample smooth distributions over continuous state space. This paper discusses the irreducibility and geometric ergodicity of the HMC algorithm. We consider cases where the number of steps of the Störmer–Verlet integrator is either fixed or random. Under mild conditions on the potential U associated with target distribution π, we first show that the Markov kernel associated to the HMC algorithm is irreducible and positive recurrent. Under more stringent conditions, we then establish that the Markov kernel is Harris recurrent. We provide verifiable conditions on U under which the HMC sampler is geometrically ergodic. Finally, we illustrate our results on several examples.

REFERENCES

[1] BETANCOURT, M., BYRNE, S., LIVINGSTONE, S. and GIROLAMI, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli* **23** 2257–2298. MR3648031 https://doi.org/10.3150/16-BEJ810

[2] LIVINGSTONE, S., BETANCOURT, M., BYRNE, S. and GIROLAMI, M. (2016). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernouilli*. Preprint. Available at arXiv:1601.08057v2.

[3] BOU-RABEE, N., EBERLE, A. and ZIMMER, R. (2018). Coupling and convergence for Hamiltonian Monte Carlo. Preprint. Available at arXiv:1805.00452.

[4] BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523 https://doi.org/10.1214/16-AAP1255

[5] BOU-RABEE, N. and SANZ-SERNA, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.* **27** 113–206. MR3826507 https://doi.org/10.1017/s0962492917000101

[6] BYRNE, S. and GIROLAMI, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scand. J. Stat.* **40** 825–845. MR3145120 https://doi.org/10.1111/sjos.12036

[7] CANCÈS, E., LEGOLL, F. and STOLTZ, G. (2007). Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM Math. Model. Numer. Anal.* **41** 351–389. MR2339633 https://doi.org/10.1051/m2an:2007014

[8] DOUC, R., MOULINES, E., PRIOURET, P. and SOULIER, P. (2018). *Markov Chains. Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

[9] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

[10] DUISTERMAAT, J. J. and KOLK, J. A. C. (2004). *Multidimensional Real Analysis. I. Differentiation. Cambridge Studies in Advanced Mathematics* **86**. Cambridge Univ. Press, Cambridge. Translated from the Dutch by J. P. van Braam Houckgeest. MR2121976 https://doi.org/10.1017/CBO9780511616716

[11] DURMUS, A., MOULINES, É. and SAKSMAN, E. (2020). Supplement to "Irreducibility and geometric ergodicity of Hamiltonian Monte Carlo." https://doi.org/10.1214/19-AOS1941SUPP

[12] FANG, Y., SANZ-SERNA, J. M. and SKEEL, R. D. (2016). Compressible generalized hybrid Monte Carlo (vol 140, 174108, 2014). *J. Chem. Phys.* **144**.

MSC2020 subject classifications. 60J05, 60J22, 65C05.
Key words and phrases. Hamiltonian Monte Carlo, Markov Chain Monte Carlo, irreducibility, geometric ergodicity.

[13] GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

[14] LUBICH, C., HAIRER, E. and WANNER, G. (2002). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed. *Springer Series in Computational Mathematics* **31**. Springer, Heidelberg.

[15] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. Springer, New York. MR2401592

[16] LIVINGSTONE, S., FAULKNER, M. F. and ROBERTS, G. O. (2019). Kinetic energy choice in Hamiltonian/hybrid Monte Carlo. *Biometrika* **106** 303–319. MR3949305 https://doi.org/10.1093/biomet/asz013

[17] LU, X., PERRONE, V., HASENCLEVER, L., TEH, Y. W. and VOLLMER, S. (2017). Relativistic Monte Carlo. In *Proceedings of Machine Learning Research* **54**. Artificial Intelligence and Statistics, AISTAT.

[18] MANGOUBI, O. and SMITH, A. (2017). Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. Preprint. Available at arXiv:1708.07114.

[19] MANGOUBI, O. and VISHNOI, N. (2018). Dimensionally tight bounds for second-order Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems* 6027–6037.

[20] MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201

[21] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 https://doi.org/10.1017/CBO9780511626630

[22] NEAL, R. M. (1993). Bayesian learning via stochastic dynamics. *Adv. Neural Inf. Process. Syst.* 475–475.

[23] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

[24] OUTERELO, E. and RUIZ, J. M. (2009). *Mapping Degree Theory*. *Graduate Studies in Mathematics* **108**. Amer. Math. Soc., Providence, RI. MR2566906 https://doi.org/10.1090/gsm/108

[25] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

[26] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158 https://doi.org/10.1093/biomet/83.1.95

[27] RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. MR0924157

[28] SCHOFIELD, M. R., BARKER, R. J., GELMAN, A., COOK, E. R. and BRIFFA, K. R. (2016). A model-based approach to climate reconstruction using tree-ring data. *J. Amer. Statist. Assoc.* **111** 93–106. MR3494640 https://doi.org/10.1080/01621459.2015.1110524

[29] TANG, Y., SRIVASTAVA, N. and SALAKHUTDINOV, R. R. (2014). Learning generative models with visual attention. In *Advances in Neural Information Processing Systems* 1808–1816.

The Annals of Statistics
2020, Vol. 48, No. 6, 3565–3588
https://doi.org/10.1214/20-AOS1943

TEST FOR HIGH DIMENSIONAL COVARIANCE MATRICES

BY YUEFENG HAN[1] AND WEI BIAO WU[2]

[1]*Department of Statistics, Rutgers University, yuefeng.han@rutgers.edu*
[2]*Department of Statistics, University of Chicago, wbwu@galton.uchicago.edu*

The paper introduces a new test for testing structures of covariances for high dimensional vectors and the data dimension can be much larger than the sample size. Under proper normalization, central and noncentral limit theorems are established. The asymptotic theory is attained without imposing any explicit restriction between data dimension and sample size. To facilitate the related statistical inference, we propose the balanced Rademacher weighted differencing scheme, which is also the delete-half jackknife, to approximate the distribution of the proposed test statistics. We also develop a new testing procedure for substructures of precision matrices. The simulation results show that the tests outperform the exiting methods both in terms of size and power. Our test procedure is applied to a colorectal cancer dataset.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

ARRAS, B., MIJOULE, G., POLY, G. and SWAN, Y. (2016). A new approach to the Stein–Tikhomirov method: With applications to the second Wiener chaos and Dickman convergence. arXiv preprint arXiv:1605.06819.

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

BENDAT, J. S. (1990). *Nonlinear System Analysis and Identification from Random Data. A Wiley-Interscience Publication*. Wiley, New York. MR1045198

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

CAI, T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

CASTI, J. L. (1985). *Nonlinear System Theory*. *Mathematics in Science and Engineering* **175**. Academic Press, Orlando, FL. MR0777143

CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859 https://doi.org/10.1214/07-AOP370

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

CHERIF, I. and FNAIECH, F. (2015). Nonlinear system identification with a real-coded genetic algorithm (RCGA). *Int. J. Appl. Math. Comput. Sci.* **25** 863–875. MR3468501 https://doi.org/10.1515/amcs-2015-0062

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

MSC2020 subject classifications. Primary 62F05; secondary 62E17.

Key words and phrases. Gaussian approximation, high dimensional covariance matrices, covariance function, resampling.

DE LA FUENTE, A. (2010). From ‘differential expression’ to ‘differential networking’—identification of dysfunctional regulatory networks in diseases. *Trends Genet*. **26** 326–333.

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991 https://doi.org/10.1016/j.jeconom.2008.09.017

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. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. MR3091653 https://doi.org/10.1111/rssb.12016

FURRER, R. and BENGTSSON, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal*. **98** 227–255. MR2301751 https://doi.org/10.1016/j.jmva.2006.08.003

GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc*. **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

HAN, Y. and WU, W. B. (2020). Supplement to “Test for high dimensional covariance matrices.” https://doi.org/10.1214/20-AOS1943SUPP.

HE, J. and CHEN, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *J. Econometrics* **194** 283–297. MR3536981 https://doi.org/10.1016/j.jeconom.2016.05.007

HE, J. and CHEN, S. X. (2018). High-dimensional two-sample covariance matrix testing via super-diagonals. *Statist. Sinica* **28** 2671–2696. MR3839879

HEDAYAT, A. and WALLIS, W. D. (1978). Hadamard matrices and their applications. *Ann. Statist*. **6** 1184–1238. MR0523759

HIMENO, T. and YAMADA, T. (2014). Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Anal*. **130** 27–44. MR3229522 https://doi.org/10.1016/j.jmva.2014.04.020

HUDSON, N. J., REVERTER, A. and DALRYMPLE, B. P. (2009). A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput. Biol*. **5** e1000382. https://doi.org/10.1371/journal.pcbi.1000382

IDEKER, T. and KROGAN, N. J. (2012). Differential network biology. *Mol. Syst. Biol*. **8** 565.

JIA, P., KAO, C.-F., KUO, P.-H. and ZHAO, Z. (2011). A comprehensive network and pathway analysis of candidate genes in major depressive disorder. *BMC Syst. Biol*. **5 Suppl 3** S12. https://doi.org/10.1186/1752-0509-5-S3-S12

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab*. **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568 https://doi.org/10.1093/biomet/58.1.123

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619 https://doi.org/10.1093/biomet/59.1.169

KALOUPTSIDIS, N. and KOUKOULAS, P. (2005). Blind identification of Volterra–Hammerstein systems. *IEEE Trans. Signal Process*. **53** 2777–2787. MR2170655 https://doi.org/10.1109/TSP.2005.850357

KELLEY, R. and IDEKER, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol*. **23** 561–566.

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, Y., AGARWAL, P. and RAJAGOPALAN, D. (2008). A global pathway crosstalk network. *Bioinformatics* **24** 1442–1447.

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist*. **40** 908–940. MR2985938 https://doi.org/10.1214/12-AOS993

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab*. **18** 1269–1283. MR1062069

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3235948 https://doi.org/10.1002/9781118573617

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press [Harcourt Brace Jovanovich, Publishers], London. MR0991969

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist*. **40** 1285–1314. MR3015026 https://doi.org/10.1214/12-AOS1002

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann*. *Statist*. **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

RUGH, W. J. (1981). *Nonlinear System Theory*: *The Volterra–Wiener Approach*. *Johns Hopkins Series in Information Sciences and Systems*. Johns Hopkins Univ. Press, Baltimore, MD. MR0713994

SCHETZEN, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York–Chichester–Brisbane. A Wiley-Interscience Publication. MR0562917

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

WIESEL, A., BIBI, O. and GLOBERSON, A. (2013). Time varying autoregressive moving average models for covariance estimation. *IEEE Trans*. *Signal Process*. **61** 2791–2801. MR3064093 https://doi.org/10.1109/TSP.2013.2256900

WU, C.-F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann*. *Statist*. **18** 1438–1452. MR1062718 https://doi.org/10.1214/aos/1176347759

WU, W. B., LOU, Z. and HAN, Y. (2018). Hypothesis testing for high-dimensional data. In *Handbook of Big Data Analytics*. *Springer Handb. Comput*. *Stat*. 203–224. Springer, Cham. MR3932004

WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist*. *Sinica* **19** 1755–1768. MR2589209

XIA, Y., CAI, T. and CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102** 247–266. MR3371002 https://doi.org/10.1093/biomet/asu074

XIA, Y., CAI, T. and CAI, T. T. (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *J*. *Amer. Statist*. *Assoc*. **113** 328–339. MR3803468 https://doi.org/10.1080/01621459.2016.1251930

XIAO, H. and WU, W. B. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Process*. *Appl*. **123** 2899–2920. MR3054550 https://doi.org/10.1016/j.spa.2013.03.012

XU, M., ZHANG, D. and WU, W. B. (2014). L^2 asymptotics for high-dimensional data. arXiv preprint arXiv:1405.7244.

YARLAGADDA, R. K. and HERSHEY, J. E. (2012). *Hadamard Matrix Analysis and Synthesis*: *With Applications to Communications and Signal/Image Processing*. Springer Science & Business Media. MR1492991 https://doi.org/10.1007/978-1-4615-6313-6

ZHANG, R., PENG, L. and WANG, R. (2013). Tests for covariance matrix with fixed or divergent dimension. *Ann*. *Statist*. **41** 2075–2096. MR3127858 https://doi.org/10.1214/13-AOS1136

ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann*. *Statist*. **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

ZHONG, P.-S., LAN, W., SONG, P. X. K. and TSAI, C.-L. (2017). Tests for covariance structures with high-dimensional repeated measurements. *Ann*. *Statist*. **45** 1185–1213. MR3662452 https://doi.org/10.1214/16-AOS1481

The Annals of Statistics
2020, Vol. 48, No. 6, 3589–3618
https://doi.org/10.1214/20-AOS1944

OPTIMAL ESTIMATION OF VARIANCE IN NONPARAMETRIC REGRESSION WITH RANDOM DESIGN

BY YANDI SHEN[1,*], CHAO GAO[2], DANIELA WITTEN[1,†] AND FANG HAN[1,‡]

[1]*Department of Statistics, University of Washington, [*]ydshen@uw.edu; [†]dwitten@uw.edu; [‡]fanghan@uw.edu*
[2]*Department of Statistics, University of Chicago, chaogao@galton.uchicago.edu*

Consider the heteroscedastic nonparametric regression model with random design

$$Y_i = f(X_i) + V^{1/2}(X_i)\varepsilon_i, \quad i = 1, 2, \dots, n,$$

with $f(\cdot)$ and $V(\cdot)$ α- and β-Hölder smooth, respectively. We show that the minimax rate of estimating $V(\cdot)$ under both local and global squared risks is of the order

$$n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}} \vee n^{-\frac{2\beta}{2\beta+1}},$$

where $a \vee b := \max\{a, b\}$ for any two real numbers a, b. This result extends the fixed design rate $n^{-4\alpha} \vee n^{-2\beta/(2\beta+1)}$ derived in (*Ann. Statist.* **36** (2008) 646–664) in a nontrivial manner, as indicated by the appearances of both α and β in the first term. In the special case of constant variance, we show that the minimax rate is $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ for variance estimation, which further implies the same rate for quadratic functional estimation and thus unifies the minimax rate under the nonparametric regression model with those under the density model and the white noise model. To achieve the minimax rate, we develop a U-statistic-based local polynomial estimator and a lower bound that is constructed over a specified distribution family of randomness designed for both ε_i and X_i.

REFERENCES

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

BHATTACHARYA, A., PATI, D. and DUNSON, D. (2014). Anisotropic function estimation using multi-bandwidth Gaussian processes. *Ann. Statist.* **42** 352–381. MR3189489 https://doi.org/10.1214/13-AOS1192

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

BIRGÉ, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291. MR0816706 https://doi.org/10.1007/BF00332312

BROWN, L. D. and LEVINE, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *Ann. Statist.* **35** 2219–2232. MR2363969 https://doi.org/10.1214/009053607000000145

CAI, T. T., LEVINE, M. and WANG, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *J. Multivariate Anal.* **100** 126–136. MR2460482 https://doi.org/10.1016/j.jmva.2008.03.007

CAI, T. T. and LOW, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34** 2298–2325. MR2291501 https://doi.org/10.1214/009053606000000849

CAI, T. T. and WANG, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *Ann. Statist.* **36** 2025–2054. MR2458178 https://doi.org/10.1214/07-AOS509

DEVROYE, L., MEHRABIAN, A. and REDDAD, T. (2018). The total variation distance between high-dimensional Gaussians. Preprint. Available at arXiv:1810.08693.

DICKER, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269–284. MR3215347 https://doi.org/10.1093/biomet/ast065

MSC2020 subject classifications. 62G08, 62G20.
Key words and phrases. Variance estimation, nonparametric regression, random design, minimax rate, U-statistics.

DOKSUM, K. and SAMAROV, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Ann. Statist.* **23** 1443–1473. MR1370291 https://doi.org/10.1214/aos/1176324307

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9

EFROMOVICH, S. and LOW, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106–1125. MR1401840 https://doi.org/10.1214/aos/1032526959

FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294. MR1126325 https://doi.org/10.1214/aos/1176348249

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. MR1209561

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173 https://doi.org/10.1214/aos/1176349022

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660. MR1665822 https://doi.org/10.1093/biomet/85.3.645

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633. MR0897854 https://doi.org/10.1093/biomet/73.3.625

GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability, II* (*Seattle, WA*, 1999). *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653 https://doi.org/10.3150/07-BEJ110

HALL, P. and CARROLL, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B* **51** 3–14. MR0984989

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842 https://doi.org/10.1093/biomet/77.3.521

HALL, P. and MARRON, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77** 415–419. MR1064818 https://doi.org/10.1093/biomet/77.2.415

HÄRDLE, W. and TSYBAKOV, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81** 223–242. MR1484586 https://doi.org/10.1016/S0304-4076(97)00044-4

HOFFMANN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. MR1902892 https://doi.org/10.1214/aos/1021379858

HUANG, L.-S. and FAN, J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli* **5** 927–949. MR1715445 https://doi.org/10.2307/3318450

IBRAGIMOV, I. A. and KHAS'MINSKIĬ, R. Z. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **108** 72–88. MR0629401

KOLCHIN, V. F., SEVAST'YANOV, B. A. and CHISTYAKOV, V. P. (1978). *Random Allocations*. V. H. Winston & Sons, Washington, DC. MR0471016

KONG, W. and VALIANT, G. (2018). Estimating learnability in the sublinear data regime. Preprint. Available at arXiv:1805.01626.

LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

LEPSKI, O. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LEPSKI, O. (1992). Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

MEYER, Y. (1990). *Ondelettes et Opérateurs. I: Ondelettes. Actualités Mathématiques.* [*Current Mathematical Topics.*] Hermann, Paris. MR1085487

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. *Statistics* **37** 179–188. MR1986175 https://doi.org/10.1080/0233188031000078051

MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625. MR0888429 https://doi.org/10.1214/aos/1176350364

MUNK, A. and RUYMGAART, F. (2002). Minimax rates for estimating the variance and its derivatives in nonparametric regression. *Aust. N. Z. J. Stat.* **44** 479–488. MR1934736 https://doi.org/10.1111/1467-842X.00249

MUNK, A., BISSANTZ, N., WAGNER, T. and FREITAG, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 19–41. MR2136637 https://doi.org/10.1111/j.1467-9868.2005.00486.x

NUSSBAUM, M. (1986). On nonparametric estimation of a regression function, being smooth on a domain in $\mathbb{R}^k$. *Theory Probab. Appl.* **31** 118–125.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230. MR0760684 https://doi.org/10.1214/aos/1176346788

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. Inst. Math. Stat.* (*IMS*) *Collect.* **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

ROBINS, J., TCHETGEN, E. T., LI, L. and VAN DER VAART, A. (2009). Semiparametric minimax rates. *Electron. J. Stat.* **3** 1305–1321. MR2566189 https://doi.org/10.1214/09-EJS479

RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270. MR1379468

RUPPERT, D., WAND, M. P., HOLST, U. and HÖSSJER, O. (1997). Local polynomial variance-function estimation. *Technometrics* **39** 262–273. MR1462587 https://doi.org/10.2307/1271131

SHEN, Y., GAO, C., WITTEN, D. and HAN, F. (2020). Supplement to "Optimal estimation of variance in nonparametric regression with random design." https://doi.org/10.1214/20-AOS1944SUPP

SPOKOINY, V. (2002). Variance estimation for high-dimensional regression models. *J. Multivariate Anal.* **82** 111–133. MR1918617 https://doi.org/10.1006/jmva.2001.2023

THOMPSON, A. M., KAY, J. W. and TITTERINGTON, D. M. (1991). Noise estimation in signal restoration using regularization. *Biometrika* **78** 475–488. MR1130916 https://doi.org/10.1093/biomet/78.3.475

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

VERZELEN, N. and GASSIAT, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli* **24** 3683–3710. MR3788186 https://doi.org/10.3150/17-BEJ975

VERZELEN, N. and VILLERS, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.* **38** 704–752. MR2604699 https://doi.org/10.1214/08-AOS629

VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.* **12** 367–395. MR0006656 https://doi.org/10.1214/aoms/1177731677

VON NEUMANN, J. (1942). A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.* **13** 86–88. MR0006657 https://doi.org/10.1214/aoms/1177731645

WANG, L., BROWN, L. D., CAI, T. T. and LEVINE, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.* **36** 646–664. MR2396810 https://doi.org/10.1214/009053607000000901

The Annals of Statistics
2020, Vol. 48, No. 6, 3619–3642
https://doi.org/10.1214/20-AOS1945

ANALYSIS OF A TWO-LAYER NEURAL NETWORK VIA DISPLACEMENT CONVEXITY

BY ADEL JAVANMARD[1], MARCO MONDELLI[2] AND ANDREA MONTANARI[3]

[1]*Data Sciences and Operations Department, Marshall School of Business, University of Southern California, ajavanma@usc.edu*

[2]*Institute of Science and Technology (IST) Austria, marco.mondelli@ist.ac.at*

[3]*Department of Electrical Engineering and Department of Statistics, Stanford University, montanar@stanford.edu*

Fitting a function by using linear combinations of a large number N of "simple" components is one of the most fruitful ideas in statistical learning. This idea lies at the core of a variety of methods, from two-layer neural networks to kernel regression, to boosting. In general, the resulting risk minimization problem is nonconvex and is solved by gradient descent or its variants. Unfortunately, little is known about global convergence properties of these approaches.

Here, we consider the problem of learning a concave function f on a compact convex domain $\Omega \subset \mathbb{R}^d$, using linear combinations of "bump-like" components (neurons). The parameters to be fitted are the centers of N bumps, and the resulting empirical risk minimization problem is highly nonconvex. We prove that, in the limit in which the number of neurons diverges, the evolution of gradient descent converges to a Wasserstein gradient flow in the space of probability distributions over Ω. Further, when the bump width δ tends to 0, this gradient flow has a limit which is a viscous porous medium equation. Remarkably, the cost function optimized by this gradient flow exhibits a special property known as *displacement convexity*, which implies exponential convergence rates for $N \to \infty$, $\delta \to 0$.

Surprisingly, this asymptotic theory appears to capture well the behavior for moderate values of δ, N. Explaining this phenomenon, and understanding the dependence on δ, N in a quantitative manner remains an outstanding challenge.

REFERENCES

[1] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

[2] ANTHONY, M. and BARTLETT, P. L. (2009). *Neural Network Learning*: *Theoretical Foundations*. Cambridge Univ. Press. MR1741038 https://doi.org/10.1017/CBO9780511624216

[3] BACH, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J*. *Mach*. *Learn*. *Res*. **18** Paper No. 19, 53. MR3634886

[4] BAKSHI, A., JAYARAM, R. and WOODRUFF, D. P. (2018). Learning two layer rectified neural networks in polynomial time. arXiv:1811.01885.

[5] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans*. *Inform*. *Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

[6] BARTLETT, P. L. (1998). The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans*. *Inform*. *Theory* **44** 525–536. MR1607706 https://doi.org/10.1109/18.661502

[7] BENGIO, Y., ROUX, N. L., VINCENT, P., DELALLEAU, O. and MARCOTTE, P. (2006). Convex neural networks. In *Advances in Neural Information Processing Systems* 123–130.

[8] BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J*. *Amer*. *Statist*. *Assoc*. **98** 324–339. MR1995709 https://doi.org/10.1198/016214503000125

MSC2020 subject classifications. Primary 62F10, 62J02; secondary 62H12.

Key words and phrases. Neural networks, stochastic gradient descent, Wasserstein gradient flow, function regression, convergence rate, displacement convexity.

[9] Candès, E. J. and Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956. MR3193963 https://doi.org/10.1002/cpa.21455

[10] Carrillo, J. A., Jüngel, A., Markowich, P. A., Toscani, G. and Unterreiter, A. (2001). Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Monatsh. Math.* **133** 1–82. MR1853037 https://doi.org/10.1007/s006050170032

[11] Carrillo, J. A., McCann, R. J. and Villani, C. (2003). Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.* **19** 971–1018. MR2053570 https://doi.org/10.4171/RMI/376

[12] Carrillo, J. A., McCann, R. J. and Villani, C. (2006). Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.* **179** 217–263. MR2209130 https://doi.org/10.1007/s00205-005-0386-1

[13] Chen, Y. and Samworth, R. J. (2016). Generalized additive and index models with shape constraints. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 729–754. MR3534348 https://doi.org/10.1111/rssb.12137

[14] Chizat, L. and Bach, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems*.

[15] Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge Univ. Press.

[16] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2** 303–314. MR1015670 https://doi.org/10.1007/BF02551274

[17] Dobrushin, R. L. (1979). Vlasov equations. *Funct. Anal. Appl.* **13** 115–123. MR0541637

[18] Donoho, D. L. (1992). Superresolution via sparsity constraints. *SIAM J. Math. Anal.* **23** 1309–1331. MR1177792 https://doi.org/10.1137/0523074

[19] Du, S. S., Zhai, X., Poczos, B. and Singh, A. (2018). Gradient descent provably optimizes over-parameterized neural networks. arXiv:1810.02054.

[20] Figalli, A. and Philipowski, R. (2008). Convergence to the viscous porous medium equation and propagation of chaos. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 185–203. MR2421181

[21] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328 https://doi.org/10.1214/aos/1013203451

[22] Hannah, L. A. and Dunson, D. B. (2013). Multivariate convex regression with adaptive partitioning. *J. Mach. Learn. Res.* **14** 3261–3294. MR3144462

[23] Javanmard, A., Mondelli, M. and Montanari, A. (2020). Supplement to "Analysis of a two-layer neural network via displacement convexity." https://doi.org/10.1214/20-AOS1945SUPP.

[24] Li, Y. and Yuan, Y. (2017). Convergence analysis of two-layer neural networks with relu activation. In *Advances in Neural Information Processing Systems* 597–607.

[25] Lions, P.-L. and Sznitman, A.-S. (1984). Stochastic differential equations with reflecting boundary conditions. *Comm. Pure Appl. Math.* **37** 511–537. MR0745330 https://doi.org/10.1002/cpa.3160370408

[26] McCann, R. J. (1997). A convexity principle for interacting gases. *Adv. Math.* **128** 153–179. MR1451422 https://doi.org/10.1006/aima.1997.1634

[27] Mei, S., Bai, Y. and Montanari, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

[28] Mei, S., Misiakiewicz, T. and Montanari, A. (2019). Mean-field theory of two-layers neural networks: Dimension-free bounds and kernel limit. In *Conference on Learning Theory* (*COLT*).

[29] Mei, S., Montanari, A. and Nguyen, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115

[30] Nitanda, A. and Suzuki, T. (2017). Stochastic particle gradient descent for infinite ensembles. arXiv:1712.05438.

[31] Oelschläger, K. (2002). Simulation of the solution of a viscous porous medium equation by a particle method. *SIAM J. Numer. Anal.* **40** 1716–1762. MR1950620 https://doi.org/10.1137/S0036142900363377

[32] Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Comput.* **3** 246–257.

[33] Philipowski, R. (2007). Interacting diffusions approximating the porous medium equation and propagation of chaos. *Stochastic Process. Appl.* **117** 526–538. MR2305385 https://doi.org/10.1016/j.spa.2006.09.003

[34] Rahimi, A. and Recht, B. (2008). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems* 1177–1184.

[35] Rosenblatt, F. (1962). *Principles of Neurodynamics*. Spartan Book.

[36] ROTSKOFF, G. M. and VANDEN-EIJNDEN, E. (2018). Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error. In *Advances in Neural Information Processing Systems*.

[37] 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

[38] SCHAPIRE, R. E. (2003). The boosting approach to machine learning: An overview. In *Nonlinear Estimation and Classification* (*Berkeley*, *CA*, 2001). *Lect*. *Notes Stat*. **171** 149–171. Springer, New York. MR2005788 https://doi.org/10.1007/978-0-387-21579-2_9

[39] SIRIGNANO, J. and SPILIOPOULOS, K. (2018). Mean field analysis of neural networks. arXiv:1805.01053.

[40] SOLTANOLKOTABI, M., JAVANMARD, A. and LEE, J. D. (2019). Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Trans*. *Inform*. *Theory* **65** 742–769. MR3904911 https://doi.org/10.1109/TIT.2018.2854560

[41] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math*. **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[42] TANAKA, H. (1979). Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math*. *J*. **9** 163–177. MR0529332

[43] THOMAS, J. W. (2013). *Numerical Partial Differential Equations*: *Finite Difference Methods* **22**. Springer Science & Business Media. MR1367964 https://doi.org/10.1007/978-1-4899-7278-1

[44] TIAN, Y. (2017). Symmetry-breaking convergence analysis of certain two-layered neural networks with ReLU nonlinearity. In *Workshop at International Conference on Learning Representation* (*ICLR*).

[45] VÁZQUEZ, J. L. (2007). *The Porous Medium Equation*: *Mathematical Theory*. *Oxford Mathematical Monographs*. Oxford University Press, Oxford. MR2286292

[46] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[47] WEI, C., LEE, J. D., LIU, Q. and MA, T. (2018). On the margin theory of feedforward neural networks. arXiv:1810.05369.

[48] ZHONG, K., SONG, Z., JAIN, P., BARTLETT, P. L. and DHILLON, I. S. (2017). Recovery guarantees for one-hidden-layer neural networks. In *International Conference on Machine Learning* 4140–4149.

The Annals of Statistics
2020, Vol. 48, No. 6, 3643–3671
https://doi.org/10.1214/20-AOS1946

BEYOND GAUSSIAN APPROXIMATION: BOOTSTRAP FOR MAXIMA OF SUMS OF INDEPENDENT RANDOM VECTORS

BY HANG DENG[*] AND CUN-HUI ZHANG[†]

Department of Statistics, Rutger University, [*]hdeng@stat.rutgers.edu; [†]czhang@stat.rutgers.edu

The Bonferroni adjustment, or the union bound, is commonly used to study rate optimality properties of statistical methods in high-dimensional problems. However, in practice, the Bonferroni adjustment is overly conservative. The extreme value theory has been proven to provide more accurate multiplicity adjustments in a number of settings, but only on an ad hoc basis. Recently, Gaussian approximation has been used to justify bootstrap adjustments in large scale simultaneous inference in some general settings when $n \gg (\log p)^7$, where p is the multiplicity of the inference problem and n is the sample size. The thrust of this theory is the validity of the Gaussian approximation for maxima of sums of independent random vectors in high dimension. In this paper, we reduce the sample size requirement to $n \gg (\log p)^5$ for the consistency of the empirical bootstrap and the multiplier/wild bootstrap in the Kolmogorov–Smirnov distance, possibly in the regime where the Gaussian approximation is not available. New comparison and anticoncentration theorems, which are of considerable interest in and of themselves, are developed as existing ones interweaved with Gaussian approximation are no longer applicable or strong enough to produce desired results.

REFERENCES

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

BENTKUS, V. Y. (1986). Dependence of the Berry–Esseen estimate on the dimension. *Litovsk. Mat. Sb.* **26** 205–210. MR0862741

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

BLANCHET, J., KANG, Y. and MURTHY, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *J. Appl. Probab.* **56** 830–857. MR4015639 https://doi.org/10.1017/jpr.2019.49

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

CHANG, J., ZHOU, W., ZHOU, W.-X. and WANG, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41. MR3632349 https://doi.org/10.1111/biom.12552

CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *Ann. Statist.* **43** 1896–1928. MR3375871 https://doi.org/10.1214/15-AOS1329

MSC2020 subject classifications. 62F40, 62E17.

Key words and phrases. Empirical bootstrap, multiplier bootstrap, wild bootstrap, Lindeberg interpolation, Gaussian approximation, multiple testing, simultaneous confidence intervals, maxima of sums, comparison theorem, anticoncentration.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. MR3750891 https://doi.org/10.1080/01621459.2016.1228536

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

DENG, H. and ZHANG, C.-H. (2020). Supplement to "Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors." https://doi.org/10.1214/20-AOS1946SUPP

DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

FAN, J. and ZHOU, W.-X. (2016). Guarding against spurious discoveries in high dimensions. *J. Mach. Learn. Res.* **17** 1–34. MR3580356

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971 https://doi.org/10.1214/aop/1176996452

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. MR0959185 https://doi.org/10.1214/aos/1176350933

HALL, P. and PRESNELL, B. (1999). Intentionally biased bootstrap methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 143–158. MR1664116 https://doi.org/10.1111/1467-9868.00168

HOROWITZ, J. L. (2019). Bootstrap methods in econometrics. *Ann. Rev. Econ.* **11**.

KLIVANS, A. R., O'DONNELL, R. and SERVEDIO, R. A. (2008). Learning geometric concepts via Gaussian surface area. In *Foundations of Computer Science*, 2008. *FOCS'*08. *IEEE* 49*th Annual IEEE Symposium on* 541–550. IEEE Press, New York.

LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR0964947 https://doi.org/10.1214/aos/1176351062

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 https://doi.org/10.1214/aos/1176349025

NAGAEV, S. V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory* (*Tashkent*, 1975). *Lecture Notes in Math.* **550** 419–438. MR0443043

NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **1807** 169–187. Springer, Berlin. MR2083397 https://doi.org/10.1007/978-3-540-36428-3_15

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

SAZONOV, V. V. (1981). *Normal Approximation—Some Recent Advances*. *Lecture Notes in Math.* **879**. Springer, Berlin. MR0643968

SENATOV, V. V. (1980). Some uniform estimates of the convergence rate in the multidimensional central limit theorem. *Teor. Veroyatn. Primen.* **25** 757–770. MR0595137

SHAO, J. and TU, D. S. (1995). *The Jackknife and Bootstrap*. *Springer Series in Statistics*. Springer, New York. MR1351010 https://doi.org/10.1007/978-1-4612-0795-5

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195. MR0630102

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183 https://doi.org/10.1002/j.1538-7305.1962.tb02419.x

STEIGER, W. L. (1969). A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Stat.* **40** 764–769. MR0240843 https://doi.org/10.1214/aoms/1177697586

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. MR0868303 https://doi.org/10.1214/aos/1176350142

ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

ZHILOVA, M. (2016). Non-classical Berry–Esseen inequality and accuracy of the weighted bootstrap. Preprint. Available at arXiv:1611.02686.

The Annals of Statistics
2020, Vol. 48, No. 6, 3672–3698
https://doi.org/10.1214/20-AOS1947

ISOTONIC REGRESSION IN MULTI-DIMENSIONAL SPACES AND GRAPHS

BY HANG DENG[*] AND CUN-HUI ZHANG[†]

Department of Statistics, Rutgers University, [*]*hdeng@stat.rutgers.edu;* [†]*czhang@stat.rutgers.edu*

In this paper, we study minimax and adaptation rates in general isotonic regression. For uniform deterministic and random designs in $[0,1]^d$ with $d \geq 2$ and $N(0,1)$ noise, the minimax rate for the ℓ_2 risk is known to be bounded from below by $n^{-1/d}$ when the unknown mean function f is nondecreasing and its range is bounded by a constant, while the least squares estimator (LSE) is known to nearly achieve the minimax rate up to a factor $(\log n)^\gamma$ where n is the sample size, $\gamma = 4$ in the lattice design and $\gamma = \max\{9/2, (d^2 + d + 1)/2\}$ in the random design. Moreover, the LSE is known to achieve the adaptation rate $(K/n)^{-2/d}\{1 \vee \log(n/K)\}^{2\gamma}$ when f is piecewise constant on K hyperrectangles in a partition of $[0,1]^d$.

Due to the minimax theorem, the LSE is identical on every design point to both the max-min and min-max estimators over all upper and lower sets containing the design point. This motivates our consideration of estimators which lie in-between the max-min and min-max estimators over possibly smaller classes of upper and lower sets, including a subclass of block estimators. Under a qth moment condition on the noise, we develop ℓ_q risk bounds for such general estimators for isotonic regression on graphs. For uniform deterministic and random designs in $[0,1]^d$ with $d \geq 3$, our ℓ_2 risk bound for the block estimator matches the minimax rate $n^{-1/d}$ when the range of f is bounded and achieves the near parametric adaptation rate $(K/n)\{1 \vee \log(n/K)\}^d$ when f is K-piecewise constant. Furthermore, the block estimator possesses the following oracle property in variable selection: When f depends on only a subset S of variables, the ℓ_2 risk of the block estimator automatically achieves up to a poly-logarithmic factor the minimax rate based on the oracular knowledge of S.

REFERENCES

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

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

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316

BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420

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

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

DENG, H. and ZHANG, C.-H. (2020). Supplement to "Isotonic regression in multi-dimensional spaces and graphs." https://doi.org/10.1214/20-AOS1947SUPP

DONOHO, D. L. (1990). Gelfand n-widths and the method of least squares. Preprint.

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

MSC2020 subject classifications. 62G05, 62G08, 62C20.

Key words and phrases. Isotonic regression, multiple isotonic regression, isotonic regression on graphs, max-min estimator, min-max estimator, block estimator, lattice design, random design, minimax rate, adaptive estimation, variable selection, oracle property.

DUROT, C. (2008). Monotone nonparametric regression with random design. *Math. Methods Statist.* **17** 327–341. MR2483461 https://doi.org/10.3103/S1066530708040042

DYKSTRA, R. L. (1983). An algorithm for restricted least squares regression. *J. Amer. Statist. Assoc.* **78** 837–842. MR0727568

FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2017). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. Preprint. Available at arXiv:1710.04813.

GAO, C., HAN, F. and ZHANG, C.-H. (2017). Minimax risk bounds for piecewise constant models. Preprint. Available at arXiv:1705.06386.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415 https://doi.org/10.1080/03461238.1956.10414944

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

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

KYNG, R., RAO, A. and SACHDEVA, S. (2015). Fast, provable algorithms for isotonic regression in all l_p-norms. In *Advances in Neural Information Processing Systems* 2719–2727.

MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104. MR1810920 https://doi.org/10.1214/aos/1015956708

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

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

STOUT, Q. F. (2015). Isotonic regression for multiple independent variables. *Algorithmica* **71** 450–470. MR3331888 https://doi.org/10.1007/s00453-013-9814-z

VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. MR1056343 https://doi.org/10.1214/aos/1176347632

VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44. MR1212164 https://doi.org/10.1214/aos/1176349013

WANG, Y. (1996). The L_2 risk of an isotonic estimate. *Comm. Statist. Theory Methods* **25** 281–294. MR1379445 https://doi.org/10.1080/03610929608831695

WOODROOFE, M. and SUN, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3** 501–515. MR1243398

YANG, F. and BARBER, R. F. (2019). Contraction and uniform convergence of isotonic regression. *Electron. J. Stat.* **13** 646–677. MR3914177 https://doi.org/10.1214/18-ejs1520

ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

The Annals of Statistics
2020, Vol. 48, No. 6, 3699–3720
https://doi.org/10.1214/20-AOS1948

ROBUST BAYES-LIKE ESTIMATION: RHO-BAYES ESTIMATION

BY YANNICK BARAUD[1] AND LUCIEN BIRGÉ[2]

[1]*Department of Mathematics, University of Luxembourg, yannick.baraud@uni.lu*
[2]*CNRS, Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, lucien.birge@upmc.fr*

We observe n independent random variables with joint distribution $\mathbf{P}$ and pretend that they are i.i.d. with some common density s (with respect to a known measure μ) that we wish to estimate. We consider a density model $\overline{S}$ for s that we endow with a prior distribution π (with support in $\overline{S}$) and build a robust alternative to the classical Bayes posterior distribution which possesses similar concentration properties around s whenever the data are truly i.i.d. and their density s belongs to the model $\overline{S}$. Furthermore, in this case, the Hellinger distance between the classical and the robust posterior distributions tends to 0, as the number of observations tends to infinity, under suitable assumptions on the model and the prior. However, unlike what happens with the classical Bayes posterior distribution, we show that the concentration properties of this new posterior distribution are still preserved when the model is misspecified or when the data are not i.i.d. but the marginal densities of their joint distribution are close enough in Hellinger distance to the model $\overline{S}$.

REFERENCES

ATCHADÉ, Y. A. (2017). On the contraction properties of some high-dimensional quasi-posterior distributions. *Ann. Statist.* **45** 2248–2273. MR3718168 https://doi.org/10.1214/16-AOS1526

BARAUD, Y. (2016). Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.* **10** 1709–1728. MR3522658 https://doi.org/10.1214/15-EJS1055

BARAUD, Y. and BIRGÉ, L. (2016). Rho-estimators for shape restricted density estimation. *Stochastic Process. Appl.* **126** 3888–3912. MR3565484 https://doi.org/10.1016/j.spa.2016.04.013

BARAUD, Y. and BIRGÉ, L. (2017). Robust bayes-like estimation: Rho-bayes estimation. Technical report, arXiv:1711.08328v1.

BARAUD, Y. and BIRGÉ, L. (2018). Rho-estimators revisited: General theory and applications. *Ann. Statist.* **46** 3767–3804. MR3852668 https://doi.org/10.1214/17-AOS1675

BARAUD, Y. and BIRGÉ, L. (2020). Supplement to "Robust Bayes-like estimation: Rho-Bayes estimation." https://doi.org/10.1214/20-AOS1948SUPP

BARAUD, Y., BIRGÉ, L. and SART, M. (2017). A new method for estimation and model selection: ρ-estimation. *Invent. Math.* **207** 425–517. MR3595933 https://doi.org/10.1007/s00222-016-0673-5

BHATTACHARYA, A., PATI, D. and YANG, Y. (2019). Bayesian fractional posteriors. *Ann. Statist.* **47** 39–66. MR3909926 https://doi.org/10.1214/18-AOS1712

BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480

BIRGÉ, L. (1984). Stabilité et instabilité du risque minimax pour des variables indépendantes équidistribuées. *Ann. Inst. Henri Poincaré Probab. Stat.* **20** 201–223. MR0762855

BIRGÉ, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291. MR0816706 https://doi.org/10.1007/BF00332312

BIRGÉ, L. (2006a). Model selection via testing: An alternative to (penalized) maximum likelihood estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 273–325. MR2219712 https://doi.org/10.1016/j.anihpb.2005.04.004

BIRGÉ, L. (2006b). Statistical estimation with model selection. *Indag. Math.* (*N.S.*) **17** 497–537. MR2320111 https://doi.org/10.1016/S0019-3577(07)00004-3

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1103–1130. MR3557191 https://doi.org/10.1111/rssb.12158

MSC2020 subject classifications. 62G35, 62F15, 62G05, 62G07, 62C20, 62F99.
Key words and phrases. Bayesian estimation, rho-Bayes estimation, robust estimation, density estimation, statistical models, metric dimension, VC-classes.

CATONI, O. (2007). *Pac-Bayesian Supervised Classification*: *The Thermodynamics of Statistical Learning*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **56**. IMS, Beachwood, OH. MR2483528

CHERNOZHUKOV, V. and HONG, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115** 293–346. MR1984779 https://doi.org/10.1016/S0304-4076(03)00100-3

DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Springer, Berlin. MR1261635 https://doi.org/10.1007/978-3-662-02888-9

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann*. *Statist*. **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York–Berlin. Translated from the Russian by Samuel Kotz. MR0620321

JIANG, W. and TANNER, M. A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann*. *Statist*. **36** 2207–2231. MR2458185 https://doi.org/10.1214/07-AOS547

KLEIJN, B. J. K. and VAN DER VAART, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann*. *Statist*. **34** 837–877. MR2283395 https://doi.org/10.1214/009053606000000029

KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein–Von-Mises theorem under misspecification. *Electron*. *J*. *Stat*. **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675

LE CAM, L. (1975). On local and global properties in the theory of asymptotic normality of experiments. In *Stochastic Processes and Related Topics* (*Proc*. *Summer Res*. *Inst*. *Statist*. *Inference for Stochastic Processes*, *Indiana Univ*., *Bloomington*, *Ind*., 1974, *Vol*. 1; *Dedicated to Jerzy Neyman*) 13–54. MR0395005

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

LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann*. *Statist*. **1** 38–53. MR0334381

MASSART, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math*. **1896**. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879

PANOV, M. and SPOKOINY, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal*. **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

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

THE ANNALS

of

STATISTICS

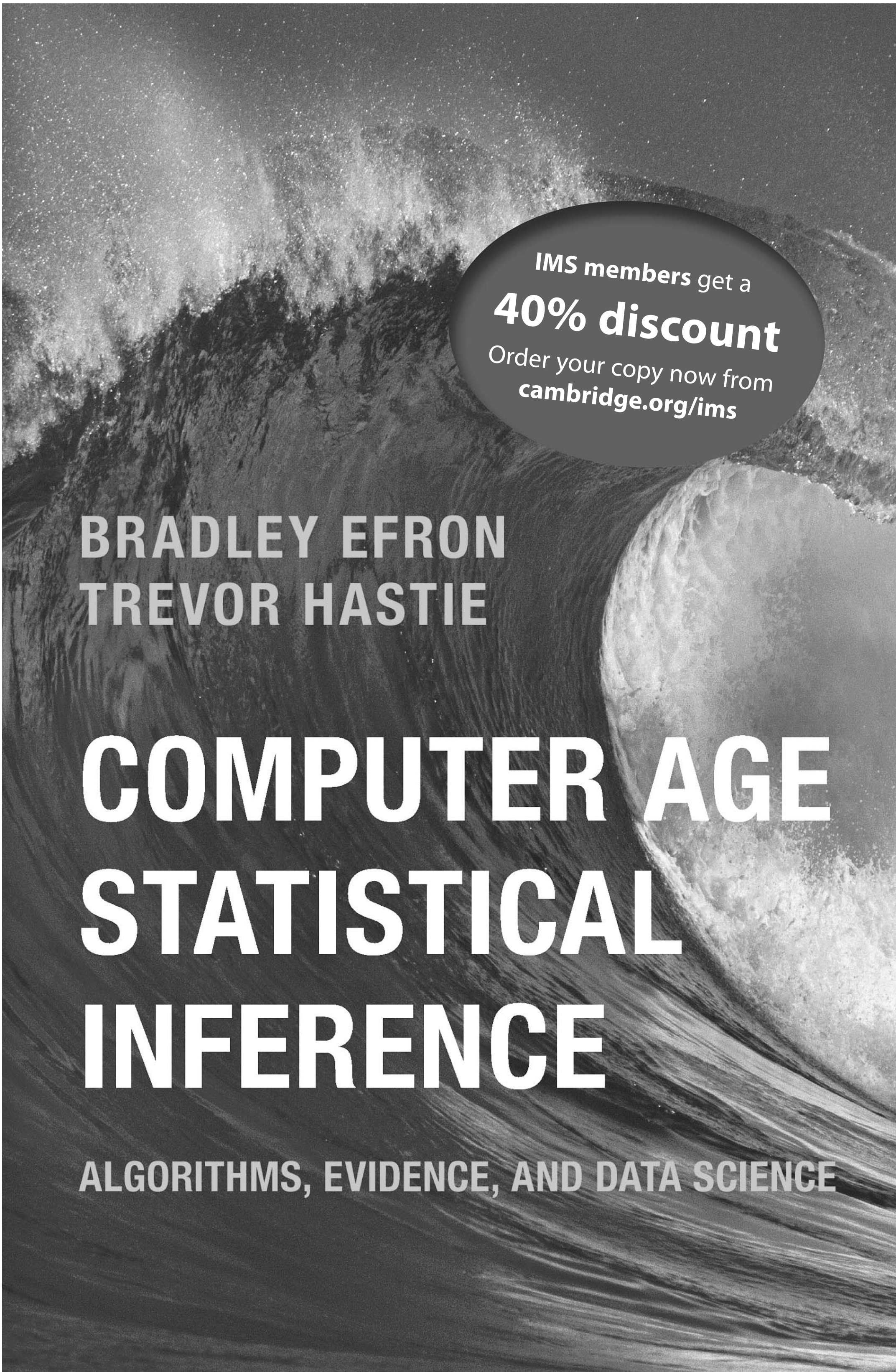
An Official Journal of the
Institute of Mathematical Statistics

VOLUME 48

2020

CONTENTS OF VOLUME 48

Articles


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE