BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty Number One February 2024 ISSN: 1350-7265

CONTENTS

(continued)

A list of forthcoming papers can be found online at https://www.bernoullisociety.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty Number One February 2024 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 30 Number 1 February 2024 Pages 1–850 ISI/BS

Volume 30 Number 1   February 2024   ISSN 1350-7265

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Aims and Scope
Issued four times per year, BERNOULLI is the flagship journal of the Bernoulli Society for Mathematical Statistics and Probability. The journal aims at publishing original research contributions of the highest quality in all subfields of Mathematical Statistics and Probability. The main emphasis of Bernoulli is on theoretical work, yet discussion of interesting applications in relation to the proposed methodology is also welcome.

Bernoulli Society for Mathematical Statistics and Probability
The Bernoulli Society was founded in 1973. It is an autonomous Association of the International Statistical Institute, ISI. According to its statutes, the object of the Bernoulli Society is the advancement, through international contacts, of the sciences of probability (including the theory of stochastic processes) and mathematical statistics and of their applications to all those aspects of human endeavour which are directed towards the increase of natural knowledge and the welfare of mankind.

Meetings: https://www.bernoullisociety.org/meetings
The Society holds a World Congress every four years; more frequent meetings, coordinated by the Society's standing committees and often organised in collaboration with other organisations, are the European Meeting of Statisticians, the Conference on Stochastic Processes and their Applications, the CLAPEM meeting (Latin-American Congress on Probability and Mathematical Statistics), the European Young Statisticians Meeting, and various meetings on special topics – in the physical sciences in particular. The Society, as an association of the ISI, also collaborates with other ISI associations in the organization of the biennial ISI World Statistics Congresses (formerly ISI Sessions).

Executive Committee
Detailed information about the members of the Executive Committee can be found on https://www.bernoullisociety.org/who-is-who

---

The papers published in Bernoulli are indexed or abstracted in Mathematical Reviews (*MathSciNet*), Zentralblatt MATH (*zbMATH Open*), Science Citation Index Expanded (*Web of Science*), *SCOPUS* and *Google Scholar*.

In 2024 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **30**(1), 2024, 1–21
https://doi.org/10.3150/22-BEJ1477

Identifiability in robust estimation of tree structured models

MARTA CASANELLAS[1,2,a], MARINA GARROTE-LÓPEZ[3,b] and PIOTR ZWIERNIK[4,c]

[1] *Institut de Matemàtiques de la UPC-BarcelonaTech (IMTech), Universitat Politècnica de Catalunya, Barcelona, Spain,* [a]*marta.casanellas@upc.edu*
[2] *Centre de Recerca Matemàtica, Barcelona, Spain*
[3] *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany,* [b]*marina.garrote@mis.mpg.de*
[4] *Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [c]*piotr.zwiernik@utoronto.ca*

Consider the problem of learning undirected graphical models on trees from corrupted data. Recently Katiyar, Shah, and Caramanis showed that it is possible to recover trees from noisy binary data up to a small equivalence class of possible trees. Another paper by Katiyar, Hoffmann, and Caramanis follows a similar pattern for the Gaussian case. By framing this as a special phylogenetic recovery problem we largely generalize these two settings. Using the framework of linear latent tree models we discuss tree identifiability for binary data under a continuous corruption model (e.g. black/white images with greyscale corruption). For the Ising and the Gaussian tree model we also provide a characterisation of when the Chow-Liu algorithm consistently learns the underlying tree from the noisy data.

Keywords: Learning tree structure; noisy data on trees; latent tree models

References

[1] Anandkumar, A., Chaudhuri, K., Hsu, D.J., Kakade, S.M., Song, L. and Zhang, T. (2011). Spectral methods for learning multivariate latent tree structure. In *Advances in Neural Information Processing Systems* 2025–2033.

[2] Badrinarayanan, V., Budvytis, I. and Cipolla, R. (2013). Semi-supervised video segmentation using tree structured graphical models. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 2751–2764. https://doi.org/10.1109/TPAMI.2013.54

[3] Bresler, G. and Karzand, M. (2020). Learning a tree-structured Ising model in order to make predictions. *Ann. Statist.* **48** 713–737. MR4102673 https://doi.org/10.1214/19-AOS1808

[4] Buneman, P. (1971). The recovery of trees from measures of dissimilarity. In *Mathematics in the Archaeological and Historical Sciences* 387–395. Edinburgh University Press.

[5] Chang, J.T. (1996). Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency. *Math. Biosci.* **137** 51–73. MR1410044 https://doi.org/10.1016/S0025-5564(96)00075-2

[6] Choi, M.J., Tan, V.Y.F., Anandkumar, A. and Willsky, A.S. (2011). Learning latent tree graphical models. *J. Mach. Learn. Res.* **12** 1771–1812. MR2813153

[7] Chow, C.K. and Liu, C.N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **14** 462–467.

[8] Crouse, M.S., Nowak, R.D. and Baraniuk, R.G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Process.* **46** 886–902. MR1665651 https://doi.org/10.1109/78.668544

[9] D'Elia, C., Poggi, G. and Scarpa, G. (2003). A tree-structured Markov random field model for Bayesian image segmentation. *IEEE Trans. Image Process.* **12** 1259–1273. MR2007536 https://doi.org/10.1109/TIP.2003.817257

1350-7265

[10] Edwards, D., Abreu, D., Gabriel, C.G. and Labouriau, R. (2010). Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests. *BMC Bioinform.* **11** 18.
[11] Everitt, W.N. (1958). A note on positive definite matrices. *Glasg. Math. J.* **3** 173–175. MR0096690
[12] Jukes, T.H. and Cantor, C.R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* 21–132.
[13] Katiyar, A., Hoffmann, J. and Caramanis, C. (2019). Robust estimation of tree structured Gaussian graphical model. In *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 3292–3300. PMLR.
[14] Katiyar, A., Shah, V. and Caramanis, C. (2020). Robust estimation of tree structured ising models. arXiv preprint. Available at arXiv:2006.05601.
[15] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: Oxford Univ. Press. MR1419991
[16] Mossel, E., Roch, S. and Sly, A. (2013). Robust estimation of latent tree graphical models: Inferring hidden states with inexact parameters. *IEEE Trans. Inf. Theory* **59** 4357–4373. MR3071334 https://doi.org/10.1109/TIT.2013.2251927
[17] Nikolakakis, K.E., Kalogerias, D.S. and Sarwate, A.D. (2021). Predictive learning on hidden tree-structured Ising models. *J. Mach. Learn. Res.* **22** Paper No. 59, 82. MR4253752
[18] Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4** 406–425.
[19] Schliep, K.P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics* **27** 592–593.
[20] Semple, C. and Steel, M. (2003). *Phylogenetics*. *Oxford Lecture Series in Mathematics and Its Applications* **24**. Oxford: Oxford Univ. Press. MR2060009
[21] Tandon, A., Tan, V.Y.F. and Zhu, S. (2020). Exact asymptotics for learning tree-structured graphical models with side information: Noiseless and noisy samples. *IEEE J. Sel. Areas Inf. Theory* **1** 760–776.
[22] Tandon, A., Yuan, A.H.J. and Tan, V.Y.F. (2021). SGA: A robust algorithm for partial recovery of tree-structured graphical models with noisy samples. arXiv preprint. Available at arXiv:2101.08917.
[23] Zhang, N.L. (2003/04). Hierarchical latent class models for cluster analysis. *J. Mach. Learn. Res.* **5** 697–723. MR2247997
[24] Zwiernik, P. (2016). *Semialgebraic Statistics and Latent Tree Models*. *Monographs on Statistics and Applied Probability* **146**. Boca Raton, FL: CRC Press/CRC. MR3379921
[25] Zwiernik, P. (2019). Latent tree models. In *Handbook of Graphical Models*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 265–288. Boca Raton, FL: CRC Press. MR3888000

Bernoulli **30**(1), 2024, 22–45
https://doi.org/10.3150/22-BEJ1513

Flexible-bandwidth needlets

CLAUDIO DURASTANTI[1,a], DOMENICO MARINUCCI[2,b] and ANNA PAOLA TODINO[3,c]

[1]*Dipartimento SBAI, La Sapienza Università di Roma,* [a]*claudio.durastanti@uniroma1.it*
[2]*Dipartimento di Matematica, Università di Roma Tor Vergata,* [b]*marinucc@mat.uniroma2.it*
[3]*Dipartimento di Scienze e Innovazione Tecnologica, Università del Piemonte Orientale,* [c]*annapaola.todino@uniupo.it*

We investigate here a generalized construction of spherical wavelets/needlets which admits extra-flexibility in the harmonic space, i.e., it allows the corresponding support in multipole (frequency) space to vary in more general forms than in the standard constructions. We study the analytic properties of this system and we investigate its behaviour when applied to isotropic random fields: more precisely, we establish asymptotic localization and uncorrelation properties (in the high-frequency sense) under broader assumptions than typically considered in the literature.

Keywords: Spherical wavelets; needlets; spherical random fields; high-frequency asymptotics

References

[1] 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**. Washington, DC: U. S. Government Printing Office. MR0167642

[2] Atkinson, K. and Han, W. (2012). *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. *Lecture Notes in Math.* **2044**. Heidelberg: Springer. MR2934227 https://doi.org/10.1007/978-3-642-25983-8

[3] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Asymptotics for spherical needlets. *Ann. Statist.* **37** 1150–1171. MR2509070 https://doi.org/10.1214/08-AOS601

[4] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395. MR2549563 https://doi.org/10.1214/09-AOS682

[5] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Subsampling needlet coefficients on the sphere. *Bernoulli* **15** 438–463. MR2543869 https://doi.org/10.3150/08-BEJ164

[6] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871 https://doi.org/10.1017/CBO9780511721434

[7] Bourguin, S. and Durastanti, C. (2017). On high-frequency limits of U-statistics in Besov spaces over compact manifolds. *Illinois J. Math.* **61** 97–125. MR3770838 https://doi.org/10.1215/ijm/1520046211

[8] Cammarota, V. and Marinucci, D. (2015). On the limiting behaviour of needlets polyspectra. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1159–1189. MR3365977 https://doi.org/10.1214/14-AIHP609

[9] Carrón Duque, J. et al. (2019). Point source detection and false discovery rate control on CMB maps. *Astron. Comput.* **28** 100310.

[10] Cheng, D., Cammarota, V., Fantaye, Y., Marinucci, D. and Schwartzman, A. (2020). Multiple testing of local maxima for detection of peaks on the (celestial) sphere. *Bernoulli* **26** 31–60. MR4036027 https://doi.org/10.3150/18-BEJ1068

[11] Cleanthous, G., Georgiadis, A.G., Kerkyacharian, G., Petrushev, P. and Picard, D. (2020). Kernel and wavelet density estimators on manifolds and more general metric spaces. *Bernoulli* **26** 1832–1862. MR4091093 https://doi.org/10.3150/19-BEJ1171

[12] Coulhon, T., Kerkyacharian, G. and Petrushev, P. (2012). Heat kernel generated frames in the setting of Dirichlet spaces. *J. Fourier Anal. Appl.* **18** 995–1066. MR2970038 https://doi.org/10.1007/s00041-012-9232-7

[13] Durastanti, C. (2016). Adaptive global thresholding on the sphere. *J. Multivariate Anal.* **151** 110–132. MR3545280 https://doi.org/10.1016/j.jmva.2016.07.009

[14] Durastanti, C. (2017). Tail behavior of Mexican needlets. *J. Math. Anal. Appl.* **447** 716–735. MR3573111 https://doi.org/10.1016/j.jmaa.2016.10.046

[15] Durastanti, C., Lan, X. and Marinucci, D. (2013). Needlet-Whittle estimates on the unit sphere. *Electron. J. Stat.* **7** 597–646. MR3035267 https://doi.org/10.1214/13-EJS782

[16] Fan, M., Paul, D., Lee, T.C.M. and Matsuo, T. (2018). A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials. *Ann. Appl. Stat.* **12** 459–489. MR3773401 https://doi.org/10.1214/17-AOAS1104

[17] Geller, D. and Mayeli, A. (2009). Nearly tight frames and space-frequency analysis on compact manifolds. *Math. Z.* **263** 235–264. MR2534117 https://doi.org/10.1007/s00209-008-0406-6

[18] Geller, D. and Pesenson, I.Z. (2011). Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds. *J. Geom. Anal.* **21** 334–371. MR2772076 https://doi.org/10.1007/s12220-010-9150-3

[19] Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

[20] Kerkyacharian, G., Nickl, R. and Picard, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Related Fields* **153** 363–404. MR2925578 https://doi.org/10.1007/s00440-011-0348-5

[21] Kerkyacharian, G., Ogawa, S., Petrushev, P. and Picard, D. (2018). Regularity of Gaussian processes on Dirichlet spaces. *Constr. Approx.* **47** 277–320. MR3769279 https://doi.org/10.1007/s00365-018-9416-8

[22] Kerkyacharian, G. and Petrushev, P. (2015). Heat kernel based decomposition of spaces of distributions in the framework of Dirichlet spaces. *Trans. Amer. Math. Soc.* **367** 121–189. MR3271256 https://doi.org/10.1090/S0002-9947-2014-05993-X

[23] Kerkyacharian, G., Pham Ngoc, T.M. and Picard, D. (2011). Localized spherical deconvolution. *Ann. Statist.* **39** 1042–1068. MR2816347 https://doi.org/10.1214/10-AOS858

[24] Lang, A. and Schwab, C. (2015). Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.* **25** 3047–3094. MR3404631 https://doi.org/10.1214/14-AAP1067

[25] Le Gia, Q.T., Sloan, I.H., Wang, Y.G. and Womersley, R.S. (2017). Needlet approximation for isotropic random fields on the sphere. *J. Approx. Theory* **216** 86–116. MR3612485 https://doi.org/10.1016/j.jat.2017.01.001

[26] Li, M., Broadbridge, P., Olenko, A. and Wang, Y.G. (2019). Fast tensor needlet transforms for tangent vector fields on the sphere. Available at arXiv:1907.13339.

[27] Marinucci, D. et al. (2008). Spherical needlets for cosmic microwave background data analysis. *Mon. Not. R. Astron. Soc.* **383** 539–545.

[28] Marinucci, D. and Peccati, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications*. *London Mathematical Society Lecture Note Series* **389**. Cambridge: Cambridge Univ. Press. MR2840154 https://doi.org/10.1017/CBO9780511751677

[29] Mayeli, A. (2010). Asymptotic uncorrelation for Mexican needlets. *J. Math. Anal. Appl.* **363** 336–344. MR2559069 https://doi.org/10.1016/j.jmaa.2009.07.044

[30] McEwen, J.D., Durastanti, C. and Wiaux, Y. (2018). Localisation of directional scale-discretised wavelets on the sphere. *Appl. Comput. Harmon. Anal.* **44** 59–88. MR3707864 https://doi.org/10.1016/j.acha.2016.03.009

[31] Møller, J., Nielsen, M., Porcu, E. and Rubak, E. (2018). Determinantal point process models on the sphere. *Bernoulli* **24** 1171–1201. MR3706791 https://doi.org/10.3150/16-BEJ896

[32] Narcowich, F., Petrushev, P. and Ward, J. (2006). Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564. MR2253732 https://doi.org/10.1016/j.jfa.2006.02.011

[33] Narcowich, F.J., Petrushev, P. and Ward, J.D. (2006). Localized tight frames on spheres. *SIAM J. Math. Anal.* **38** 574–594. MR2237162 https://doi.org/10.1137/040614359

[34] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. MR2962301 https://doi.org/10.1017/CBO9781139084659

[35] Oppizzi, F., Renzi, A., Liguori, M., Hansen, F.K., Marinucci, D., Baccigalupi, C., Bertacca, D. and Poletti, D. (2020). Needlet thresholding methods in component separation. *J. Cosmol. Astropart. Phys.* **2020** 054, 29 pp. MR4127710 https://doi.org/10.1088/1475-7516/2020/03/054

[36] Pietrobon, D., Amblard, A., Balbi, A., Cabella, P., Cooray, A. and Marinucci, D. (2008). Needlet detection of features in CMB sky and the impact on anisotropies and hemispherical ssymmetries. *Phys. Rev. D* **78** 103504.

[37] Pietrobon, D., Balbi, A. and Marinucci, D. (2006). Integrated sachs-Wolfe effect from the cross correlation of WMAP3 year and the NRAO VLA sky survey data: New results and constraints on dark energy. *Phys. Rev. D* **74** 043524.

[38] Planck Collaboration (2014). Planck 2013 results. XV. CMB power spectra and likelihood. *Astron. Astrophys.* **571** id.A15.

[39] Planck Collaboration (2016). Planck 2015 results. XVI. Isotropy and statistics of the CMB. *Astron. Astrophys.* **594** id.A16.

[40] Porteous, I.R. (2001). *Geometric Differentiation: For the Intelligence of Curves and Surfaces*, 2nd ed. Cambridge: Cambridge Univ. Press. MR1871900

[41] Rudjord, O., Hansen, F.K., Lan, X., Liguori, M., Marinucci, D. and Matarrese, S. (2009). An estimate of the primordial non-Gaussianity parameter f_{NL} using the needlet bispectrum from WMAP. *Astrophys. J.* **701** 369–376.

[42] Shevchenko, R. and Todino, A.P. (2023). Asymptotic behaviour of level sets of needlet random fields. *Stochastic Process. Appl.* **155** 268–318. MR4509490 https://doi.org/10.1016/j.spa.2022.10.011

[43] Trübner, M. and Ziegel, J.F. (2017). Derivatives of isotropic positive definite functions on spheres. *Proc. Amer. Math. Soc.* **145** 3017–3031. MR3637950 https://doi.org/10.1090/proc/13561

[44] Wang, Y.G., Le Gia, Q.T., Sloan, I.H. and Womersley, R.S. (2017). Fully discrete needlet approximation on the sphere. *Appl. Comput. Harmon. Anal.* **43** 292–316. MR3668040 https://doi.org/10.1016/j.acha.2016.01.003

[45] Wang, Y.G., Sloan, I.H. and Womersley, R.S. (2018). Riemann localisation on the sphere. *J. Fourier Anal. Appl.* **24** 141–183. MR3750311 https://doi.org/10.1007/s00041-016-9496-4

[46] Wiaux, Y., McEwen, J.D., Vandergheynst, P. and Blanc, O. (2008). Exact reconstruction with directional wavelets on the sphere. *Mon. Not. R. Astron. Soc.* **388** 770–788.

Bernoulli **30**(1), 2024, 46–71
https://doi.org/10.3150/22-BEJ1560

Statistics for heteroscedastic time series extremes

AXEL BÜCHER[a] and TOBIAS JENNESSEN[b]

Heinrich-Heine-University, Düsseldorf, Germany, [a]*axel.buecher@hhu.de,* [b]*tobias.jennessen@hhu.de*

Einmahl, de Haan and Zhou (2016, Journal of the Royal Statistical Society: Series B, 78(1), 31–51) recently introduced a stochastic model that allows for heteroscedasticity of extremes. The model is extended to the situation where the observations are serially dependent, which is crucial for many practical applications. We prove a local limit theorem for a kernel estimator for the scedasis function, and a functional limit theorem for an estimator for the integrated scedasis function. We further prove consistency of a bootstrap scheme that allows to test for the null hypothesis that the extremes are homoscedastic. Finally, we propose an estimator for the extremal index governing the dynamics of the extremes and prove its consistency. All results are illustrated by Monte Carlo simulations. An important intermediate result concerns the sequential tail empirical process under serial dependence.

Keywords: Extremal index; kernel estimator; multiplier bootstrap; non-stationary extremes; regular varying time series

References

[1] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

[2] Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

[3] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2108013 https://doi.org/10.1002/0470012382

[4] Berghaus, B. and Bücher, A. (2017). Goodness-of-fit tests for multivariate copula-based time series models. *Econometric Theory* **33** 292–330. MR3600045 https://doi.org/10.1017/S0266466615000419

[5] Berghaus, B. and Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* **46** 2307–2335. MR3845019 https://doi.org/10.1214/17-AOS1621

[6] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1324786

[7] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

[8] Bücher, A. and Jennessen, T. (2024). Supplement to "Statistics for heteroscedastic time series extremes." https://doi.org/10.3150/22-BEJ1560SUPP

[9] Bücher, A. and Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *J. Theoret. Probab.* **32** 1145–1165. MR3979663 https://doi.org/10.1007/s10959-018-0823-3

[10] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. London: Springer. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

[11] Dahlhaus, R. (2012). Locally stationary processes. In *Time Series Analysis: Methods and Applications* (T. Subba Rao, S. Subba Rao and C.R. Rao, eds.). *Handbook of Statistics* **30** 351–413. Elsevier.

[12] Davis, R.A., Drees, H., Segers, J. and Warchoł, M. (2018). Inference on the tail process with application to financial time series modeling. *J. Econometrics* **205** 508–525. MR3813529 https://doi.org/10.1016/j.jeconom.2018.01.009

[13] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52** 393–442. With discussion and a reply by the authors. MR1086795

[14] de Haan, L., Klein Tank, A. and Neves, C. (2015). On tail trend detection: Modeling relative risk. *Extremes* **18** 141–178. MR3351812 https://doi.org/10.1007/s10687-014-0207-8

[15] de Haan, L., Resnick, S.I., Rootzén, H. and de Vries, C.G. (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Process. Appl.* **32** 213–224. MR1014450 https://doi.org/10.1016/0304-4149(89)90076-8

[16] de Haan, L. and Zhou, C. (2021). Trends in extreme value indices. *J. Amer. Statist. Assoc.* **116** 1265–1279. MR4309271 https://doi.org/10.1080/01621459.2019.1705307

[17] Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875 https://doi.org/10.1214/aoap/1019487617

[18] Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. ArXiv Preprint. Available at arXiv:1511.00420.

[19] Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886 https://doi.org/10.1214/09-AOS788

[20] Drees, H., Segers, J. and Warchoł, M. (2015). Statistics for tail processes of Markov chains. *Extremes* **18** 369–402. MR3377814 https://doi.org/10.1007/s10687-015-0217-1

[21] Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2** 291–293. MR0777842 https://doi.org/10.1016/0167-7152(84)90067-1

[22] Einmahl, J.H.J., de Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

[23] Einmahl, J.H.J., Ferreira, A., de Haan, L., Neves, C. and Zhou, C. (2022). Spatial dependence and space-time trend in extreme events. *Ann. Statist.* **50** 30–52. MR4383006 https://doi.org/10.1214/21-aos2067

[24] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statist. Sci.* **15** 153–167. MR1788730 https://doi.org/10.1214/ss/1009212755

[25] Hoga, Y. (2017). Change point tests for the tail index of β-mixing random variables. *Econometric Theory* **33** 915–954. MR3667041 https://doi.org/10.1017/S0266466616000189

[26] Hoga, Y. (2018). A structural break test for extremal dependence in β-mixing random vectors. *Biometrika* **105** 627–643. MR3842889 https://doi.org/10.1093/biomet/asy030

[27] Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Stat. Comput.* **3** 135–146.

[28] Kojadinovic, I. and Naveau, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes* **20** 417–450. MR3638372 https://doi.org/10.1007/s10687-016-0273-1

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

[30] Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

[31] Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. MR0722133 https://doi.org/10.1007/BF00532484

[32] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. *Springer Series in Statistics*. New York-Berlin: Springer. MR0691492

[33] Northrop, P.J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18** 585–603. MR3418769 https://doi.org/10.1007/s10687-015-0221-5

[34] Segers, J. (2003). Functionals of clusters of extremes. *Adv. in Appl. Probab.* **35** 1028–1045. MR2014268 https://doi.org/10.1239/aap/1067436333

[35] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[36] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes – Springer Series in Statistics*. New York: Springer.

Bernoulli **30**(1), 2024, 72–87
https://doi.org/10.3150/22-BEJ1564

A note on the empty balls of a critical super-Brownian motion

SHUXIONG ZHANG[1,a] and JIE XIONG[2,b]

[1] *School of Mathematics and Statistics, Anhui Normal University, Wuhu, China,*
[a] *shuxiong.zhang@mail.bnu.edu.cn*
[2] *Department of Mathematics and SUSTech International Center for Mathematics, Southern University of Science and Technology, Shenzhen, China,* [b] *xiongj@sustech.edu.cn*

Let $\{X_t\}_{t\geq 0}$ be a d-dimensional critical super-Brownian motion started from a Poisson random measure whose intensity is the Lebesgue measure. Denote by $R_t := \sup\{u > 0 : X_t(\{x \in \mathbb{R}^d : |x| < u\}) = 0\}$ the radius of the largest empty ball centered at the origin of X_t. In this work, we prove that for $r > 0$,

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{R_t}{t^{(1/d)\wedge(3-d)^+}} \geq r\right) = e^{-A_d(r)},$$

where $A_d(r)$ satisfies

$$\lim_{r\to\infty} \frac{A_d(r)}{r^{|d-2|+d\mathbf{1}_{\{d=2\}}}} = C$$

for some $C \in (0,\infty)$ depending only on d.

Keywords: Empty ball; scaling property; Super-Brownian motion

References

Dawson, D.A. (1977). The critical measure diffusion process. *Z. Wahrsch. Verw. Gebiete* **40** 125–145. MR0478374 https://doi.org/10.1007/BF00532877

Dawson, D.A., Iscoe, I. and Perkins, E.A. (1989). Super-Brownian motion: Path properties and hitting probabilities. *Probab. Theory Related Fields* **83** 135–205. MR1012498 https://doi.org/10.1007/BF00333147

Engländer, J. (2004). Large deviations for the growth rate of the support of supercritical super-Brownian motion. *Statist. Probab. Lett.* **66** 449–456. MR2045138 https://doi.org/10.1016/j.spl.2003.12.003

Etheridge, A.M. (2000). *An Introduction to Superprocesses*. *University Lecture Series* **20**. Providence, RI: Amer. Math. Soc. MR1779100 https://doi.org/10.1090/ulect/020

Gorostiza, L.G. and Wakolbinger, A. (1991). Persistence criteria for a class of critical branching particle systems in continuous time. *Ann. Probab.* **19** 266–288. MR1085336

Hu, Y. (2005). A note on the empty balls left by a critical branching Wiener process. *Period. Math. Hungar.* **50** 165–174. MR2162807 https://doi.org/10.1007/s10998-005-0010-z

Iscoe, I. (1988). On the supports of measure-valued critical branching Brownian motion. *Ann. Probab.* **16** 200–221. MR0920265

Kyprianou, A.E., Liu, R.-L., Murillo-Salas, A. and Ren, Y.-X. (2012). Supercritical super-Brownian motion with a general branching mechanism and travelling waves. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 661–687. MR2976558 https://doi.org/10.1214/11-AIHP448

Le Gall, J.-F. (1999). *Spatial Branching Processes, Random Snakes and Partial Differential Equations*. *Lectures in Mathematics ETH Zürich*. Basel: Birkhäuser. MR1714707 https://doi.org/10.1007/978-3-0348-8683-3

Li, Z. (2011). *Measure-Valued Branching Markov Processes*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2760602 https://doi.org/10.1007/978-3-642-15004-3

Li, Z. (2020). Continuous-state branching processes with immigration. In *From Probability to Finance, Mathematical Lectures from Peking University* (Y. Jiao, ed.). Singapore: Springer.

Perkins, E.A. (2002). *Dawson-Watanabe Superprocesses and Measure-Valued Diffusions*. Berlin: Springer.

Pinsky, R.G. (1995). On the large time growth rate of the support of supercritical super-Brownian motion. *Ann. Probab.* **23** 1748–1754. MR1379166

Pinsky, R.G. (2001). Invariant probability distributions for measure-valued diffusions. *Ann. Probab.* **29** 1476–1514. MR1880229 https://doi.org/10.1214/aop/1015345759

Ren, Y.-X., Song, R. and Zhang, R. (2015). Limit theorems for some critical superprocesses. *Illinois J. Math.* **59** 235–276. MR3459635

Ren, Y.-X., Song, R. and Zhang, R. (2021). The extremal process of super-Brownian motion. *Stochastic Process. Appl.* **137** 1–34. MR4242982 https://doi.org/10.1016/j.spa.2021.03.007

Révész, P. (2002). Large balls left empty by a critical branching Wiener field. *Stat. Neerl.* **56** 195–205.

Zhou, X. (2008). A zero-one law of almost sure local extinction for $(1+\beta)$-super-Brownian motion. *Stochastic Process. Appl.* **118** 1982–1996. MR2462283 https://doi.org/10.1016/j.spa.2007.11.011

Bernoulli **30**(1), 2024, 88–116
https://doi.org/10.3150/22-BEJ1574

On Lasso and Slope drift estimators for Lévy-driven Ornstein–Uhlenbeck processes

NIKLAS DEXHEIMER[a] and CLAUDIA STRAUCH[b]

Aarhus University, Department of Mathematics, Ny Munkegade 118, 8000 Aarhus C, Denmark,
[a]*dexheimer@math.au.dk,* [b]*strauch@math.au.dk*

We investigate the problem of estimating the drift parameter of a high-dimensional Lévy-driven Ornstein–Uhlenbeck process under sparsity constraints. It is shown that both Lasso and Slope estimators achieve the minimax optimal rate of convergence (up to numerical constants), for tuning parameters chosen independently of the confidence level, which improves the previously obtained results for standard Ornstein–Uhlenbeck processes.

Keywords: High-dimensional statistics; Lasso; Ornstein–Uhlenbeck process; parametric statistics; Slope; sparse estimation

References

[1] Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge: Cambridge Univ. Press. MR2512800 https://doi.org/10.1017/CBO9780511809781

[2] Barndorff-Nielsen, O.E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2** 41–68. MR1804664 https://doi.org/10.1007/s007800050032

[3] Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

[4] Bellec, P.C., Lecué, G. and Tsybakov, A.B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663 https://doi.org/10.1214/17-AOS1670

[5] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

[6] Bogdan, M., van den Berg, E., Sabatti, C., Su, W. and Candès, E.J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842

[7] Carmona, R., Fouque, J.-P. and Sun, L.-H. (2015). Mean field games and systemic risk. *Commun. Math. Sci.* **13** 911–933. MR3325083 https://doi.org/10.4310/CMS.2015.v13.n4.a4

[8] Cattiaux, P. and Guillin, A. (2008). Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.* **12** 12–29. MR2367991 https://doi.org/10.1051/ps:2007032

[9] Ciołek, G., Marushkevych, D. and Podolskij, M. (2020). On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.* **14** 4395–4420. MR4194266 https://doi.org/10.1214/20-EJS1775

[10] Fouque, J.-P. and Ichiba, T. (2013). Stability in a model of interbank lending. *SIAM J. Financial Math.* **4** 784–803. MR3119315 https://doi.org/10.1137/110841096

[11] Gaïffas, S. and Matulewicz, G. (2019). Sparse inference of the drift of a high-dimensional Ornstein-Uhlenbeck process. *J. Multivariate Anal.* **169** 1–20. MR3875583 https://doi.org/10.1016/j.jmva.2018.08.005

[12] Ledoux, M. and Talagrand, M. (2011). *Probability in Banach Spaces: Isoperimetry and Processes*. *Classics in Mathematics*. Berlin: Springer. MR2814399

[13] Mai, H. (2012). Drift Estimation for Jump Diffusions Ph.D. thesis Humboldt-Univ. zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.

[14] Mai, H. (2014). Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **20** 919–957. MR3178522 https://doi.org/10.3150/13-BEJ510

1350-7265

[15] Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli* **10** 97–120. MR2044595 https://doi.org/10.3150/bj/1077544605

[16] Nourdin, I. and Viens, F.G. (2009). Density formula and concentration inequalities with Malliavin calculus. *Electron. J. Probab.* **14** 2287–2309. MR2556018 https://doi.org/10.1214/EJP.v14-707

[17] Protter, P.E. (2004). *Stochastic Integration and Differential Equations: Stochastic Modelling and Applied Probability*, 2nd ed. *Applications of Mathematics (New York)* **21**. Berlin: Springer. MR2020294

[18] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR1739520

[19] Sato, K. and Yamazato, M. (1984). Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Process. Appl.* **17** 73–100. MR0738769 https://doi.org/10.1016/0304-4149(84)90312-0

[20] Sørensen, M. (1991). Likelihood methods for diffusions with jumps. In *Statistical Inference in Stochastic Processes*. *Probab. Pure Appl.* **6** 67–105. New York: Dekker. MR1138259

[21] Talagrand, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **60**. Heidelberg: Springer. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

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

[23] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge: Cambridge Univ. Press. MR3837109 https://doi.org/10.1017/9781108231596

Bernoulli **30**(1), 2024, 117–142
https://doi.org/10.3150/23-BEJ1591

Estimation of functional ARMA models

THOMAS KUENZER[1,2,a]

[1]*Institute of Statistics, Graz University of Technology, Graz, Austria.*
[2]*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria,*
[a]*thomas.kuenzer@medunigraz.at*

Functional auto-regressive moving average (FARMA or ARMAH) models allow for flexible and natural modelling of functional time series. While there are many results on pure autoregressive (FAR) models in Hilbert spaces, results on estimation and prediction of FARMA models are considerably more scarce. We devise a simple two-step method to estimate ARMA models in separable Hilbert spaces. Estimation is based on dimension-reduction using principal components analysis of the functional time series. We explore two different approaches to selecting principal component subspaces for regularization and establish consistency of the proposed estimators both under minimal assumptions and in a practical setting. The empirical performance of the estimation algorithm is evaluated in a simulation study, where it performs better than competing methods.

Keywords: FARMA model; functional data analysis; functional time series; model estimation; moving average

References

Aue, A. and Klepsch, J. (2017). Estimating functional time series by moving average model fitting. arXiv preprint. arXiv:1701.00770.

Aue, A., Norinho, D.D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510 https://doi.org/10.1080/01621459.2014.909317

Bosq, D. (2000). *Linear Processes in Function Spaces*. *Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* **19** 2627–2651. MR3160566 https://doi.org/10.3150/12-BEJ469

Dufour, J.-M. and Jouini, T. (2005). Asymptotic distribution of a simple linear estimator for VARMA models in echelon form. In *Statistical Modeling and Analysis for Complex Data Problems*. *GERAD 25th Anniv. Ser.* **1** 209–240. New York: Springer. MR2189538 https://doi.org/10.1007/0-387-24555-3_11

Dufour, J.-M. and Jouini, T. (2014). Asymptotic distributions for quasi-efficient estimators in echelon VARMA models. *Comput. Statist. Data Anal.* **73** 69–86. MR3147975 https://doi.org/10.1016/j.csda.2013.11.002

Gabrys, R., Horváth, L. and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *J. Amer. Statist. Assoc.* **105** 1113–1125. MR2752607 https://doi.org/10.1198/jasa.2010.tm09794

Hannan, E.J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems*. *Classics in Applied Mathematics* **70**. Philadelphia, PA: SIAM. MR3397291 https://doi.org/10.1137/1.9781611972191.ch1

Hannan, E.J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69** 81–94. MR0655673 https://doi.org/10.1093/biomet/69.1.81

Hörmann, S. and Kidziński, Ł. (2015). A note on estimation in Hilbertian linear models. *Scand. J. Stat.* **42** 43–62. MR3318024 https://doi.org/10.1111/sjos.12094

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

Horváth, L., Hušková, M. and Rice, G. (2013). Test of independence for functional data. *J. Multivariate Anal.* **117** 100–119. MR3053537 https://doi.org/10.1016/j.jmva.2013.02.005

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

Izumino, S. (1983). Convergence of generalized inverses and spline projectors. *J. Approx. Theory* **38** 269–278. MR0705545 https://doi.org/10.1016/0021-9045(83)90133-8

Klepsch, J. and Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *J. Multivariate Anal.* **155** 252–271. MR3607894 https://doi.org/10.1016/j.jmva.2017.01.005

Klepsch, J., Klüppelberg, C. and Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econom. Stat.* **1** 128–149. MR3669993 https://doi.org/10.1016/j.ecosta.2016.10.009

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *J. Time Series Anal.* **34** 116–129. MR3008019 https://doi.org/10.1111/j.1467-9892.2012.00816.x

Kuenzer, T., Hörmann, S. and Kokoszka, P. (2021). Principal component analysis of spatially indexed functions. *J. Amer. Statist. Assoc.* **116** 1444–1456. MR4309284 https://doi.org/10.1080/01621459.2020.1732395

Kühnert, S. (2020). Functional ARCH and GARCH models: A Yule-Walker approach. *Electron. J. Stat.* **14** 4321–4360. MR4187136 https://doi.org/10.1214/20-EJS1778

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer. MR2172368 https://doi.org/10.1007/978-3-540-27752-1

Lütkepohl, H. and Poskitt, D.S. (1996). Specification of echelon-form VARMA models. *J. Bus. Econom. Statist.* **14** 69–79.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *J. Multivariate Anal.* **98** 1231–1261. MR2326249 https://doi.org/10.1016/j.jmva.2006.05.010

Mestre, G., Portela, J., Muñoz San Roque, A. and Alonso, E. (2020). Forecasting hourly supply curves in the Italian Day-Ahead electricity market with a double-seasonal SARMAHX model. *Int. J. Electr. Power Energy Syst.* **121** 106083.

Mestre, G., Portela, J., Rice, G., Muñoz San Roque, A. and Alonso, E. (2021). Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis. *Comput. Statist. Data Anal.* **155** Paper No. 107108. MR4163024 https://doi.org/10.1016/j.csda.2020.107108

Mitchell, H. (1996). Topics in Multiple Time Series Ph.D. thesis Royal Melbourne Institute of Technology.

Portela González, J., Muñoz San Roque, A. and Alonso, E. (2018). Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Trans. Power Syst.* **33** 545–556.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

Ruiz-Medina, M.D. (2011). Spatial autoregressive and moving average Hilbertian processes. *J. Multivariate Anal.* **102** 292–305. MR2739116 https://doi.org/10.1016/j.jmva.2010.09.005

Spangenberg, F. (2013). Strictly stationary solutions of ARMA equations in Banach spaces. *J. Multivariate Anal.* **121** 127–138. MR3090473 https://doi.org/10.1016/j.jmva.2013.06.007

Turbillon, C., Bosq, D., Marion, J.-M. and Pumo, B. (2008). Estimation du paramètre des moyennes mobiles hilbertiennes. *C. R. Math. Acad. Sci. Paris* **346** 347–350. MR2414183 https://doi.org/10.1016/j.crma.2008.01.008

Bernoulli **30**(1), 2024, 143–169
https://doi.org/10.3150/23-BEJ1592

Supermartingale shadow couplings: The decreasing case

ERHAN BAYRAKTAR[1,a], SHUOQING DENG[2,c] and DOMINYKAS NORGILAS[1,b]

[1]*Department of Mathematics, University of Michigan, Ann Arbor, USA,* [a]*erhan@umich.edu,* [b]*dnorgila@umich.edu*
[2]*Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China,* [c]*masdeng@ust.hk*

For two measures μ and ν that are in convex-decreasing order, Nutz and Stebegg (Canonical supermartingale couplings, Ann. Probab. 46(6) 3351–3398, 2018) studied the optimal transport problem with supermartingale constraints and introduced two canonical couplings, namely the increasing and decreasing transport plans, that are optimal for a large class of cost functions. In the present paper we provide an explicit construction of the decreasing coupling π^D by establishing a Brenier-type result: (a generalised version of) π^D concentrates on the graphs of two functions. Our construction is based on the concept of the supermartingale *shadow* measure and requires a suitable extension of the results by Juillet (Stability of the shadow projection and the left-curtain coupling, Ann. Inst. H. Poincaré Probab. Statist. 52(4) 1823–1843, November 2016) and Beiglböck and Juillet (Shadow couplings, Trans. Amer. Math. Soc. 374 4973–5002, 2021) established in the martingale setting. In particular, we prove the stability of the supermartingale shadow measure with respect to initial and target measures μ, ν, introduce an infinite family of lifted supermartingale couplings that arise via shadow measure, and show how to explicitly determine the *martingale points* of each such coupling.

Keywords: Brenier's theorem; convex-decreasing order; optimal transport; peacocks; stability; supermartingales

References

[1] Bayraktar, E., Deng, S. and Norgilas, D. (2021). A potential-based construction of the increasing supermartingale coupling. Preprint (to appear in Ann. Appl. Probab.), available online at arXiv:2108.03450.
[2] Bayraktar, E., Deng, S. and Norgilas, D. (2022). Supermartingale Brenier's Theorem with full-marginals constraint. Preprint (to appear in Frontiers of Mathematical Finance), available online at arXiv:2212.14174.
[3] Beiglböck, M., Cox, A.M.G. and Huesmann, M. (2017). Optimal transport and Skorokhod embedding. *Invent. Math.* **208** 327–400. MR3639595 https://doi.org/10.1007/s00222-016-0692-2
[4] Beiglböck, M., Cox, A.M.G. and Huesmann, M. (2020). The geometry of multi-marginal Skorokhod embedding. *Probab. Theory Related Fields* **176** 1045–1096. MR4087489 https://doi.org/10.1007/s00440-019-00935-z
[5] Beiglböck, M., Henry-Labordère, P. and Penkner, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985 https://doi.org/10.1007/s00780-013-0205-8
[6] Beiglböck, M., Henry-Labordère, P. and Touzi, N. (2017). Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.* **127** 3005–3013. MR3682121 https://doi.org/10.1016/j.spa.2017.01.004
[7] Beiglböck, M., Hobson, D. and Norgilas, D. (2022). The potential of the shadow measure. *Electron. Commun. Probab.* **27** Paper No. 16, 12. MR4389158 https://doi.org/10.1214/22-ecp457
[8] Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332 https://doi.org/10.1214/14-AOP966
[9] Beiglböck, M. and Juillet, N. (2021). Shadow couplings. *Trans. Amer. Math. Soc.* **374** 4973–5002. MR4273182 https://doi.org/10.1090/tran/8380

1350-7265

[10] Beiglböck, M. and Nutz, M. (2014). Martingale inequalities and deterministic counterparts. *Electron. J. Probab.* **19** no. 95, 15. MR3272328 https://doi.org/10.1214/EJP.v19-3270
[11] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[12] Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.* **305** 805–808. MR0923203
[13] Brückerhoff, M., Huesmann, M. and Juillet, N. (2022). Shadow martingales—a stochastic mass transport approach to the peacock problem. *Electron. J. Probab.* **27** Paper No. 127, 62. MR4490406 https://doi.org/10.1214/22-ejp846
[14] Campi, L., Laachir, I. and Martini, C. (2017). Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.* **21** 471–486. MR3626622 https://doi.org/10.1007/s00780-016-0322-2
[15] Chacon, R.V. (1977). Potential processes. *Trans. Amer. Math. Soc.* **226** 39–58. MR0501374 https://doi.org/10.2307/1997940
[16] Chacon, R.V. and Walsh, J.B. (1976). One-dimensional potential embedding. In *Séminaire de Probabilités, X*. *Lecture Notes in Math.* **511** 19–23. Berlin: Springer. MR0445598
[17] Föllmer, H. and Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time*, 4th ed. *De Gruyter Graduate*. Berlin: de Gruyter. MR3859905 https://doi.org/10.1515/9783110463453
[18] Galichon, A., Henry-Labordère, P. and Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** 312–336. MR3161649 https://doi.org/10.1214/13-AAP925
[19] Henry-Labordère, P., Obłój, J., Spoida, P. and Touzi, N. (2016). The maximum maximum of a martingale with given *n* marginals. *Ann. Appl. Probab.* **26** 1–44. MR3449312 https://doi.org/10.1214/14-AAP1084
[20] Henry-Labordère, P., Tan, X. and Touzi, N. (2016). An explicit martingale version of the one-dimensional Brenier's theorem with full marginals constraint. *Stochastic Process. Appl.* **126** 2800–2834. MR3522302 https://doi.org/10.1016/j.spa.2016.03.003
[21] Henry-Labordère, P. and Touzi, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.* **20** 635–668. MR3519164 https://doi.org/10.1007/s00780-016-0299-x
[22] Hirsch, F. and Roynette, B. (2012). A new proof of Kellerer's theorem. *ESAIM Probab. Stat.* **16** 48–60. MR2911021 https://doi.org/10.1051/ps/2011164
[23] Hobson, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*. *Lecture Notes in Math.* **2003** 267–318. Berlin: Springer. MR2762363 https://doi.org/10.1007/978-3-642-14660-2_4
[24] Hobson, D. and Klimmek, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.* **19** 189–214. MR3292129 https://doi.org/10.1007/s00780-014-0249-4
[25] Hobson, D. and Neuberger, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879 https://doi.org/10.1111/j.1467-9965.2010.00473.x
[26] Hobson, D. and Norgilas, D. (2019). Robust bounds for the American put. *Finance Stoch.* **23** 359–395. MR3933425 https://doi.org/10.1007/s00780-019-00385-4
[27] Hobson, D. and Norgilas, D. (2022). A construction of the left-curtain coupling. *Electron. J. Probab.* **27** Paper No. 147, 46. MR4512390 https://doi.org/10.1214/22-ejp868
[28] Hobson, D.G. (1998). The maximum maximum of a martingale. In *Séminaire de Probabilités, XXXII*. *Lecture Notes in Math.* **1686** 250–263. Berlin: Springer. MR1655298 https://doi.org/10.1007/BFb0101762
[29] Hobson, D.G. and Norgilas, D. (2019). The left-curtain martingale coupling in the presence of atoms. *Ann. Appl. Probab.* **29** 1904–1928. MR3914560 https://doi.org/10.1214/18-AAP1450
[30] Juillet, N. (2016). Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1823–1843. MR3573297 https://doi.org/10.1214/15-AIHP700
[31] Juillet, N. (2018). Martingales associated to peacocks using the curtain coupling. *Electron. J. Probab.* **23** Paper No. 8, 29. MR3771745 https://doi.org/10.1214/18-EJP138
[32] Källblad, S., Tan, X. and Touzi, N. (2017). Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks. *Ann. Appl. Probab.* **27** 686–719. MR3655851 https://doi.org/10.1214/16-AAP1191
[33] Kellerer, H.G. (1973). Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ. Prague, Prague, 1971; Dedicated to the Memory of Antonín Špaček)* 341–374. Prague: Academia. MR0356249

[34] Nutz, M. and Stebegg, F. (2018). Canonical supermartingale couplings. *Ann. Probab.* **46** 3351–3398. MR3857858 https://doi.org/10.1214/17-AOP1249

[35] Nutz, M., Stebegg, F. and Tan, X. (2020). Multiperiod martingale transport. *Stochastic Process. Appl.* **130** 1568–1615. MR4058283 https://doi.org/10.1016/j.spa.2019.05.010

[36] Obłój, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–390. MR2068476 https://doi.org/10.1214/154957804100000060

[37] Obłój, J., Spoida, P. and Touzi, N. (2015). Martingale inequalities for the maximum via pathwise arguments. In *Memoriam Marc Yor—Séminaire de Probabilités XLVII*. *Lecture Notes in Math.* **2137** 227–247. Cham: Springer. MR3444301 https://doi.org/10.1007/978-3-319-18585-9_11

[38] Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton, N.J.: Princeton Univ. Press. MR0274683

[39] Rüschendorf, L. and Rachev, S.T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* **32** 48–54. MR1035606 https://doi.org/10.1016/0047-259X(90)90070-X

[40] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423–439. MR0177430 https://doi.org/10.1214/aoms/1177700153

[41] Villani, C. (2009). *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Berlin: Springer. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

Bernoulli **30**(1), 2024, 170–197
https://doi.org/10.3150/23-BEJ1593

Two-sample contamination model test

XAVIER MILHAUD[1,a], DENYS POMMERET[2,b], YAHIA SALHI[3,c] and PIERRE VANDEKERKHOVE[4,d]

[1]*Aix-Marseille University, Centrale Marseille, 13013 Marseille, France,* [a]*xavier.milhaud@univ-amu.fr*
[2]*Aix-Marseille University, Campus de Luminy, 13288 Marseille cedex 9, France,* [b]*denys.pommeret@univ-amu.fr*
[3]*Univ Lyon, UCBL, ISFA LSAF EA2429, F-69007, Lyon, France,* [c]*yahia.salhi@univ-lyon1.fr*
[4]*Université Gustave Eiffel, LAMA (UMR 8050), 77420 Champs-sur-Marne, France,* [d]*pierre.vandekerkhove@univ-eiffel.fr*

In this paper, we consider two-component mixture models having one single known component. This type of model is of particular interest when a known random phenomenon is contaminated by an unknown random effect. We propose in this setup to test the equality in distribution of the unknown random sources involved in two separate samples generated from such a model. For this purpose, we introduce the so-called IBM (Inversion-Best Matching) approach resulting in a tuning-free relaxed semiparametric Cramér-von Mises type two-sample test requiring minimal assumptions about the unknown distributions. The accomplishment of our work lies in the fact that we establish, under some natural and interpretable mutual-identifiability conditions specific to the two-sample case, a functional central limit theorem about the proportion parameters along with the unknown cumulative distribution functions of the model. An intensive numerical study is carried out from a large range of simulation setups to illustrate the asymptotic properties of our test. Finally, our testing procedure, implemented in the `admixRpackage`, is applied to a real-life situation through pairwise post COVID-19 mortality excess profile testing across a panel of European countries.

Keywords: Cramér-von Mises; finite mixture model; mortality excess; semiparametric estimation

References

Beaney, T., Clarke, J.M., Jain, V., Golestaneh, A.K., Lyons, G., Salman, D. and Majeed, A. (2020). Excess mortality: The gold standard in measuring the impact of Covid-19 worldwide? *J. R. Soc. Med.* **113** 329–334. https://doi.org/10.1177/0141076820956802

Bordes, L., Delmas, C. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scand. J. Stat.* **33** 733–752. MR2300913 https://doi.org/10.1111/j.1467-9469.2006.00515.x

Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356 https://doi.org/10.1214/009053606000000353

Bordes, L. and Vandekerkhove, P. (2010). Semiparametric two-component mixture model with a known component: An asymptotically normal estimator. *Math. Methods Statist.* **19** 22–41. MR2682853 https://doi.org/10.3103/S1066530710010023

Broët, P., Lewin, A., Richardson, S., Dalmasso, C. and Magdelenat, H. (2004). A mixture model-based strategy for selecting sets of genes in multiclass response microarray experiments. *Bioinformatics* **20** 2562–2571. https://doi.org/10.1093/bioinformatics/bth285

Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.* **41** 227–239. MR3181141 https://doi.org/10.1111/sjos.12015

Cai, T.T. and Jin, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100–145. MR2589318 https://doi.org/10.1214/09-AOS696

Celisse, A. and Robin, S. (2010). A cross-validation based estimation of the proportion of true null hypotheses. *J. Statist. Plann. Inference* **140** 3132–3147. MR2659843 https://doi.org/10.1016/j.jspi.2010.04.014

Chakraborty, R. and Weiss, K.M. (1988). Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci.* **85** 9119–9123.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetics* **23** 70–86.

Genest, C., Nešlehová, J.G., Rémillard, B. and Murphy, O.A. (2019). Testing for independence in arbitrary distributions. *Biometrika* **106** 47–68. MR3912383 https://doi.org/10.1093/biomet/asy059

Gribkova, S. and Lopez, O. (2015). Non-parametric copula estimation under bivariate censoring. *Scand. J. Stat.* **42** 925–946. MR3426303 https://doi.org/10.1111/sjos.12144

Higham, D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43** 525–546. MR1872387 https://doi.org/10.1137/S0036144500378302

Klingenberg, C., Pirner, M. and Puppo, G. (2017). A consistent kinetic model for a two-component mixture with an application to plasma. *Kinet. Relat. Models* **10** 445–465. MR3579578 https://doi.org/10.3934/krm.2017017

Kontis, V., Bennett, J.E., Rashid, T., Parks, R.M., Pearson-Stuttard, J., Guillot, M., Asaria, P., Zhou, B., Battaglini, M., Corsetti, G. et al. (2020). Magnitude, demographics and dynamics of the effect of the first wave of the Covid-19 pandemic on all-cause mortality in 21 industrialized countries. *Nat. Med.* 1–10.

Lai, Y., Adam, B.-L., Podolsky, R. and She, J.-X. (2007). A mixture model approach to the tests of concordance and discordance between two large-scale experiments with two-sample groups. *Bioinformatics* **23** 1243–1250.

Lai, Y., Zhang, F., Nayak, T.K., Modarres, R., Lee, N.H. and McCaffrey, T.A. (2017). An efficient concordant integrative analysis of multiple large-scale two-sample expression data sets. *Bioinformatics* **33** 3852–3860.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D. and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193** 1233–1254. https://doi.org/10.1534/genetics.112.147330

Mannucci, E., Nreu, B. and Monami, M. (2020). Factors associated with increased all-cause mortality during the Covid-19 pandemic in Italy. *Int. J. Infect. Dis.* **98** 121–124. https://doi.org/10.1016/j.ijid.2020.06.077

McLachlan, G.J., Bean, R. and Jones, L.B.-T. (2006). A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics* **22** 1608–1615.

Milhaud, X., Pommeret, D., Salhi, Y. and Vandekerkhove, P. (2022). Semiparametric two-sample admixture components comparison test: The symmetric case. *J. Statist. Plann. Inference* **216** 135–150. MR4278139 https://doi.org/10.1016/j.jspi.2021.05.010

Milhaud, X., Pommeret, D., Salhi, Y. and Vandekerkhove, P. (2024). Supplement to “Two-sample contamination model test.” https://doi.org/10.3150/23-BEJ1593SUPP

Nguyen, V.H. and Matias, C. (2014). On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. *Scand. J. Stat.* **41** 1167–1194. MR3277044 https://doi.org/10.1111/sjos.12091

Patra, R.K. and Sen, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 869–893. MR3534354 https://doi.org/10.1111/rssb.12148

Podlaski, R. and Roesch, F.A. (2014). Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species: A two-component mixture model approach. *Math. Biosci.* **249** 60–74. MR3173073 https://doi.org/10.1016/j.mbs.2014.01.007

Pommeret, D. and Vandekerkhove, P. (2019). Semiparametric density testing in the contamination model. *Electron. J. Stat.* **13** 4743–4793. MR4038725 https://doi.org/10.1214/19-EJS1650

Suesse, T., Rayner, J.C.W. and Thas, O. (2017). Assessing the fit of finite mixture distributions. *Aust. N. Z. J. Stat.* **59** 463–483. MR3760154 https://doi.org/10.1111/anzs.12213

Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Stat.* **34** 1265–1269. MR0155376 https://doi.org/10.1214/aoms/1177703862

van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Walker, M.G., Mateo, M., Olszewski, E.W., Sen, B. and Woodroofe, M. (2009). Clean kinematic samples in dwarf spheroidals: An algorithm for evaluating membership and estimating distribution parameters when contamination is present. *Astron. J.* **137** 3109.

Xiang, S., Yao, W. and Yang, G. (2019). An overview of semiparametric extensions of finite mixture models. *Statist. Sci.* **34** 391–404. MR4017520 https://doi.org/10.1214/19-STS698

Bernoulli **30**(1), 2024, 198–226
https://doi.org/10.3150/23-BEJ1594

Concentration of measure bounds for matrix-variate data with missing values

SHUHENG ZHOU[a]

Department of Statistics, University of California, Riverside, CA, USA, [a]*szhou@ucr.edu*

We consider the following data perturbation model, where the covariates incur multiplicative errors. For two $n \times m$ random matrices U, X, we denote by $U \circ X$ the Hadamard or Schur product, which is defined as $(U \circ X)_{ij} = (U_{ij}) \cdot (X_{ij})$. In this paper, we study the subgaussian matrix variate model, where we observe the matrix variate data X through a random mask U:

$$\mathcal{X} = U \circ X \quad \text{where} \quad X = B^{1/2} \mathbb{Z} A^{1/2},$$

where $\mathbb{Z}$ is a random matrix with independent subgaussian entries, and U is a mask matrix with either zero or positive entries, where $\mathbb{E}U_{ij} \in [0,1]$ and all entries are mutually independent. Under the assumption of independence between U and X, we introduce componentwise unbiased estimators for estimating covariance A and B, and prove the concentration of measure bounds in the sense of guaranteeing the restricted eigenvalue (RE) conditions to hold on the unbiased estimator for B, when columns of data matrix X are sampled with different rates. We further develop multiple regression methods for estimating the inverse of B and show statistical rate of convergence. Our results provide insight for sparse recovery for relationships among entities (samples, locations, items) when features (variables, time points, user ratings) are present in the observed data matrix $\mathcal{X}$ with heterogeneous rates. Our proof techniques can certainly be extended to other scenarios. We provide simulation evidence illuminating the theoretical predictions.

Keywords: Concentration of measure; covariance estimation; inverse covariance estimation; matrix variate data; missing values; multiple regression; restricted Eigenvalue conditions; space-time model; sparse Hanson-Wright inequality; sparse quadratic forms; subgaussian concentration; subsampling

References

[1] Agarwal, A., Negahban, S. and Wainwright, M.J. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *Ann. Statist.* **40** 2452–2482. MR3097609 https://doi.org/10.1214/12-AOS1032
[2] Allen, G.I. and Tibshirani, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.* **4** 764–790. MR2758420 https://doi.org/10.1214/09-AOAS314
[3] Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243
[4] Belloni, A., Rosenbaum, M. and Tsybakov, A.B. (2017). Linear and conic programming estimators in high dimensional errors-in-variables models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 939–956. MR3641415 https://doi.org/10.1111/rssb.12196
[5] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620
[6] Brockwell, P.J. and Davis, R.A. (2006). *Time Series: Theory and Methods*. *Springer Series in Statistics*. New York: Springer. Reprint of the second (1991) edition. MR2839251
[7] Bryson, J., Vershynin, R. and Zhao, H. (2021). Marchenko-Pastur law with relaxed independence conditions. *Random Matrices Theory Appl.* **10** 2150040. MR4379545 https://doi.org/10.1142/S2010326321500404
[8] Cai, T., Cai, T.T. and Zhang, A. (2016). Structured matrix completion with applications to genomic data integration. *J. Amer. Statist. Assoc.* **111** 621–633. MR3538692 https://doi.org/10.1080/01621459.2015.1021005

[9] Candès, E.J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

[10] Candès, E.J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061

[11] Carroll, R.J., Gail, M.H. and Lubin, J.H. (1993). Case-control studies with errors in covariates. *J. Amer. Statist. Assoc.* **88** 185–199. MR1212486

[12] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. Boca Raton, FL: CRC Press/CRC. MR2243417 https://doi.org/10.1201/9781420010138

[13] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094 https://doi.org/10.1137/S1064827596304010

[14] Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2848400

[15] Dawid, A.P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68** 265–274. MR0614963 https://doi.org/10.1093/biomet/68.1.265

[16] 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

[17] Duncan, G. and Pearson, R. (1991). Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Stat. Sci.* **6** 219–232.

[18] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441. https://doi.org/10.1093/biostatistics/kxm045

[19] Fuller, W.A. (1987). *Measurement Error Models*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0898653 https://doi.org/10.1002/9780470316665

[20] Greenewald, K., Zhou, S. and Hero, A. III (2019). Tensor graphical lasso (TeraLasso). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 901–931. MR4025402 https://doi.org/10.1111/rssb.12339

[21] Gupta, A.K. and Varga, T. (1992). Characterization of matrix variate normal distributions. *J. Multivariate Anal.* **41** 80–88. MR1156682 https://doi.org/10.1016/0047-259X(92)90058-N

[22] Hanson, D.L. and Wright, F.T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864 https://doi.org/10.1214/aoms/1177693335

[23] Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *Amer. J. Polit. Sci.* **54** 561–581.

[24] Hornstein, M., Fan, R., Shedden, K. and Zhou, S. (2019). Joint mean and covariance estimation with unreplicated matrix-variate data. *J. Amer. Statist. Assoc.* **114** 682–696. MR3963172 https://doi.org/10.1080/01621459.2018.1429275

[25] Hwang, J.T. (1986). Multiplicative errors-in-variables models with applications to recent data released by the U.S. Department of Energy. *J. Amer. Statist. Assoc.* **81** 680–688. MR0860499

[26] Kalaitzis, A., Lafferty, J., Lawrence, N. and Zhou, S. (2013). The bigraphical lasso. In *Proc. 30th Int. Conf. Mach. Learn.* 1229–1237.

[27] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: Oxford University Press. MR1419991

[28] Leng, C. and Tang, C.Y. (2012). Sparse matrix graphical models. *J. Amer. Statist. Assoc.* **107** 1187–1200. MR3010905 https://doi.org/10.1080/01621459.2012.706133

[29] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR1925014 https://doi.org/10.1002/9781119013563

[30] Loh, P. and Wainwright, M. (2012). Corrupted and missing predictors: Minimax bounds for high-dimensional linear regression. In *Proc. IEEE Int. Symp. Inf. Theory* 2601–2605.

[31] Loh, P.-L. and Wainwright, M.J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

[32] Loh, P.-L. and Wainwright, M.J. (2012). Supplementary material for: High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

[33] 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
[34] Pigott, T.D. (2001). A review of methods for missing data. *Educ. Res. Eval.* **7** 353–383.
[35] Plan, Y., Vershynin, R. and Yudovina, E. (2017). High-dimensional estimation with geometric constraints. *Inf. Inference* **6** 1–40. MR3636866 https://doi.org/10.1093/imaiai/iaw015
[36] Raskutti, G., Wainwright, M.J. and Yu, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. MR2719855
[37] Ravikumar, P., Wainwright, M.J., Raskutti, G. and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766 https://doi.org/10.1214/11-EJS631
[38] 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
[39] Rosenbaum, M. and Tsybakov, A.B. (2010). Sparse recovery under matrix uncertainty. *Ann. Statist.* **38** 2620–2651. MR2722451 https://doi.org/10.1214/10-AOS793
[40] Rosenbaum, M. and Tsybakov, A.B. (2013). Improved matrix uncertainty selector. In *From Probability to Statistics and Back: High-Dimensional Models and Processes*. *Inst. Math. Stat. (IMS) Collect.* **9** 276–290. Beachwood, OH: IMS. MR3202640 https://doi.org/10.1214/12-IMSCOLL920
[41] Rothman, A.J., Bickel, P.J., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. MR2417391 https://doi.org/10.1214/08-EJS176
[42] Rudelson, M. (2013). Private Communication.
[43] Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** 82. MR3125258 https://doi.org/10.1214/ECP.v18-2865
[44] Rudelson, M. and Zhou, S. (2013). Reconstruction from anisotropic random measurements. *IEEE Trans. Inf. Theory* **59** 3434–3447. MR3061256 https://doi.org/10.1109/TIT.2013.2243201
[45] Rudelson, M. and Zhou, S. (2017). Errors-in-variables models with dependent measurements. *Electron. J. Stat.* **11** 1699–1797. MR3639561 https://doi.org/10.1214/17-EJS1234
[46] Sigrist, F., Künsch, H.R. and Stahel, W.A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 3–33. MR3299397 https://doi.org/10.1111/rssb.12061
[47] Smith, R., Kolenikov, S. and Cox, L. (2003). Spatiotemporal modeling of $PM_{2.5}$ data with missing values. *J. Geophys. Res.* **108**.
[48] Städler, N. and Bühlmann, P. (2012). Missing values: Sparse inverse covariance estimation and an extension to sparse regression. *Stat. Comput.* **22** 219–235. MR2865066 https://doi.org/10.1007/s11222-010-9219-7
[49] Städler, N., Stekhoven, D.J. and Bühlmann, P. (2014). Pattern alternating maximization algorithm for missing data in high-dimensional problems. *J. Mach. Learn. Res.* **15** 1903–1928. MR3231598
[50] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
[51] Tsiligkaridis, T., Hero, A.O. III and Zhou, S. (2013). On convergence of Kronecker graphical lasso algorithms. *IEEE Trans. Signal Process.* **61** 1743–1755. MR3038387 https://doi.org/10.1109/TSP.2013.2240157
[52] 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
[53] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge: Cambridge Univ. Press. MR3837109 https://doi.org/10.1017/9781108231596
[54] Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge: Cambridge Univ. Press. MR3967104 https://doi.org/10.1017/9781108627771
[55] Wright, F.T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** 1068–1070. MR0353419 https://doi.org/10.1214/aop/1176996815
[56] Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. MR2719856

[57] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018
[58] Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.* **42** 532–562. MR3210978 https://doi.org/10.1214/13-AOS1187
[59] Zhou, S. (2019). Sparse Hanson-Wright inequalities for subgaussian quadratic forms. *Bernoulli* **25** 1603–1639. MR3961224 https://doi.org/10.3150/17-BEJ978
[60] Zhou, S. (2024). Supplement to "Concentration of measure bounds for matrix-variate data with missing values." https://doi.org/10.3150/23-BEJ1594SUPP
[61] Zhou, S., Lafferty, J. and Wasserman, L. (2009). Compressed and privacy-sensitive sparse regression. *IEEE Trans. Inf. Theory* **55** 846–866. MR2597272 https://doi.org/10.1109/TIT.2008.2009605
[62] Zhou, S., Lafferty, J. and Wasserman, L. (2010). Time varying undirected graphs. *Mach. Learn.* **80** 295–319. MR3108169 https://doi.org/10.1007/s10994-010-5180-0
[63] Zhou, S., Rütimann, P., Xu, M. and Bühlmann, P. (2011). High-dimensional covariance estimation based on Gaussian graphical models. *J. Mach. Learn. Res.* **12** 2975–3026. MR2854354

Bernoulli **30**(1), 2024, 227–251
https://doi.org/10.3150/23-BEJ1595

On bivariate distributions of the local time of Itô-McKean diffusions

JACEK JAKUBOWSKI[a] and MACIEJ WIŚNIEWOLSKI[b]

Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland, [a]*jakub@mimuw.edu.pl,* [b]*wisniewolski@mimuw.edu.pl*

Denote as L the local time at 0 of an Itô-McKean diffusion X. We present a new explicit description of the distribution of L_t in terms of convolution exponent and, using the excursion theory, we describe the transition density of the pair (X, L). We provide a simple connection formula for the distribution of excursions of a bivariate Itô-McKean diffusion from a hyperplane. Examples involving the distribution of a local time are presented, including a formula for the distribution of (X_t, L_∞) for a transient diffusion.

Keywords: Convolution algebra; excursion theory; Itô-McKean diffusion; local time

References

Blumenthal, R.M. (1992). *Excursions of Markov Processes*. *Probability and Its Applications*. Boston, MA: Birkhäuser, Inc. MR1138461 https://doi.org/10.1007/978-1-4684-9412-9

Borodin, A.N. and Salminen, P. (2002). *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. *Probability and Its Applications*. Basel: Birkhäuser. MR1912205 https://doi.org/10.1007/978-3-0348-8163-0

Burdzy, K. (1986). Brownian excursions from hyperplanes and smooth surfaces. *Trans. Amer. Math. Soc.* **295** 35–57. MR0831187 https://doi.org/10.2307/2000144

Çetin, U. (2018). Path transformations for local times of one-dimensional diffusions. *Stochastic Process. Appl.* **128** 3439–3465. MR3849815 https://doi.org/10.1016/j.spa.2017.11.005

Eqworld – The World of Mathematical equations http://eqworld.ipmnet.ru/index.htm.

Fitzsimmons, P.J. and Pitman, J. (1999). Kac's moment formula and the Feynman-Kac formula for additive functionals of a Markov process. *Stochastic Process. Appl.* **79** 117–134. MR1670526 https://doi.org/10.1016/S0304-4149(98)00081-7

Forman, J.L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scand. J. Statist.* **35** 438–465. MR2446729 https://doi.org/10.1111/j.1467-9469.2007.00592.x

Getoor, R.K. (1979). Excursions of a Markov process. *Ann. Probab.* **7** 244–266. MR0525052

Getoor, R.K. and Sharpe, M.J. (1979). Excursions of Brownian motion and Bessel processes. *Z. Wahrsch. Verw. Gebiete* **47** 83–106. MR0521534 https://doi.org/10.1007/BF00533253

Getoor, R.K. and Sharpe, M.J. (1982). Excursions of dual processes. *Adv. in Math.* **45** 259–309. MR0673804 https://doi.org/10.1016/S0001-8708(82)80006-6

Itô, K. (1972). Poisson point processes attached to Markov processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability Theory* 225–239. Berkeley, Calif.: Univ. California Press. MR0402949

Itô, K. and McKean, H.P. Jr. (1974). *Diffusion Processes and Their Sample Paths*. *Die Grundlehren der Mathematischen Wissenschaften, Band 125*. Berlin: Springer. MR0345224

Jakubowski, J. and Wiśniewolski, M. (2020). Volterra integral equations of the first kind and applications to linear diffusions. *Trans. Amer. Math. Soc.* **373** 7455–7472. MR4155213 https://doi.org/10.1090/tran/8169

Jakubowski, J. and Wiśniewolski, M. (2021). A convolution formula for the local time of an Itô diffusion reflecting at 0 and a generalized Stroock-Williams equation. *Bernoulli* **27** 1870–1898. MR4260510 https://doi.org/10.3150/20-bej1295

1350-7265

Jakubowski, J. and Wiśniewolski, M. (2024). Supplement to "On bivariate distributions of the local time of Itô-McKean diffusions." https://doi.org/10.3150/23-BEJ1595SUPP

Kasahara, Y. and Watanabe, S. (2006). Brownian representation of a class of Lévy processes and its application to occupation times of diffusion processes. *Illinois J. Math.* **50** 515–539. MR2247838

Kent, J.T. (1980). Eigenvalue expansions for diffusion hitting times. *Z. Wahrsch. Verw. Gebiete* **52** 309–319. MR0576891 https://doi.org/10.1007/BF00538895

Kesten, H. (1969). *Hitting Probabilities of Single Points for Processes with Stationary Independent Increments*. *Memoirs of the American Mathematical Society, No. 93*. Providence, R.I.: Amer. Math. Soc. MR0272059

Knight, F.B. (1981). Characterization of the Levy measures of inverse local times of gap diffusion. In *Seminar on Stochastic Processes, 1981 (Evanston, Ill., 1981)*. *Progr. Prob. Statist.* **1** 53–78. Boston, Mass.: Birkhäuser. MR0647781

Lew, J.S. (1972). On linear Volterra integral equations of convolution type. *Proc. Amer. Math. Soc.* **35** 450–456. MR0308699 https://doi.org/10.2307/2037627

Maisonneuve, B. (1975). Exit systems. *Ann. Probab.* **3** 399–411. MR0400417 https://doi.org/10.1214/aop/1176996348

Mansuy, R. and Yor, M. (2008). *Aspects of Brownian Motion*. *Universitext*. Berlin: Springer. MR2454984 https://doi.org/10.1007/978-3-540-49966-4

Naĭmark, M.A. (1959). *Normed Rings*. Groningen: P. Noordhoff N. V. Translated from the first Russian edition by Leo F. Boron. MR0110956

NIST (2016). *Digital Library of Mathematical Functions*, http://dlmf.nist.gov/, Release 1.0.11 of 2016-06-08.

Pang, H. and Stroock, D.W. (2007). A peculiar two point boundary value problem. *Ann. Probab.* **35** 1623–1641. MR2349570 https://doi.org/10.1214/009117906000000818

Peskir, G. (2014). A probabilistic solution to the Stroock-Williams equation. *Ann. Probab.* **42** 2197–2206. MR3262501 https://doi.org/10.1214/13-AOP865

Pitman, J. and Yor, M. (1999). Laplace transforms related to excursions of a one-dimensional diffusion. *Bernoulli* **5** 249–255. MR1681697 https://doi.org/10.2307/3318434

Pitman, J. and Yor, M. (2003). Hitting, occupation and inverse local times of one-dimensional diffusions: Martingale and excursion approaches. *Bernoulli* **9** 1–24. MR1963670 https://doi.org/10.3150/bj/1068129008

Revuz, D. and Yor, M. (2005). *Continuous Martingales and Brownian Motion*, 3rd ed. Berlin: Springer.

Rogers, L.C.G. (1989). A guided tour through excursions. *Bull. Lond. Math. Soc.* **21** 305–341. MR0998631 https://doi.org/10.1112/blms/21.4.305

Rogers, L.C.G. and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales. Vol. 2: Itô Calculus*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0921238

Rudin, W. (1991). *Functional Analysis*, 2nd ed. *International Series in Pure and Applied Mathematics*. New York: McGraw-Hill, Inc. MR1157815

Salminen, P., Vallois, P. and Yor, M. (2007). On the excursion theory for linear diffusions. *Jpn. J. Math.* **2** 97–127. MR2295612 https://doi.org/10.1007/s11537-007-0662-y

Stroock, D.W. and Williams, D. (2005). A simple PDE and Wiener-Hopf Riccati equations. *Comm. Pure Appl. Math.* **58** 1116–1148. MR2143528 https://doi.org/10.1002/cpa.20081

Stroock, D.W. and Williams, D. (2006). Further study of a simple PDE. *Illinois J. Math.* **50** 961–989. MR2247850

Titchmarsh, E.C. (1926). The Zeros of Certain Integral Functions. *Proc. Lond. Math. Soc.* **25** 283–302. MR1575285 https://doi.org/10.1112/plms/s2-25.1.283

Walsh, J.B. (1978). Excursions and local time. *Astérisque* **52–53** 159–192.

Williams, D. (1974). Path decomposition and continuity of local time for one-dimensional diffusions. I. *Proc. Lond. Math. Soc. (3)* **28** 738–768. MR0350881 https://doi.org/10.1112/plms/s3-28.4.738

Williams, D. and Andrews, S. (2006). Indefinite inner products: A simple illustrative example. *Math. Proc. Cambridge Philos. Soc.* **141** 127–159. MR2238647 https://doi.org/10.1017/S0305004106009364

Bernoulli **30**(1), 2024, 252–277
https://doi.org/10.3150/23-BEJ1596

The infinite Viterbi alignment and decay-convexity

NICK WHITELEY[1,3,a], MATT W. JONES[2,b] and ALEKS P.F. DOMANSKI[2,3,c]

[1]*School of Mathematics, University of Bristol, Bristol, UK,* [a]*nick.whiteley@bristol.ac.uk*
[2]*School of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, UK,* [b]*matt.jones@bristol.ac.uk,* [c]*aleks.domanski@bristol.ac.uk*
[3]*Alan Turing Institute, London, UK*

The infinite Viterbi alignment is the limiting maximum a-posteriori estimate of the unobserved path in a hidden Markov model as the length of the time horizon grows. For models on state-space $\mathbb{R}^d$ satisfying a new "decay-convexity" condition, we develop an approach to existence of the infinite Viterbi alignment in an infinite dimensional Hilbert space. Quantitative bounds on the distance to the Viterbi process, which are the first of their kind, are derived and used to illustrate how approximate estimation via parallelization can be accurate and scaleable to high-dimensional problems because the rate of convergence to the infinite Viterbi alignment does not necessarily depend on d. The results are applied to approximate estimation via parallelization and a model of neural population activity.

Keywords: Convex optimization; hidden Markov models; MAP estimation; parallelization

References

[1] Andrews, B. and Hopper, C. (2011). *The Ricci Flow in Riemannian Geometry: A Complete Proof of the Differentiable 1/4-Pinching Sphere Theorem*. *Lecture Notes in Math.* **2011**. Heidelberg: Springer. MR2760593 https://doi.org/10.1007/978-3-642-16286-2
[2] Caliebe, A. (2006). Properties of the maximum a posteriori path estimator in hidden Markov models. *IEEE Trans. Inf. Theory* **52** 41–51. MR2237334 https://doi.org/10.1109/TIT.2005.860425
[3] Caliebe, A. and Rösler, U. (2002). Convergence of the maximum a posteriori path estimator in hidden Markov models. *IEEE Trans. Inf. Theory* **48** 1750–1758. MR1929991 https://doi.org/10.1109/TIT.2002.1013123
[4] Chigansky, P. and Ritov, Y. (2011). On the Viterbi process with continuous state space. *Bernoulli* **17** 609–627. MR2787607 https://doi.org/10.3150/10-BEJ294
[5] Deimling, K. (1977). *Ordinary Differential Equations in Banach Spaces*. *Lecture Notes in Mathematics* **596**. Berlin: Springer. MR0463601
[6] Donner, C., Obermayer, K. and Shimazaki, H. (2017). Approximate inference for time-varying interactions and macroscopic dynamics of neural populations. *PLoS Comput. Biol.* **13** 1–27.
[7] Douc, R., Fort, G. , Moulines, E. and Priouret, P. (2009). Forgetting the initial distribution for hidden Markov models. *Stochastic Process. Appl.* **119** 1235–1256.
[8] Douc, R. and Moulines, E. (2012). Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. *Ann. Statist.* **40** 2697–2732.
[9] Douc, R., Moulines, E. and Stoffer, D.S. (2014). *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*. *Chapman & Hall/CRC Texts in Statistical Science Series*. Boca Raton, FL: CRC Press/CRC. MR3289095
[10] Durbin, J. and Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*, 2nd ed. *Oxford Statistical Science Series* **38**. Oxford: Oxford Univ. Press. MR3014996 https://doi.org/10.1093/acprof:oso/9780199641178.001.0001
[11] Jun, J.J. et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551** 232.

1350-7265

[12] Koloydenko, A. and Lember, J. (2008). Infinite Viterbi alignments in the two state hidden Markov models. *Acta Comment. Univ. Tartu. Math.* **12** 109–124. MR2488711

[13] Kuljus, K. and Lember, J. (2012). Asymptotic risks of Viterbi segmentation. *Stochastic Process. Appl.* **122** 3312–3341. MR2946444 https://doi.org/10.1016/j.spa.2012.05.014

[14] Lember, J. and Koloydenko, A. (2008). The adjusted Viterbi training for hidden Markov models. *Bernoulli* **14** 180–206. MR2401659 https://doi.org/10.3150/07-BEJ105

[15] Lember, J. and Koloydenko, A.A. (2010). A constructive proof of the existence of Viterbi processes. *IEEE Trans. Inf. Theory* **56** 2017–2033. MR2654491 https://doi.org/10.1109/TIT.2010.2040897

[16] Lember, J. and Sova, J. (2020). Existence of infinite Viterbi path for pairwise Markov models. *Stochastic Process. Appl.* **130** 1388–1425. MR4058277 https://doi.org/10.1016/j.spa.2019.05.004

[17] Lember, J. and Sova, J. (2021). Regenerativity of Viterbi process for pairwise Markov models. *J. Theoret. Probab.* **34** 1–33. MR4207478 https://doi.org/10.1007/s10959-020-01022-z

[18] Saumard, A. and Wellner, J.A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

[19] Shimazaki, H., Amari, S., Brown, E.N. and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput. Biol.* **8** e1002385, 27. MR2921292 https://doi.org/10.1371/journal.pcbi.1002385

[20] West, M. and Harrison, J. (2006). *Bayesian Forecasting and Dynamic Models*. New York: Springer.

Bernoulli **30**(1), 2024, 278–303
https://doi.org/10.3150/23-BEJ1597

Central limit theorem and near classical Berry-Esseen rate for self normalized sums in high dimensions

DEBRAJ DAS[a]

Department of Mathematics, Indian Institute of Technology, Bombay, Maharashtra 400076, India,
[a]*debrajdas@math.iitb.ac.in*

In this article, we are interested in the normal approximation of

$$T_n = \Big(\sum_{i=1}^{n} X_{i1} / \Big(\sqrt{\sum_{i=1}^{n} X_{i1}^2} \Big), \dots, \sum_{i=1}^{n} X_{ip} / \Big(\sqrt{\sum_{i=1}^{n} X_{ip}^2} \Big) \Big)$$

in $\mathcal{R}^p$ uniformly over the class of hyper-rectangles $\mathcal{A}^{re} = \big\{ \prod_{j=1}^{p} [a_j, b_j] \cap \mathcal{R} : -\infty \le a_j \le b_j \le \infty, j = 1, \dots, p \big\}$, where $X_1, \dots, X_n$ are non-degenerate independent p-dimensional random vectors. We assume that the components of X_i are independent and identically distributed (iid) and investigate the optimal cut-off rate of $\log p$ in the uniform central limit theorem (UCLT) for T_n over $\mathcal{A}^{re}$. The aim is to reduce the exponential moment conditions, generally assumed for exponential growth of the dimension with respect to the sample size in high dimensional CLT, to some polynomial moment conditions. Indeed, we establish that only the existence of some polynomial moment of order $\in [2, 4]$ is sufficient for exponential growth of p. However the rate of growth of $\log p$ cannot further be improved from $o\big(n^{1/2}\big)$ as a power of n even if X_{ij}'s are iid across (i, j) and X_{11} is bounded. We also establish near$-n^{-\kappa/2}$ Berry-Esseen rate for T_n in high dimension under the existence of $(2+\kappa)$th absolute moments of X_{ij} for $0 < \kappa \le 1$. When $\kappa = 1$, the obtained Berry-Esseen rate is also shown to be optimal. As an application, we find respective versions for componentwise Student's t-statistic, which may be useful in high dimensional statistical inference.

Keywords: Berry-Esseen theorem; self-normalized sum; student t-statistic; UCLT

References

[1] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *Ann. Probab.* **17** 9–25. MR0972770

[2] Bentkus, V., Bloznelis, M. and Götze, F. (1996). A Berry-Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.* **9** 765–796. MR1400598 https://doi.org/10.1007/BF02214086

[3] Bentkus, V. and Götze, F. (1996). The Berry-Esseen bound for Student's statistic. *Ann. Probab.* **24** 491–503. MR1387647 https://doi.org/10.1214/aop/1042644728

[4] Birnbaum, Z.W. (1942). An inequality for Mill's ratio. *Ann. Math. Stat.* **13** 245–246. MR0006640 https://doi.org/10.1214/aoms/1177731611

[5] 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

[6] Chernozhukov, V., Chetverikov, D. and Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. MR3565470 https://doi.org/10.1016/j.spa.2016.04.009

[7] 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

1350-7265

[8] Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions. Available at https://arxiv.org/pdf/1912.10529.pdf.

[9] Chistyakov, G.P. and Götze, F. (2004). On bounds for moderate deviations for Student's statistic. *Theory Probab. Appl.* **48** 528–535.

[10] Chistyakov, G.P. and Götze, F. (2004). Limit distributions of Studentized means. *Ann. Probab.* **32** 28–77. MR2040775 https://doi.org/10.1214/aop/1078415828

[11] Chow, Y.S. and Teicher, H. (1997). *Probability Theory-Independence, Interchangeability, Martingales*. Springer.

[12] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2003). Donsker's theorem for self-normalized partial sums processes. *Ann. Probab.* **31** 1228–1240. MR1988470 https://doi.org/10.1214/aop/1055425777

[13] Darling, D.A. (1952). The influence of the maximum term in the addition of independent random variables. *Trans. Amer. Math. Soc.* **73** 95–107. MR0048726 https://doi.org/10.2307/1990824

[14] Das, D. (2024). Supplement to "Central limit theorem and near classical Berry-Esseen rate for self normalized sums in high dimensions." https://doi.org/10.3150/23-BEJ1597SUPP

[15] Das, D. and Lahiri, S. (2021). Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** 6991–7009. MR4315595 https://doi.org/10.1090/tran/8459

[16] de la Peña, V.H., Lai, T.L. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications*. *Probability and Its Applications (New York)*. Berlin: Springer. MR2488094 https://doi.org/10.1007/978-3-540-85636-8

[17] Efron, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc.* **64** 1278–1302. MR0251826

[18] Fang, X. and Koike, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

[19] Giné, E., Götze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. MR1457629 https://doi.org/10.1214/aop/1024404523

[20] Hall, P. (1988). On the effect of random norming on the rate of convergence in the central limit theorem. *Ann. Probab.* **16** 1265–1280. MR0942767

[21] Jing, B.-Y., Shao, Q.-M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616 https://doi.org/10.1214/aop/1068646382

[22] Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7

[23] Kuchibhotla, A.K. and Rinaldo, D. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Working paper. Available at arXiv:2009.13673.

[24] Logan, B.F., Mallows, C.L., Rice, S.O. and Shepp, L.A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **1** 788–809. MR0362449 https://doi.org/10.1214/aop/1176996846

[25] Lopes, M.E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

[26] Mason, D.M. and Zhou, H.H. (2012). Quantile coupling inequalities and their applications. *Probab. Surv.* **9** 439–479. MR3007210 https://doi.org/10.1214/12-PS198

[27] Nagaev, S.V. (2002). The Berry-Esseen bound for self-normalized sums. *Siberian Adv. Math.* **12** 79–125. MR1964394

[28] Robbins, H. (1955). A remark on Stirling's formula. *Amer. Math. Monthly* **62** 26–29. MR0069328 https://doi.org/10.2307/2308012

[29] Robinson, J. and Wang, Q. (2005). On the self-normalized Cramér-type large deviation. *J. Theoret. Probab.* **18** 891–909. MR2300002 https://doi.org/10.1007/s10959-005-7531-5

[30] Sang, H. and Ge, L. (2017). Further refinement of self-normalized Cramér-type moderate deviations. *ESAIM Probab. Stat.* **21** 201–219. MR3716127 https://doi.org/10.1051/ps/2017010

[31] Shao, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328. MR1428510 https://doi.org/10.1214/aop/1024404289

[32] Shao, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750 https://doi.org/10.1023/A:1021626127372

[33] Shao, Q.-M. (2005). An explicit Berry-Esseen bound for Student's t-statistic via Stein's method. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 143–155. Singapore: Singapore Univ. Press. MR2205333 https://doi.org/10.1142/9789812567673_0009
[34] Shao, Q.-M. and Wang, Q. (2013). Self-normalized limit theorems: A survey. *Probab. Surv.* **10** 69–93. MR3161676 https://doi.org/10.1214/13-PS216
[35] Slavova, V.V. (1985). On the Berry-Esseen bound for Student's statistic. In *Stability Problems for Stochastic Models (Uzhgorod, 1984)* (V.V. Kalashnikov and V.M. Zolotarev, eds.). *Lecture Notes in Math.* **1155** 355–390. Berlin: Springer. MR0825335 https://doi.org/10.1007/BFb0074828
[36] Wang, Q. (2005). Limit theorems for self-normalized large deviation. *Electron. J. Probab.* **10** 1260–1285. MR2176384 https://doi.org/10.1214/EJP.v10-289
[37] Wang, Q. (2011). Refined self-normalized large deviations for independent random variables. *J. Theoret. Probab.* **24** 307–329. MR2795041 https://doi.org/10.1007/s10959-011-0347-6
[38] Wang, Q. and Hall, P. (2009). Relative errors in central limit theorems for Student's t statistic, with applications. *Statist. Sinica* **19** 343–354. MR2487894
[39] Wang, Q. and Jing, B.-Y. (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums. *Ann. Probab.* **27** 2068–2088. MR1742902 https://doi.org/10.1214/aop/1022677562

Bernoulli **30**(1), 2024, 304–331
https://doi.org/10.3150/23-BEJ1598

Statistical inference for function-on-function linear regression

HOLGER DETTE[a] and JIAJUN TANG[b]

Fakultät für Mathematik, Ruhr-Universität Bochum, D-44801 Bochum, Germany,
[a]*holger.dette@ruhr-uni-bochum.de,* [b]*Jiajun.Tang@ruhr-uni-bochum.de*

We propose a reproducing kernel Hilbert space approach for statistical inference regarding the slope in a function-on-function linear regression via penalised least squares, regularized by the thin-plate spline smoothness penalty. We derive a Bahadur expansion for the slope surface estimator and prove its weak convergence as a process in the space of all continuous functions. As a consequence of these results, we construct minimax optimal estimates, simultaneous confidence regions for the slope surface and simultaneous prediction bands. Moreover, we derive new tests for the hypothesis that the maximum deviation between the "true" slope surface and a given surface is less than or equal to a given threshold. In other words, we are not trying to test for exact equality (because in many applications this hypothesis is hard to justify), but rather for pre-specified deviations under the null hypothesis. To ensure practicability, non-standard bootstrap procedures are developed addressing particular features that arise in these testing problems. We also demonstrate that the new methods have good finite sample properties by means of a simulation study and illustrate their practicability by analyzing a data example.

Keywords: Bootstrap; function-on-function linear regression; maximum deviation; minimax optimality; simultaneous confidence regions; relevant hypotheses; reproducing kernel Hilbert space

References

Aitchison, J. (1964). Confidence-region tests. *J. Roy. Statist. Soc. Ser. B* **26** 462–476. MR0174119

Aue, A., Rice, G. and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 509–529. MR3798876 https://doi.org/10.1111/rssb.12257

Avery, M., Wu, Y., Zhang, H. and Zhang, J. (2014). RKHS-based functional nonparametric regression for sparse and irregular longitudinal data. *Canad. J. Statist.* **42** 204–216. MR3208336 https://doi.org/10.1002/cjs.11215

Benatia, D., Carrasco, M. and Florens, J.-P. (2017). Functional linear regression with functional response. *J. Econometrics* **201** 269–291. MR3717564 https://doi.org/10.1016/j.jeconom.2017.08.008

Berger, J.O. and Delampady, M. (1987). Testing precise hypotheses. *Statist. Sci.* **2** 317–352. MR0920141

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. *Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

Cai, T.T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.* **107** 1201–1216. MR3010906 https://doi.org/10.1080/01621459.2012.716337

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Stat.* **30** 241–255. MR1965105 https://doi.org/10.1111/1467-9469.00329

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

Cardot, H., Goia, A. and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33** 179–199. MR2044864 https://doi.org/10.1081/SAC-120028440

Cressie, N.A.C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: Wiley. MR3559472

Dette, H. and Kokot, K. (2021). Bio-equivalence tests in functional data by maximum deviation. *Biometrika* **108** 895–913. MR4341358 https://doi.org/10.1093/biomet/asaa096

Dette, H. and Kokot, K. (2022). Detecting relevant differences in the covariance operators of functional time series: A sup-norm approach. *Ann. Inst. Statist. Math.* **74** 195–231. MR4396673 https://doi.org/10.1007/s10463-021-00795-2

Dette, H., Kokot, K. and Aue, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Statist.* **48** 1168–1192. MR4102692 https://doi.org/10.1214/19-AOS1842

Dette, H., Möllenhoff, K., Volgushev, S. and Bretz, F. (2018). Equivalence of regression curves. *J. Amer. Statist. Assoc.* **113** 711–729. MR3832221 https://doi.org/10.1080/01621459.2017.1281813

Dette, H. and Tang, J. (2024). Supplement to "Statistical inference for function-on-function linear regression." https://doi.org/10.3150/23-BEJ1598SUPP

Ferraty, F. and Vieu, P. (2010). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. New York: Springer. MR2229687

Fogarty, C.B. and Small, D.S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.* **8** 2002–2026. MR3292487 https://doi.org/10.1214/14-AOAS763

Gsteiger, S., Bretz, F. and Liu, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *J. Biopharm. Statist.* **21** 708–725. MR2800311 https://doi.org/10.1080/10543406.2011.551332

Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269 https://doi.org/10.1214/009053606000000957

Hao, M., Liu, K., Xu, W. and Zhao, X. (2021). Semiparametric inference for the functional Cox model. *J. Amer. Statist. Assoc.* **116** 1319–1329. MR4309275 https://doi.org/10.1080/01621459.2019.1710155

Hilgert, N., Mas, A. and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* **41** 838–869. MR3099123 https://doi.org/10.1214/13-AOS1093

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

Imaizumi, M. and Kato, K. (2019). A simple method to construct confidence bands in functional linear regression. *Statist. Sinica* **29** 2055–2081. MR3970347

James, G.M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 411–432. MR1924298 https://doi.org/10.1111/1467-9868.00342

Kong, D., Staicu, A.-M. and Maity, A. (2016). Classical testing in functional linear models. *J. Nonparametr. Stat.* **28** 813–838. MR3555459 https://doi.org/10.1080/10485252.2016.1231806

Lei, J. (2014). Adaptive global testing for functional linear models. *J. Amer. Statist. Assoc.* **109** 624–634. MR3223738 https://doi.org/10.1080/01621459.2013.856794

Li, T. and Zhu, Z. (2020). Inference for generalized partial functional linear regression. *Statist. Sinica* **30** 1379–1397. MR4257538 https://doi.org/10.5705/ss.202018.0155

Lian, H. (2007). Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces. *Canad. J. Statist.* **35** 597–606. MR2381399 https://doi.org/10.1002/cjs.5550350410

Lian, H. (2015). Minimax prediction for functional linear regression with functional responses in reproducing kernel Hilbert spaces. *J. Multivariate Anal.* **140** 395–402. MR3372576 https://doi.org/10.1016/j.jmva.2015.06.005

Liu, W., Bretz, F., Hayter, A.J. and Wynn, H.P. (2009). Assessing nonsuperiority, noninferiority, or equivalence when comparing two regression models over a restricted covariate region. *Biometrics* **65** 1279–1287. MR2756516 https://doi.org/10.1111/j.1541-0420.2008.01192.x

Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression. *J. Amer. Statist. Assoc.* **112** 690–705. MR3671763 https://doi.org/10.1080/01621459.2016.1164053

Meister, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* **39** 1471–1495. MR2850209 https://doi.org/10.1214/10-AOS872

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159 https://doi.org/10.1214/009053604000001156

Qu, S. and Wang, X. (2017). Optimal global test for functional regression. Preprint. Available at arXiv:1710.02269.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

Ramsay, J.O., Wickham, H., Graves, S. and Hooker, G. (2014). fda: Functional data analysis. R package version 2.4.3. Available at http://CRAN.R-project.org/package=fda.

Scheipl, F. and Greven, S. (2016). Identifiability in penalized function-on-function regression models. *Electron. J. Stat.* **10** 495–526. MR3471986 https://doi.org/10.1214/16-EJS1123

Shang, Z. and Cheng, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist.* **43** 1742–1773. MR3357877 https://doi.org/10.1214/15-AOS1322

Sun, X., Du, P., Wang, X. and Ma, P. (2018). Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. *J. Amer. Statist. Assoc.* **113** 1601–1611. MR3902232 https://doi.org/10.1080/01621459.2017.1356320

Tekbudak, M.Y., Alfaro-Córdoba, M., Maity, A. and Staicu, A.-M. (2019). A comparison of testing methods in scalar-on-function regression. *AStA Adv. Stat. Anal.* **103** 411–436. MR3997720 https://doi.org/10.1007/s10182-018-00337-x

Wahba, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. Philadelphia, PA: SIAM. MR1045442 https://doi.org/10.1137/1.9781611970128

Wellner, J.A. (2003). Gaussian white noise models: Some results for monotone functions. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*. *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **43** 87–104. Beachwood, OH: IMS. MR2125049 https://doi.org/10.1214/lnms/1215092392

Wood, S.N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 95–114. MR1959095 https://doi.org/10.1111/1467-9868.00374

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

Yuan, M. and Cai, T.T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444. MR2766857 https://doi.org/10.1214/09-AOS772

Bernoulli **30**(1), 2024, 332–345
https://doi.org/10.3150/23-BEJ1599

Large deviation principles for SDEs under locally weak monotonicity conditions

JIAN WANG[1,a], HAO YANG[2,d], JIANLIANG ZHAI[1,b] and TUSHENG ZHANG[1,c]

[1]*School of Mathematical Sciences, University of Science and Technology of China, Hefei, Anhui, China,* [a]*wg1995@mail.ustc.edu.cn,* [b]*zhaijl@ustc.edu.cn,* [c]*tushengz@ustc.edu.cn*
[2]*School of Mathematics, Hefei University of Technology, Hefei, Anhui, China,* [d]*yanghao@hfut.edu.cn*

This paper establishes a Freidlin-Wentzell large deviation principle for stochastic differential equations(SDEs) under locally weak monotonicity conditions and Lyapunov conditions. We illustrate the main result of the paper by showing that it can be applied to SDEs with non-Lipschitzian coefficients, which can not be covered in the existing literature. These include the interesting biological models like stochastic Duffing-van der Pol oscillator model, stochastic SIR model, etc.

Keywords: Freidlin-Wentzell large deviation principle; locally weak monotonicity condition; non-Lipschitzian coefficients

References

[1] Anderson, R.M. and May, R.M. (1979). Population biology of infectious diseases: Part I. *Nature* **280** 361–367.
[2] Budhiraja, A., Dupuis, P. and Maroulas, V. (2008). Large deviations for infinite dimensional stochastic dynamical systems. *Ann. Probab.* **36** 1390–1420. MR2435853 https://doi.org/10.1214/07-AOP362
[3] Budhiraja, A., Dupuis, P. and Maroulas, V. (2011). Variational representations for continuous time processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 725–747. MR2841073 https://doi.org/10.1214/10-AIHP382
[4] Cerrai, S. and Röckner, M. (2004). Large deviations for stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term. *Ann. Probab.* **32** 1100–1139. MR2044675 https://doi.org/10.1214/aop/1079021473
[5] Chen, L., Dong, Z., Jiang, J., Niu, L. and Zhai, J. (2019). Decomposition formula and stationary measures for stochastic Lotka-Volterra system with applications to turbulent convection. *J. Math. Pures Appl. (9)* **125** 43–93. MR3944199 https://doi.org/10.1016/j.matpur.2019.02.013
[6] Cheng, L. and Huang, X. (2020). Large deviation principle for SDEs with Dini continuous drifts. *Stochastics* **93** 239–251.
[7] Cox, S.G., Hutzenthaler, M. and Jentze, A. (2021). Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. arXiv preprint, 90 pages, arXiv:1309.5595.
[8] Donsker, M.D. and Varadhan, S.R.S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.* **28** 1–47; ibid. 28 (1975), 279–301. MR0386024 https://doi.org/10.1002/cpa.3160280102
[9] Donsker, M.D. and Varadhan, S.R.S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461. MR0428471 https://doi.org/10.1002/cpa.3160290405
[10] Donsker, M.D. and Varadhan, S.R.S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212. MR0690656 https://doi.org/10.1002/cpa.3160360204
[11] Fang, S. and Zhang, T. (2005). A study of a class of stochastic differential equations with non-Lipschitzian coefficients. *Probab. Theory Related Fields* **132** 356–390. MR2197106 https://doi.org/10.1007/s00440-004-0398-z

[12] Freidlin, M.I. and Wentzell, A.D. (2012). *Random Perturbations of Dynamical Systems*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **260**. Heidelberg: Springer. MR2953753 https://doi.org/10.1007/978-3-642-25847-3

[13] Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge Univ. Press. MR1635735 https://doi.org/10.1017/CBO9781139173179

[14] Holmes, P. and Rand, D. (1980). Phase portraits and bifurcations of the nonlinear oscillator: $\ddot{x} + (\alpha + \gamma x^2)\dot{x} + \beta x + \delta x^3 = 0$. *Internat. J. Non-Linear Mech.* **15** 449–458. MR0605732 https://doi.org/10.1016/0020-7462(80)90031-1

[15] Lan, G. (2013). Large deviation principle of stochastic differential equations with non-Lipschitzian coefficients. *Front. Math. China* **8** 1307–1321. MR3122313 https://doi.org/10.1007/s11464-013-0318-6

[16] Lan, G. and Wu, J.-L. (2014). New sufficient conditions of existence, moment estimations and non confluence for SDEs with non-Lipschitzian coefficients. *Stochastic Process. Appl.* **124** 4030–4049. MR3264438 https://doi.org/10.1016/j.spa.2014.07.010

[17] Matoussi, A., Sabbagh, W. and Zhang, T. (2021). Large deviation principles of obstacle problems for quasilinear stochastic PDEs. *Appl. Math. Optim.* **83** 849–879. MR4239802 https://doi.org/10.1007/s00245-019-09570-5

[18] Ren, J. and Zhang, X. (2005). Schilder theorem for the Brownian motion on the diffeomorphism group of the circle. *J. Funct. Anal.* **224** 107–133. MR2139106 https://doi.org/10.1016/j.jfa.2004.08.006

[19] Ren, J. and Zhang, X. (2005). Freidlin-Wentzell's large deviations for homeomorphism flows of non-Lipschitz SDEs. *Bull. Sci. Math.* **129** 643–655. MR2166732 https://doi.org/10.1016/j.bulsci.2004.12.005

[20] Situ, R. (2005). *Theory of Stochastic Differential Equations with Jumps and Applications*. *Mathematical and Analytical Techniques with Applications to Engineering*. New York: Springer. MR2160585

[21] Varadhan, S.R.S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** 261–286. MR0203230 https://doi.org/10.1002/cpa.3160190303

[22] Zhang, X. (2008). Euler schemes and large deviations for stochastic Volterra equations with singular kernels. *J. Differential Equations* **244** 2226–2250. MR2413840 https://doi.org/10.1016/j.jde.2008.02.019

Bernoulli **30**(1), 2024, 346–370
https://doi.org/10.3150/23-BEJ1600

Logarithmic law of large random correlation matrices

NESTOR PAROLYA[1,a], JOHANNES HEINY[2,c] and DOROTA KUROWICKA[1,b]

[1]*Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands,* [a]*n.parolya@tudelft.nl,* [b]*d.kurowicka@tudelft.nl*
[2]*Department of Mathematics, Stockholm University, Stockholm, Sweden,* [c]*johannes.heiny@math.su.se*

Consider a random vector $\mathbf{y} = \mathbf{\Sigma}^{1/2}\mathbf{x}$, where the p elements of the vector $\mathbf{x}$ are i.i.d. real-valued random variables with zero mean and finite fourth moment, and $\mathbf{\Sigma}^{1/2}$ is a deterministic $p \times p$ matrix such that the eigenvalues of the population correlation matrix $\mathbf{R}$ of $\mathbf{y}$ are uniformly bounded away from zero and infinity. In this paper, we find that the log determinant of the sample correlation matrix $\hat{\mathbf{R}}$ based on a sample of size n from the distribution of $\mathbf{y}$ satisfies a CLT (central limit theorem) for $p/n \to \gamma \in (0,1]$ and $p \le n$. Explicit formulas for the asymptotic mean and variance are provided. In case the mean of $\mathbf{y}$ is unknown, we show that after re-centering by the empirical mean the obtained CLT holds with a shift in the asymptotic mean. This result is of independent interest in both large dimensional random matrix theory and high-dimensional statistical literature of large sample correlation matrices for non-normal data. Finally, the obtained findings are applied for testing of uncorrelatedness of p random variables. Surprisingly, in the null case $\mathbf{R} = \mathbf{I}$, the test statistic becomes distribution-free and the extensive simulations show that the obtained CLT also holds if the moments of order four do not exist at all, which conjectures a promising and robust test statistic for heavy-tailed high-dimensional data.

Keywords: CLT; dependent data; large-dimensional asymptotic; log determinant; random matrix theory; sample correlation matrix

References

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. New Jersey: Wiley.

Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

Bao, Z., Pan, G. and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 88. MR2988403 https://doi.org/10.1214/EJP.v17-1962

Bao, Z., Pan, G. and Zhou, W. (2015). The logarithmic law of random determinant. *Bernoulli* **21** 1600–1628. MR3352055 https://doi.org/10.3150/14-BEJ615

Bodnar, T., Dette, H. and Parolya, N. (2019). Testing for independence of large dimensional vectors. *Ann. Statist.* **47** 2977–3008. MR3988779 https://doi.org/10.1214/18-AOS1771

Bodnar, T., Parolya, N. and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European J. Oper. Res.* **266** 371–390. MR3737003 https://doi.org/10.1016/j.ejor.2017.09.028

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

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

1350-7265

Dette, H. and Dörnemann, N. (2020). Likelihood ratio tests for many groups in high dimensions. *J. Multivariate Anal.* **178** 104605. MR4079037 https://doi.org/10.1016/j.jmva.2020.104605

Dörnemann, N. (2023). Likelihood ratio tests under model misspecification in high dimensions. *J. Multivariate Anal.* **193** Paper No. 105122. MR4504582 https://doi.org/10.1016/j.jmva.2022.105122

El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248 https://doi.org/10.1214/08-AAP548

Erdős, L. and Yau, H.-T. (2017). *A Dynamical Approach to Random Matrix Theory*. *Courant Lecture Notes in Mathematics* **28**. New York: Courant Institute of Mathematical Sciences. Available at http://www.math.harvard.edu/~htyau/RM-Aug-2016.pdf. MR3699468

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

Goodman, N.R. (1963). The distribution of the determinant of a complex Wishart distributed matrix. *Ann. Math. Stat.* **34** 178–180. MR0145619 https://doi.org/10.1214/aoms/1177704251

Grote, J., Kabluchko, Z. and Thäle, C. (2019). Limit theorems for random simplices in high dimensions. *ALEA Lat. Am. J. Probab. Math. Stat.* **16** 141–177. MR3903027 https://doi.org/10.30757/alea.v16-06

Hanea, A.M. and Nane, G.F. (2018). The asymptotic distribution of the determinant of a random correlation matrix. *Stat. Neerl.* **72** 14–33. MR3746785 https://doi.org/10.1111/stan.12113

Heiny, J. (2022). Large sample correlation matrices: A comparison theorem and its applications. *Electron. J. Probab.* **27** Paper No. 94. MR4456777 https://doi.org/10.1214/22-ejp817

Heiny, J., Johnston, S. and Prochno, J. (2022). Thin-shell theory for rotationally invariant random simplices. *Electron. J. Probab.* **27** Paper No. 2. MR4362288 https://doi.org/10.1214/21-ejp734

Heiny, J. and Mikosch, T. (2018). Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. *Stochastic Process. Appl.* **128** 2779–2815. MR3811704 https://doi.org/10.1016/j.spa.2017.10.002

Heiny, J., Mikosch, T. and Yslas, J. (2021). Point process convergence for the off-diagonal entries of sample covariance matrices. *Ann. Appl. Probab.* **31** 538–560. MR4254488 https://doi.org/10.1214/20-aap1597

Heiny, J. and Parolya, N. (2023). Log determinant of large correlation matrices under infinite fourth moment. *Ann. Inst. Henri Poincaré B Probab. Stat.*

Jiang, T. (2004a). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48. MR2082906

Jiang, T. (2004b). 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. (2019). Determinant of sample correlation matrix with application. *Ann. Appl. Probab.* **29** 1356–1397. MR3914547 https://doi.org/10.1214/17-AAP1362

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. MR3426306 https://doi.org/10.1111/sjos.12147

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

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *J. Multivariate Anal.* **97** 2177–2189. MR2301633 https://doi.org/10.1016/j.jmva.2005.05.010

Li, D., Qi, Y. and Rosalsky, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111** 256–270. MR2944420 https://doi.org/10.1016/j.jmva.2012.04.002

Liu, W.-D., Lin, Z. and Shao, Q.-M. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539 https://doi.org/10.1214/08-AAP527

Mao, G. (2014). A new test of independence for high-dimensional data. *Statist. Probab. Lett.* **93** 14–18. MR3244549 https://doi.org/10.1016/j.spl.2014.05.024

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0652932

Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.* **26** 1837–1887. MR3513608 https://doi.org/10.1214/15-AAP1135

Nguyen, H.H. and Vu, V. (2014). Random matrices: Law of the determinant. *Ann. Probab.* **42** 146–167. MR3161483 https://doi.org/10.1214/12-AOP791

Nielsen, J. (1999). The distribution of volume reductions induced by isotropic random projections. *Adv. in Appl. Probab.* **31** 985–994. MR1747452 https://doi.org/10.1239/aap/1029955254

Parolya, N., Heiny, J. and Kurowicka, D. (2024). Supplement to "Logarithmic law of large random correlation matrices." https://doi.org/10.3150/23-BEJ1600SUPP

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

Tao, T. and Vu, V. (2012). A central limit theorem for the determinant of a Wigner matrix. *Adv. Math.* **231** 74–101. MR2935384 https://doi.org/10.1016/j.aim.2012.05.006

Wang, X., Han, X. and Pan, G. (2018). The logarithmic law of sample covariance matrices near singularity. *Bernoulli* **24** 80–114. MR3706751 https://doi.org/10.3150/16-BEJ867

Wang, Q. and Yao, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916 https://doi.org/10.1214/13-EJS842

Wiens, D.P. (1992). On moments of quadratic forms in non-spherically distributed variables. *Statistics* **23** 265–270. MR1237804 https://doi.org/10.1080/02331889208802374

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**. New York: Cambridge Univ. Press. MR3468554 https://doi.org/10.1017/CBO9781107588080

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

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

Bernoulli **30**(1), 2024, 371–387
https://doi.org/10.3150/23-BEJ1601

Refined behaviour of a conditioned random walk in the large deviations regime

SØREN ASMUSSEN[1,a] and PETER W. GLYNN[2,b]

[1]*Department of Mathematics, Aarhus University, 8000 Aarhus C, Denmark,* [a]*asmus@math.au.dk*
[2]*Department of Management Science and Engineering, Stanford University, Stanford, CA 94305-4121, USA,* [b]*glynn@stanford.edu*

Conditioned limit theorems as $n \to \infty$ are given for the increments $X_1, \dots, X_n$ of a random walk $S_n = X_1 + \cdots + X_n$, subject to the conditionings $S_n \ge nb$ or $S_n = nb$ with $b > \mathbb{E}X$. The probabilities of these conditioning events are given by saddlepoint approximations, corresponding to the exponential tilting $f_\theta(x) = \mathrm{e}^{\theta x - \psi(\theta)} f(x)$ of the increment density $f(x)$, with θ satisfying $b = \mathbb{E}_\theta X = \psi'(\theta)$ where $\psi(\theta) = \log \mathbb{E}\mathrm{e}^{\theta X}$. It has been noted in various formulations that conditionally, the increment density somehow is close to $f_\theta(x)$. Sharp versions of such statements are given, including correction terms for segments $(X_1, \dots, X_k)$ with k fixed. Similar correction terms are given for the mean and variance of $\widehat{F}_n(x) - F_\theta(x)$ where $\widehat{F}_n$ is the empirical c.d.f. of $X_1, \dots, X_n$. Also a result on the total variation distance for segments with $k/n \to c \in (0,1)$ is derived. Further functional limit theorems for $\big(\widehat{F}_k(x), S_k\big)_{k \le n}$ are given, involving a bivariate conditioned Brownian limit.

Keywords: Boltzmann law; conditioned Brownian motion; empirical c.d.f.; exponential tilting; functional limit theorem; Gibbs conditioning; saddlepoint approximation; total variation distance

References

[1] Aboulalaâ, A. (1996). The Gibbs principle for Markov jump processes. *Stochastic Process. Appl.* **64** 257–271. MR1424238 https://doi.org/10.1016/S0304-4149(96)00092-0
[2] Asmussen, S. (1982). Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the $GI/G/1$ queue. *Adv. in Appl. Probab.* **14** 143–170. MR0644012 https://doi.org/10.2307/1426737
[3] Asmussen, S. and Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. New York: Springer. MR2331321
[4] Asmussen, S. and Klüppelberg, C. (1996). Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Process. Appl.* **64** 103–125. MR1419495 https://doi.org/10.1016/S0304-4149(96)00087-7
[5] Asmussen, S. and Steffensen, M. (2020). *Risk and Insurance—a Graduate Text*. *Probability Theory and Stochastic Modelling* **96**. Cham: Springer. MR4180989 https://doi.org/10.1007/978-3-030-35176-2
[6] Bahadur, R.R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Stat.* **31** 1015–1027. MR0117775 https://doi.org/10.1214/aoms/1177705674
[7] Bar-Lev, S.K., Schulte-Geers, E. and Stadje, W. (2013). Conditional limit theorems for the terms of a random walk revisited. *J. Appl. Probab.* **50** 871–882. MR3102520 https://doi.org/10.1239/jap/1378401242
[8] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. *Monographs on Statistics and Applied Probability*. London: CRC Press. MR1010226 https://doi.org/10.1007/978-1-4899-3424-6
[9] Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. *Monographs on Statistics and Applied Probability* **52**. London: CRC Press. MR1317097 https://doi.org/10.1007/978-1-4899-3210-5
[10] Bhattacharya, R.N. and Rao, R.R. (2010). *Normal Approximation and Asymptotic Expansions*. *Classics in Applied Mathematics* **64**. Philadelphia, PA: SIAM. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

[11] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[12] Borisov, I.S. (1978). On the rate of convergence in the "conditional" invariance principle. *Teor. Veroyatn. Primen.* **23** 67–79. MR0471011
[13] Borovkov, A.A. and Borovkov, K.A. (2008). *Asymptotic Analysis of Random Walks: Heavy-Tailed Distributions*. *Encyclopedia of Mathematics and Its Applications* **118**. Cambridge: Cambridge Univ. Press. MR2424161 https://doi.org/10.1017/CBO9780511721397
[14] Broniatowski, M. and Cao, Z. (2014). Light tails: All summands are large when the empirical mean is large. *Extremes* **17** 305–336. MR3207719 https://doi.org/10.1007/s10687-014-0187-8
[15] Bucklew, J.A., Ney, P. and Sadowsky, J.S. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59. MR1039183 https://doi.org/10.1017/s0021900200038419
[16] Dembo, A. and Zeitouni, O. (1996). Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields* **104** 1–14. MR1367663 https://doi.org/10.1007/BF01303799
[17] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. New York: Springer. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4
[18] Diaconis, P. and Freedman, D.A. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.* **1** 381–410. MR0958245 https://doi.org/10.1007/BF01048727
[19] Dobrushin, R. and Hryniv, O. (1996). Fluctuations of shapes of large areas under paths of random walks. *Probab. Theory Related Fields* **105** 423–458. MR1402652 https://doi.org/10.1007/BF01191908
[20] Durrett, R. (1978). Conditioned limit theorems for some null recurrent Markov processes. *Ann. Probab.* **6** 798–828. MR0503953
[21] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838085 https://doi.org/10.1002/9780470316658
[22] Grendar, M. (2004). Gibbs conditioning extended, Boltzmann conditioning introduced. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. *AIP Conf. Proc.* **735** 470–477. Melville, NY: Amer. Inst. Phys. MR2266280 https://doi.org/10.1063/1.1835245
[23] Hryniv, O. (1998). On conditional invariance principle for random walks. *Mat. Stud.* **9** 102–109, 112. MR1685988
[24] Iglehart, D.L. (1974). Functional central limit theorems for random walks conditioned to stay positive. *Ann. Probab.* **2** 608–619. MR0362499 https://doi.org/10.1214/aop/1176996607
[25] Jensen, J.L. (1995). *Saddlepoint Approximations*. *Oxford Statistical Science Series* **16**. New York: The Clarendon Press, Oxford University Press. MR1354837
[26] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8
[27] Lauritzen, S.L. (1988). *Extremal Families and Systems of Sufficient Statistics*. *Lecture Notes in Statistics* **49**. New York: Springer. MR0971253 https://doi.org/10.1007/978-1-4612-1023-8
[28] Lehtomaa, J. (2015). Limiting behaviour of constrained sums of two variables and the principle of a single big jump. *Statist. Probab. Lett.* **107** 157–163. MR3412770 https://doi.org/10.1016/j.spl.2015.08.017
[29] Lehtomaa, J. (2017). Large deviations of means of heavy-tailed random variables with finite moments of all orders. *J. Appl. Probab.* **54** 66–81. MR3632606 https://doi.org/10.1017/jpr.2016.87
[30] Léonard, C. and Najim, J. (2002). An extension of Sanov's theorem: Application to the Gibbs conditioning principle. *Bernoulli* **8** 721–743. MR1963659
[31] Liggett, T.M. (1968). An invariance principle for conditioned sums of independent random variables. *J. Math. Mech.* **18** 559–570. MR0238373 https://doi.org/10.1512/iumj.1969.18.18043
[32] Liggett, T.M. (1970). Weak convergence of conditioned sums of independent random vectors. *Trans. Amer. Math. Soc.* **152** 195–213. MR0268940 https://doi.org/10.2307/1995646
[33] Martin-Löf, P. (1970). Statistical Models. In Swedish. Unpublished lecture notes compiled by R. Sundberg. Stockholm University, available from http://www.math.su.se/publikationer.
[34] Mikosch, T. and Nagaev, A.V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1** 81–110. MR1652936 https://doi.org/10.1023/A:1009913901219

[35] Palmowski, Z. and Rolski, T. (2004). Markov processes conditioned to never exit a subspace of the state space. *Probab. Math. Statist.* **24** 339–353. MR2157210
[36] Petrov, V.V. (1965). On the probabilities of large deviations for sums of independent random variables. *Teor. Veroyatn. Primen.* **10** 310–322. MR0185645
[37] Rhee, C.-H., Blanchet, J. and Zwart, B. (2019). Sample path large deviations for Lévy processes and random walks with regularly varying increments. *Ann. Probab.* **47** 3551–3605. MR4038038 https://doi.org/10.1214/18-aop1319
[38] Stroock, D.W. and Zeitouni, O. (1991). Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In *Random Walks, Brownian Motion, and Interacting Particle Systems*. *Progress in Probability* **28** 399–424. Boston, MA: Birkhäuser. MR1146461
[39] Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. *Institute of Mathematical Statistics Textbooks* **12**. Cambridge: Cambridge Univ. Press. MR3969949 https://doi.org/10.1017/9781108604574
[40] Zabell, S.L. (1993). A limit theorem for expectations conditional on a sum. *J. Theoret. Probab.* **6** 267–283. MR1215658 https://doi.org/10.1007/BF01047574

Bernoulli **30**(1), 2024, 388–418
https://doi.org/10.3150/23-BEJ1602

Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals

FANGZHENG XIE[a]

Department of Statistics, Indiana University, Bloomington, United States, [a]*fxie@iu.edu*

We establish a finite-sample Berry-Esseen theorem for the entrywise limits of the eigenvectors for a broad collection of signal-plus-noise random matrix models under challenging weak signal regimes. The signal strength is characterized by a scaling factor ρ_n through $n\rho_n$, where n is the dimension of the random matrix, and we allow $n\rho_n$ to grow at the rate of $\log n$. The key technical contribution is a sharp finite-sample entrywise eigenvector perturbation bound. The existing error bounds on the two-to-infinity norms of the higher-order remainders are not sufficient when $n\rho_n$ is proportional to $\log n$. We apply the general entrywise eigenvector analysis results to the symmetric noisy matrix completion problem, random dot product graphs, and two subsequent inference tasks for random graphs: the estimation of pure nodes in mixed membership stochastic block models and the hypothesis testing of the equality of latent positions in random graphs.

Keywords: Berry-Esseen theorem; entrywise eigenvector analysis; random dot product graphs; signal-plus-noise matrix model; symmetric noisy matrix completion

References

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

[2] 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

[3] Abbe, E., Fan, J., Wang, K. and Zhong, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[4] Agterberg, J., Lubberts, Z. and Priebe, C.E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inf. Theory* **68** 4618–4650. MR4449064

[5] Airoldi, E.M., Blei, D.M., Fienberg, S.E. and Xing, E.P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[6] Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M. and Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832. MR3270750

[7] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR1990662

[8] Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C.E. and Vogelstein, J.T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142, 49. MR4318498

[9] Athreya, A., Fishkind, D.E., Tang, M., Priebe, C.E., Park, Y., Vogelstein, J.T., Levin, K., Lyzinski, V., Qin, Y. and Sussman, D.L. (2017). Statistical inference on random dot product graphs: A survey. *J. Mach. Learn. Res.* **18** Paper No. 226, 92. MR3827114

[10] Athreya, A., Priebe, C.E., Tang, M., Lyzinski, V., Marchette, D.J. and Sussman, D.L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. MR3494576 https://doi.org/10.1007/s13171-015-0071-x

1350-7265

[11] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8
[12] Bean, D., Bickel, P.J., El Karoui, N. and Yu, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568. https://doi.org/10.1073/pnas.1307845110
[13] Benaych-Georges, F. and Nadakuditi, R.R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007
[14] Bennett, J., Lanning, S. et al. (2007). The Netflix prize. In *Proceedings of KDD Cup and Workshop* **2007** 35–35. https://doi.org/10.1145/1345448.1345459
[15] Bhatia, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. New York: Springer. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8
[16] Cai, T.T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541
[17] Candès, E.J. and Plan, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory* **57** 2342–2359. MR2809094 https://doi.org/10.1109/TIT.2011.2111771
[18] Candès, E.J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5
[19] Candès, E.J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061
[20] Cape, J. (2020). A note on the orthogonal Procrustes problem and norm-dependent optimality. *Electron. J. Linear Algebra* **36** 158–168. MR4089048 https://doi.org/10.13001/ela.2020.5009
[21] Cape, J., Tang, M. and Priebe, C.E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070
[22] Cape, J., Tang, M. and Priebe, C.E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752
[23] Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272
[24] Chung, K.L. (2001). *A Course in Probability Theory*. New York-London: Academic Press. MR0346858
[25] Davis, C. and Kahan, W.M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001
[26] Donoho, D. and Gavish, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257
[27] Donoho, D.L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* **52** 1289–1306. MR2241189 https://doi.org/10.1109/TIT.2006.871582
[28] Eldar, Y. C. and Kutyniok, G. (2012). *Compressed Sensing: Theory and Applications*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511794308
[29] Eldridge, J., Belkin, M. and Wang, Y. (2018). Unperturbed: Spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory 2018*. *Proc. Mach. Learn. Res. (PMLR)* **83** 38. Proceedings of Machine Learning Research PMLR, [place of publication not identified]. MR3857310
[30] Fan, J., Fan, Y., Han, X. and Lv, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990
[31] Fan, J., Fan, Y., Han, X. and Lv, J. (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505
[32] Fan, J., Wang, W. and Zhong, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42. MR3827095
[33] Gao, C., Ma, Z., Zhang, A.Y. and Zhou, H.H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45. MR3687603
[34] Gillis, N. and Vavasis, S.A. (2014). Fast and robust recursive algorithms for separable nonnegative matrix factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 698–714. https://doi.org/10.1109/TPAMI.2013.226

[35] Girvan, M. and Newman, M.E.J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826. MR1908073 https://doi.org/10.1073/pnas.122653799

[36] Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM* **35** 61–70. https://doi.org/10.1145/138859.138867

[37] Gopalan, P.K. and Blei, D.M. (2013). Efficient discovery of overlapping communities in massive networks. *Proc. Natl. Acad. Sci. USA* **110** 14534–14539. MR3105375 https://doi.org/10.1073/pnas.1221839110

[38] Gross, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999

[39] 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

[40] Hopkins, S.B. and Steurer, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 379–390. Los Alamitos, CA: IEEE Computer Soc. MR3734245 https://doi.org/10.1109/FOCS.2017.42

[41] Jain, P., Netrapalli, P. and Sanghavi, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing* 665–674. New York: ACM. MR3210828 https://doi.org/10.1145/2488608.2488693

[42] Javanmard, A. and Montanari, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630

[43] Jin, J., Ke, Z.T. and Luo, S. (2023). Mixed membership estimation for social networks. *J. Econometrics*. https://doi.org/10.1016/j.jeconom.2022.12.003

[44] Johnstone, I.M. and Lu, A.Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

[45] Karrer, B. and Newman, M.E.J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)* **83** 016107, 10. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[46] Keshavan, R.H., Montanari, A. and Oh, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022

[47] 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

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

[49] Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

[50] Lei, L. (2019). Unified $\ell_{2\to\infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. Preprint. Available at arXiv:1909.04798. https://doi.org/10.48550/arXiv.1909.04798

[51] Levin, K., Athreya, A., Tang, M., Lyzinski, V. and Priebe, C. E. (2017). A central limit theorem for an omnibus embedding of random dot product graphs. Preprint. Available at arXiv:1705.09355. https://doi.org/10.48550/arXiv.1705.09355

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

[53] Lyzinski, V., Sussman, D.L., Tang, M., Athreya, A. and Priebe, C.E. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electron. J. Stat.* **8** 2905–2922. MR3299126 https://doi.org/10.1214/14-EJS978

[54] Lyzinski, V., Tang, M., Athreya, A., Park, Y. and Priebe, C.E. (2017). Community detection and classification in hierarchical stochastic blockmodels. *IEEE Trans. Netw. Sci. Eng.* **4** 13–26. MR3625952 https://doi.org/10.1109/TNSE.2016.2634322

[55] Mao, X., Sarkar, P. and Chakrabarti, D. (2017). On mixed memberships and symmetric nonnegative matrix factorizations. In *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y.W. Teh, eds.). *Proc. Mach. Learn. Res.* **70** 2324–2333. PMLR.

[56] Mao, X., Sarkar, P. and Chakrabarti, D. (2021). Estimating mixed memberships with sharp eigenvector deviations. *J. Amer. Statist. Assoc.* **116** 1928–1940. MR4353723 https://doi.org/10.1080/01621459.2020.1751645

[57] Neil, J., Uphoff, B., Hash, C. and Storlie, C. (2013). Towards improved detection of attackers in computer networks: New edges, fast updating, and host agents. In *2013 6th International Symposium on Resilient Control Systems (ISRCS)* 218–224. https://doi.org/10.1109/ISRCS.2013.6623779

[58] Nickel, C.L.M. (2007). *Random Dot Product Graphs: A Model for Social Networks*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)–The Johns Hopkins University. MR2710154

[59] 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

[60] Priebe, C.E., Park, Y., Tang, M., Athreya, A., Lyzinski, V., Vogelstein, J.T., Qin, Y., Cocanougher, B., Eichler, K. and Zlatic, M. et al. (2017). Semiparametric spectral modeling of the Drosophila connectome. Preprint. Available at arXiv:1705.03297. https://doi.org/10.48550/arXiv.1705.03297

[61] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[62] Rubin-Delanchy, P., Adams, N.M. and Heard, N.A. (2016). Disassortativity of computer networks. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)* 243–247. https://doi.org/10.1109/ISI.2016.7745482

[63] Sarkar, P. and Bickel, P.J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43** 962–990. MR3346694 https://doi.org/10.1214/14-AOS1285

[64] Schönemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika* **31** 1–10. MR0215870 https://doi.org/10.1007/BF02289451

[65] Shabalin, A.A. and Nobel, A.B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. MR3054091 https://doi.org/10.1016/j.jmva.2013.03.005

[66] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905. https://doi.org/10.1109/34.868688

[67] Stewart, G.W. and Sun, J.G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Boston, MA: Academic Press. MR1061154

[68] Sun, Y., Baricz, Á. and Zhou, S. (2010). On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q-functions. *IEEE Trans. Inf. Theory* **56** 1166–1186. MR2723669 https://doi.org/10.1109/TIT.2009.2039048

[69] Sussman, D.L., Tang, M., Fishkind, D.E. and Priebe, C.E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899 https://doi.org/10.1080/01621459.2012.699795

[70] Sussman, D.L., Tang, M. and Priebe, C.E. (2014). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 48–57. https://doi.org/10.1109/TPAMI.2013.135

[71] Tang, M., Athreya, A., Sussman, D.L., Lyzinski, V. and Priebe, C.E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872 https://doi.org/10.3150/15-BEJ789

[72] Tang, M. and Priebe, C.E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.* **46** 2360–2415. MR3845021 https://doi.org/10.1214/17-AOS1623

[73] Tang, M., Sussman, D.L. and Priebe, C.E. (2013). Universally consistent vertex classification for latent positions graphs. *Ann. Statist.* **41** 1406–1430. MR3113816 https://doi.org/10.1214/13-AOS1112

[74] Tang, R., Ketcha, M., Badea, A., Calabrese, E.D., Margulies, D.S., Vogelstein, J.T., Priebe, C.E. and Sussman, D.L. (2019). Connectome smoothing via low-rank approximations. *IEEE Trans. Med. Imag.* **38** 1446–1456. https://doi.org/10.1109/TMI.2018.2885968

[75] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

[77] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

[78] Wedin, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

[79] Xia, D. and Zhou, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** Paper No. 61, 42. MR3960915

[80] Xie, F. (2024). Supplement to "Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals." https://doi.org/10.3150/23-BEJ1602SUPP

[81] Xie, F. and Xu, Y. (2020). Optimal Bayesian estimation for random dot product graphs. *Biometrika* **107** 875–889. MR4186493 https://doi.org/10.1093/biomet/asaa031

[82] Xie, F. and Xu, Y. (2023). Efficient estimation for random dot product graphs via a one-step procedure. *J. Amer. Statist. Assoc.* **118** 651–664. MR4571148 https://doi.org/10.1080/01621459.2021.1948419

[83] 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**. New York: Cambridge Univ. Press. MR3468554 https://doi.org/10.1017/CBO9781107588080

[84] Young, S.J. and Scheinerman, E.R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph*. *Lecture Notes in Computer Science* **4863** 138–149. Berlin: Springer. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

[85] Zhong, Y. and Boumal, N. (2018). Near-optimal bounds for phase synchronization. *SIAM J. Optim.* **28** 989–1016. MR3782406 https://doi.org/10.1137/17M1122025

Bernoulli **30**(1), 2024, 419–447
https://doi.org/10.3150/23-BEJ1603

Limit theorems for Fréchet mean sets

STEVEN N. EVANS[a] and ADAM Q. JAFFE[b]

Department of Statistics, University of California, Berkeley, USA, [a]*evans@stat.berkeley.edu,*
[b]*aqjaffe@berkeley.edu*

For $1 \leq p \leq \infty$, the Fréchet p-mean of a probability measure on a metric space is an important notion of central tendency that generalizes the usual notions in the real line of mean ($p = 2$) and median ($p = 1$). In this work we prove a collection of limit theorems for Fréchet means and related objects, which, in general, constitute a sequence of random closed sets. On the one hand, we show that many limit theorems (a strong law of large numbers, an ergodic theorem, and a large deviations principle) can be simply descended from analogous theorems on the space of probability measures via purely topological considerations. On the other hand, we provide the first sufficient conditions for the strong law of large numbers to hold in a T_2 topology (in particular, the Fell topology), and we show that this condition is necessary in some special cases. We also discuss statistical and computational implications of the results herein.

Keywords: Kuratowski convergence; Hausdorff metric; Wasserstein metric; random sets; Karcher mean; non-Euclidean statistics; medoids

References

[1] Adamaszek, M., Adams, H. and Frick, F. (2018). Metric reconstruction via optimal transport. *SIAM J. Appl. Algebra Geom.* **2** 597–619. MR3871057 https://doi.org/10.1137/17M1148025

[2] Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5

[3] Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0

[4] Bagaria, V., Kamath, G., Ntranos, V., Zhang, M. and Tse, D. (2018). Medoids in almost-linear time via multi-armed bandits (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 500–509. Playa Blanca, Lanzarote, Canary Islands: PMLR.

[5] Baharav, T. and Tse, D. (2019). Ultra fast medoid identification via correlated sequential halving. In *Advances in Neural Information Processing Systems* 3655–3664.

[6] Beer, G. (1993). *Topologies on Closed and Closed Convex Sets*. *Mathematics and Its Applications* **268**. Dordrecht: Kluwer Academic. MR1269778 https://doi.org/10.1007/978-94-015-8149-3

[7] Bhattacharya, A. (2008). Statistical analysis on manifolds: A nonparametric approach for inference on shape spaces. *Sankhyā* **70** 223–266. MR2551814

[8] Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds: With Applications to Shape Spaces*. *Institute of Mathematical Statistics (IMS) Monographs* **2**. Cambridge: Cambridge Univ. Press. MR2934285 https://doi.org/10.1017/CBO9781139094764

[9] Bhattacharya, R. and Lin, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. MR3565392 https://doi.org/10.1090/proc/13216

[10] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

[11] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

[12] Blanchard, M. and Jaffe, A.Q. Fréchet Mean Set Estimation in the Hausdorff Metric, via Relaxation. Preprint. Available at arXiv:2212.12057.

1350-7265

[13] Cartan, É. (1963). *Leçons sur la Géométrie des Espaces de Riemann*. Paris: Gauthier-Villars. MR0020842
[14] Chakraborty, R. and Vemuri, B.C. (2015). Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
[15] Chakraborty, R. and Vemuri, B.C. (2019). Statistics on the Stiefel manifold: Theory and applications. *Ann. Statist.* **47** 415–438. MR3909937 https://doi.org/10.1214/18-AOS1692
[16] Dembo, A. and Zeitouni, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Berlin: Springer. MR2571413 https://doi.org/10.1007/978-3-642-03311-7
[17] Dryden, I.L., Koloydenko, A. and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249
[18] Engelking, R. (1956). Sur l'impossibilité de définir la limite topologique inférieure à l'aide des opérations dénombrables de l'algèbre de Boole et de l'opération de fermeture. *Bull. Acad. Pol. Sci. Cl. III* **4** 659–662. MR0089398
[19] Fiori, S. (2009). Learning the Fréchet mean over the manifold of symmetric positive-definite matrices. *Cogn. Comput.* **1** 279.
[20] Ginestet, C.E. (2012). Strong consistency of Fréchet sample mean sets for graph-valued random variables. Preprint. Available at arXiv:1204.3183.
[21] Grove, K. and Karcher, H. (1973). How to conjugate C^1-close group actions. *Math. Z.* **132** 11–20. MR0356104 https://doi.org/10.1007/BF01214029
[22] Hotz, T., Huckemann, S., Le, H., Marron, J.S., Mattingly, J.C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. and Skwerer, S. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23** 2238–2258. MR3127934 https://doi.org/10.1214/12-AAP899
[23] Huckemann, S.F. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349 https://doi.org/10.1214/10-AOS862
[24] Huckemann, S.F. (2014). Stickiness and smeariness. In *Mini-Workshop: Asymptotic Statistics on Stratified Spaces*. *Mathematisches Forschungsinstitut Oberwolfach*. Zurich: EMS Pub. House.
[25] Huckemann, S.F. (2015). (Semi-)Intrinsic Statistical Analysis on Non-Euclidean Spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics* 103–118. Cham: Springer.
[26] Jaffe, A.Q. (2022). Strong Consistency for a Class of Adaptive Clustering Procedures. Preprint. Available at arXiv:2202.13423.
[27] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8
[28] Kaufman, L. and Rousseeuw, P.J. (1987). Clustering by means of medoids: Statistical data analysis based on the L^1 norm. *Y. Dodge, Ed* 405–416.
[29] Kelley, J.L. (1955). *General Topology*. Toronto–New York: D. Van Nostrand Co., Inc. MR0070144
[30] Kolaczyk, E.D., Lin, L., Rosenberg, S., Walters, J. and Xu, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Statist.* **48** 514–538. MR4065172 https://doi.org/10.1214/19-AOS1820
[31] Kuratowski, K. (1966). *Topology. Vol. I*. New York–London: Academic Press; Warsaw: Państwowe Wydawnictwo Naukowe [Polish Scientific Publishers]. MR0217751
[32] Le, H. and Kume, A. (2000). The Fréchet mean shape and the shape of the means. *Adv. in Appl. Probab.* **32** 101–113. MR1765168 https://doi.org/10.1239/aap/1013540025
[33] Schötz, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618
[34] Schötz, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics* **56** 34–52. MR4392146 https://doi.org/10.1080/02331888.2022.2032063
[35] Srivastava, S.M. (1998). *A Course on Borel Sets*. *Graduate Texts in Mathematics* **180**. New York: Springer. MR1619545 https://doi.org/10.1007/978-3-642-85473-6
[36] 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. Providence, RI: Amer. Math. Soc. MR2039961 https://doi.org/10.1090/conm/338/06080

[37] Sverdrup-Thygeson, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. *Ann. Statist.* **9** 141–145. MR0600540

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

[39] Turner, K., Mileyko, Y., Mukherjee, S. and Harer, J. (2014). Fréchet means for distributions of persistence diagrams. *Discrete Comput. Geom.* **52** 44–70. MR3231030 https://doi.org/10.1007/s00454-014-9604-7

[40] Varadarajan, V.S. (1958). On the convergence of sample probability distributions. *Sankhyā* **19** 23–26. MR0094839

[41] Villani, C. (2009). *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Berlin: Springer. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[42] Wang, R., Wang, X. and Wu, L. (2010). Sanov's theorem in the Wasserstein distance: A necessary and sufficient condition. *Statist. Probab. Lett.* **80** 505–512. MR2593592 https://doi.org/10.1016/j.spl.2009.12.003

[43] Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A* 591–602. Dordrecht: Reidel. MR0501230

Bernoulli **30**(1), 2024, 448–474
https://doi.org/10.3150/23-BEJ1604

Tracy-Widom law for the extreme eigenvalues of large signal-plus-noise matrices

ZHIXIANG ZHANG[1,a], YIMING LIU[2,c] and GUANGMING PAN[1,b]

[1]*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore,* [a]*S160011@e.ntu.edu.sg,* [b]*gmpan@ntu.edu.sg*
[2]*Department of Statistics, School of Economics, Jinan University, Guangzhou, China,* [c]*liuyiming@jnu.edu.cn*

Let $\mathbf{S} = \mathbf{R} + \mathbf{X}$ be an $M \times N$ matrix where $\mathbf{R}$ is the signal matrix and $\mathbf{X}$ is the noise matrix consisting of *i.i.d.* standardized entries. The signal matrix $\mathbf{R}$ is allowed to be full rank, which is rarely studied in literature compared with the low rank cases. Under a regularity condition of $\mathbf{R}$ that assures the square root behaviour of the spectral density near the edge, we prove that the largest eigenvalue of $\mathbf{SS}^*$ has the Tracy-Widom distribution under a tail condition on the entries of $\mathbf{X}$. Moreover, such a condition is proved to be necessary and sufficient to assure the Tracy-Widom law.

Keywords: Edge universality; extreme eigenvalues; signal-plus-noise matrix; Tracy-Widom law

References

[1] Banna, M., Najim, J. and Yao, J. (2020). A CLT for linear spectral statistics of large random information-plus-noise matrices. *Stochastic Process. Appl.* **130** 2250–2281. MR4074699 https://doi.org/10.1016/j.spa.2019.06.017
[2] Bao, Z., Ding, X. and Wang, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960
[3] Bao, Z., Hu, J., Pan, G. and Zhou, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 https://doi.org/10.1214/18-AOS1704
[4] Benaych-Georges, F. and Nadakuditi, R.R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007
[5] Cape, J., Tang, M. and Priebe, C.E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070
[6] Chang, C.-I. and Du, Q. (1999). Interference and noise-adjusted principal components analysis. *IEEE Trans. Geosci. Remote Sens.* **37** 2387–2396.
[7] Chaudhuri, K., Sarwate, A.D. and Sinha, K. (2013). A near-optimal algorithm for differentially-private principal components. *J. Mach. Learn. Res.* **14** 2905–2943. MR3121662
[8] Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129
[9] Ding, X. and Yang, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.* **28** 1679–1738. MR3809475 https://doi.org/10.1214/17-AAP1341
[10] Ding, X. and Yang, F. (2022). Edge statistics of large dimensional deformed rectangular matrices. *J. Multivariate Anal.* **192** Paper No. 105051, 22 pp. MR4438212 https://doi.org/10.1016/j.jmva.2022.105051
[11] Ding, X. and Yang, F. (2022). Tracy-Widom distribution for heterogeneous Gram matrices with applications in signal detection. *IEEE Trans. Inf. Theory* **68** 6682–6715. MR4502560
[12] Dozier, R.B. and Silverstein, J.W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123 https://doi.org/10.1016/j.jmva.2006.09.006

1350-7265

[13] Dozier, R.B. and Silverstein, J.W. (2007). Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *J. Multivariate Anal.* **98** 1099–1122. MR2326242 https://doi.org/10.1016/j.jmva.2006.12.005

[14] Erdős, L., Knowles, A. and Yau, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y

[15] Erdős, L., Yau, H.-T. and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 https://doi.org/10.1016/j.aim.2011.12.010

[16] Erlingsson, Ú., Pihur, V. and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proc. ACM Conf. Comput. Commun. Secur.* 1054–1067.

[17] Fan, Z. and Johnstone, I.M. (2022). Tracy-Widom at each edge of real covariance and MANOVA estimators. *Ann. Appl. Probab.* **32** 2967–3003. MR4474524 https://doi.org/10.1214/21-aap1754

[18] Hachem, W., Loubaton, P. and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925

[19] Hachem, W., Loubaton, P., Najim, J. and Vallet, P. (2013). On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 36–63. MR3060147 https://doi.org/10.1214/11-AIHP450

[20] Han, X., Pan, G. and Yang, Q. (2018). A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli* **24** 3447–3468. MR3788178 https://doi.org/10.3150/17-BEJ965

[21] Han, X., Pan, G. and Zhang, B. (2016). The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933 https://doi.org/10.1214/15-AOS1427

[22] Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991 https://doi.org/10.1007/s002200050027

[23] 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

[24] Knowles, A. and Yin, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909 https://doi.org/10.1002/cpa.21450

[25] 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

[26] Landon, B. and Yau, H.-T. (2017). Edge statistics of Dyson Brownian motion. Preprint. Available at arXiv:1712.03881.

[27] Lee, J.O. and Schnelli, K. (2013). Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *J. Math. Phys.* **54** 103504, 62 pp. MR3134604 https://doi.org/10.1063/1.4823718

[28] Lee, J.O. and Schnelli, K. (2015). Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** 1550018, 94 pp. MR3405746 https://doi.org/10.1142/S0129055X1550018X

[29] Lee, J.O. and Schnelli, K. (2016). Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818 https://doi.org/10.1214/16-AAP1193

[30] Lee, J.O. and Yin, J. (2014). A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.* **163** 117–173. MR3161313 https://doi.org/10.1215/00127094-2414767

[31] Loubaton, P. and Vallet, P. (2011). Almost sure localization of the eigenvalues in a Gaussian information plus noise model – Application to the spiked models. *Electron. J. Probab.* **16** 1934–1959. MR2851051 https://doi.org/10.1214/EJP.v16-943

[32] Mani, V.V. and Bose, R. (2010). Direction of arrival estimation and beamforming of multiple coherent UWB signals. In *2010 IEEE Int. Conf. Commun.* 1–5. IEEE.

[33] Montanari, A., Reichman, D. and Zeitouni, O. (2015). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank-one perturbations of Gaussian tensors. *Adv. Neural Inf. Process. Syst.* **28** 217–225.

[34] Nadakuditi, R.R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661

[35] Onatski, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070 https://doi.org/10.3982/ECTA6964
[36] Svensson, A. and Jakobsson, A. (2011). Adaptive detection of a partly known signal corrupted by strong interference. *IEEE Signal Process. Lett.* **18** 729–732.
[37] Tao, T. and Vu, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449 https://doi.org/10.1007/s00220-010-1044-5
[38] Team, A.D.P. (2017). Learning with privacy at scale. *Apple Mach. Learn. J.* **1** 1–25.
[39] Tracy, C.A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246
[40] Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083
[41] Vallet, P., Loubaton, P. and Mestre, X. (2012). Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Trans. Inf. Theory* **58** 1043–1068. MR2918009 https://doi.org/10.1109/TIT.2011.2173718
[42] Vinogradova, J., Björnson, E. and Larsson, E.G. (2016). On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination. In *Proc. – ICASSP IEEE Int. Conf. Acoust.* 3421–3425. IEEE.
[43] Yang, D., Ma, Z. and Buja, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** Paper No. 92, 27 pp. MR3543498
[44] Yang, F. (2019). Edge universality of separable covariance matrices. *Electron. J. Probab.* **24** Paper No. 123, 57 pp. MR4029426 https://doi.org/10.1214/19-ejp381
[45] Yang, M., Lyu, L., Zhao, J., Zhu, T. and Lam, K.-Y. (2020). Local differential privacy and its applications: A comprehensive survey. Preprint. Available at arXiv:2008.03686.
[46] Zhang, Z., Liu, Y. and Pan, G. (2024). Supplement to "Tracy-Widom law for the extreme eigenvalues of large signal-plus-noise matrices." https://doi.org/10.3150/23-BEJ1604SUPP

Bernoulli **30**(1), 2024, 475–502
https://doi.org/10.3150/23-BEJ1605

Convergence rates for shallow neural networks learned by gradient descent

ALINA BRAUN[1,a], MICHAEL KOHLER[1,b], SOPHIE LANGER[2,c] and HARRO WALK[3,d]

[1]*Department of Mathematics, Technical University of Darmstadt, Darmstadt, Germany,* [a]*braun@mathematik.tu-darmstadt.de,* [b]*kohler@mathematik.tu-darmstadt.de*
[2]*Department of Applied Mathematics, University of Twente, Enschede, The Netherlands,* [c]*s.langer@utwente.nl*
[3]*Department of Mathematics, University of Stuttgart, Stuttgart, Germany,* [d]*walk@mathematik.uni-stuttgart.de*

In this paper we analyze the L_2 error of neural network regression estimates with one hidden layer. Under the assumption that the Fourier transform of the regression function decays suitably fast, we show that an estimate, where all initial weights are chosen according to proper uniform distributions and where the weights are learned by gradient descent, achieves a rate of convergence of $1/\sqrt{n}$ (up to a logarithmic factor). Our statistical analysis implies that the key aspect behind this result is the proper choice of the initial inner weights and the adjustment of the outer weights via gradient descent. This indicates that we can also simply use linear least squares to choose the outer weights. We prove a corresponding theoretical result and compare our new linear least squares neural network estimate with standard neural network estimates via simulated data. Our simulations show that our theoretical considerations lead to an estimate with an improved performance in many cases.

Keywords: Deep learning; gradient descent; rate of convergence; neural networks

References

Allen-Zhu, Z., Li, Y. and Liang, Y. (2019). Learning and generalization in overparameterized neural networks, going beyond two layers. In *Advances in Neural Information Processing Systems* **32**.

Allen-Zhu, Z., Li, Y. and Song, Z. (2019). A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning* **97** 242–252.

Andoni, A., Panigraphy, R., Valiant, G. and Zhang, L. (2014). Learning polynomials with neural networks. In *International Conference on Machine Learning* 1908–1916.

Arora, S., Cohen, N., Golowich, N. and Hu, W. (2018). A convergence analysis of gradient descent for deep linear neural networks. In *International Conference on Learning Representations*.

Arora, S., Du, S., Hu, W., Li, Z., Salakhutdinov, R. and Wang, R. (2019a). On exact computation with an infinitely wide neural net. In *Advances in Neural Information Processing Systems* **32**.

Arora, S., Du, S., Hu, W., Li, Z. and Wang, R. (2019b). Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning* **97** 322–332.

Barron, A.R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

Barron, A.R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.

Bartlett, P.L., Harvey, N., Liaw, C. and Mehrabian, A. (2019). Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.* **20** Paper No. 63, 17. MR3960917

Bauer, B. and Kohler, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747

Braun, A., Kohler, M., Langer, S. and Walk, H. (2024). Supplement to "Convergence rates for shallow neural networks learned by gradient descent." https://doi.org/10.3150/23-BEJ1605SUPP

1350-7265

Brutzkus, A., Globerson, A., Malach, E. and Shalew-Shwartz, S. (2018). SGD learn overparametrized networks that provably generalize on linearly seperable data. In *International Conference on Learning Representation*.

Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B. and LeCun, Y. (2015). The loss surface of multilayer networks. In *International Conference on Artificial Intelligence and Statistics* **38** 192–204.

Du, S. and Lee, J. (2018). On the power of over-parametrization in neural networks with quadratic activation. In *International Conference on Machine Learning* **80** 1329–1338.

Du, S., Lee, J., Li, H., Wang, I. and Zhai, X. (2019). Gradient descent finds global minima of deep neural networks. In *International Conference on Machine Learning* **97** 1675–1685.

Dudek, G. (2019). Generating random weights and biases in feedforward neural networks with random hidden nodes. *Inf. Sci.* **481** 33–56.

Epstein, C.L. (2008). *Introduction to the Mathematics of Medical Imaging*, 2nd ed. Philadelphia, PA: SIAM. MR2378706 https://doi.org/10.1137/1.9780898717792

Ghorbani, B., Mei, S., Misiakiewicz, T. and Montanari, A. (2019). Limitations of lazy training of two-layer neural networks. In *Advances in Neural Information Processing Systems* **32**.

Gonon, L. (2021). Random feature neural networks learn Black-Scholes type PDEs without curse of dimensionality. Preprint, arXiv:2106.08900.

Gonon, L., Grigoryeva, L. and Ortega, J.-P. (2023). Approximation bounds for random neural networks and reservoir systems. *Ann. Appl. Probab.* **33** 28–69. MR4551542 https://doi.org/10.1214/22-aap1806

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR3617773

Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution–Free Theory of Nonparametric Regression*. New York: Springer.

Huang, G.-B., Chen, L. and Siew, C.-K. (2006). Universal approximation using incremental contractive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **17** 879–892.

Igelnik, B. and Pao, Y.H. (1995). Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans. Neural Netw.* **6** 1320–1329. https://doi.org/10.1109/72.471375

Imaizumi, M. and Fukamizu, K. (2018). Deep neural networks learn non-smooth functions effectively. In *International Conference on Artificial Intelligence and Statistics* 869–878.

Jacot, A., Gabriel, F. and Hongler, C. (2020). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* **31**.

Javanmard, A., Mondelli, M. and Montanari, A. (2020). Analysis of a two-layer neural network via displacement convexity. *Ann. Statist.* **48** 3619–3642. MR4185822 https://doi.org/10.1214/20-AOS1945

Kawaguchi, K. (2016). Deep learning without poor local minima. *Adv. Neural Inf. Process. Syst.* **29**.

Kawaguchi, K. and Huang, J. (2019). Gradient descent finds global minima for generalizable deep neural networks of practical sizes. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 92–99.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Empirical Methods in Natural Language Processing* 1746–1751.

Kohler, M. and Krzyżak, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Trans. Inf. Theory* **63** 1620–1630. MR3625984 https://doi.org/10.1109/TIT.2016.2634401

Kohler, M. and Krzyżak, A. (2021). Over-parametrized deep neural networks minimizing the empirical risk do not generalize well. *Bernoulli* **27** 2564–2597. MR4303896 https://doi.org/10.3150/21-BEJ1323

Kohler, M., Krzyżak, A. and Langer, S. (2022). Estimation of a function of low local dimensionality by deep neural networks. *IEEE Trans. Inf. Theory* **68** 4032–4042. MR4433267

Kohler, M. and Langer, S. (2020). Discussion of: "Nonparametric regression using deep neural networks with ReLU activation function". *Ann. Statist.* **48** 1906–1910. MR4134777 https://doi.org/10.1214/19-AOS1912

Kohler, M. and Langer, S. (2021). On the rate of convergence of fully connected deep neural network regression estimates. *Ann. Statist.* **49** 2231–2249. MR4319248 https://doi.org/10.1214/20-aos2034

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

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** 1097–1105.

Kůrková, V. and Sanguineti, M. (2008). Geometric upper bounds on rates of variable-basis approximation. *IEEE Trans. Inf. Theory* **54** 5681–5688. MR2596808 https://doi.org/10.1109/TIT.2008.2006383

Lee, W.S. (1996). Agnostic Learning and Single Hidden Layer Neural Networks PhD Thesis, Australian National Univ.

Li, Y. and Liang, Y. (2018). Learning overparameterized neural networks via stochastic gradient descent on structured data. In *Advances in Neural Information Processing Systems* 8168–8177.

Liang, S., Sun, R., Lee, J. and Srikant, R. (2018). Adding one neuron can eliminate all bad local minima. In *Advances in Neural Information Processing Systems* 4355–4365.

McCaffrey, D.F. and Gallant, A.R. (1994). Convergence rates for single hidden layer feedforward networks. *Neural Netw.* **7** 147–158.

Pisier, G. (1981). Remarques sur un résultat non publié de B. Maurey. In *Seminar on Functional Analysis, 1980–1981* Exp. No. V, 13. Palaiseau: École Polytech. MR0659306

Poggio, T., Banburski, A. and Liao, Q. (2020). Theoretical issues in deep networks. *Proc. Natl. Acad. Sci. USA* **117** 30039–30045. MR4263286 https://doi.org/10.1073/pnas.1907369117

Rahimi, A. and Recht, B. (2009). Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. *Adv. Neural Inf. Process. Syst.* **21** 1313–1320.

Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Huber, T. et al. (2017). Mastering the game of go without human knowledge. *Nature* **550** 354–359.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D. and Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems* **33** 7462–7473.

Sonoda, S., Ishikawa, I. and Ikeda, M. (2021). Ridge regression with overparametrized two-layer networks converge to ridgelet spectrum. In *International Conference on Artificial Intelligence and Statistics* **130** 2674–2682.

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

Suzuki, T. and Nitanda, A. (2021). Deep learning is adaptive to intrinsic diemsnionality of model smoothness in anisotropic Besov space. *Adv. Neural Inf. Process. Syst.* **34** 3609–3621.

Widrow, B., Greenblatt, A., Kim, Y. and Park, D. (2013). The No-Prop algorithm: A new learning algorithm for multilayer neural networks. *Neural Netw.* **37** 182–188.

Woodworth, B., Gunasekar, S., Lee, J., Moroshko, E., Savarese, P., Golan, I., Soudry, D. and Srebro, N. (2020). Kernel and rich regimes in overparametrized models. In *Conference on Learning Theory* **125** 3635–3673.

Wu, Y., Schuster, M., Chen, Z., Le, Q., Norouzi, M., Macherey, W., Krikum, M. et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv:1609.08144.

Yosida, K. (1968). *Functional Analysis*, 2nd ed. *Die Grundlehren der Mathematischen Wissenschaften* **123**. New York: Springer New York, Inc. MR0239384

Zou, D., Cao, Y., Zhou, D. and Gu, Q. (2020). Gradient descent optimizes over-parameterized deep ReLU networks. *Mach. Learn.* **109** 467–492. MR4075425 https://doi.org/10.1007/s10994-019-05839-6

Bernoulli **30**(1), 2024, 503–533
https://doi.org/10.3150/23-BEJ1606

Tail inverse regression: Dimension reduction for prediction of extremes

ANASS AGHBALOU[1,a], FRANÇOIS PORTIER[2,b], ANNE SABOURIN[3,c] and CHEN ZHOU[4,5,d]

[1]*LTCI, Télécom Paris, Institut polytechnique de Paris, Palaiseau, France,* [a]*anass.aghbalou@telecom-paris.fr*
[2]*Ensai, CREST - UMR 9194, Rennes, France,* [b]*francois.portier@gmail.com*
[3]*Université Paris Cité, CNRS, MAP5, F-75006 Paris, France,* [c]*anne.sabourin@u-paris.fr*
[4]*Erasmus University Rotterdam, Rotterdam, Netherlands,* [d]*zhou@ese.eur.nl*
[5]*Tinbergen Institute, Rotterdam, Netherlands*

We consider the problem of supervised dimension reduction with a particular focus on extreme values of the target $Y \in \mathbb{R}$ to be explained by a covariate vector $X \in \mathbb{R}^p$. The general purpose is to define and estimate a projection on a lower dimensional subspace of the covariate space which is sufficient for predicting exceedances of the target above high thresholds. We propose an original definition of Tail Conditional Independence which matches this purpose. Inspired by Sliced Inverse Regression (SIR) methods, we develop a novel framework (TIREX, Tail Inverse Regression for EXtreme response) in order to estimate an extreme sufficient dimension reduction (SDR) space of potentially smaller dimension than that of a classical SDR space. We prove the weak convergence of tail empirical processes involved in the estimation procedure and we illustrate the relevance of the proposed approach on simulated and real world data.

Keywords: Dimension reduction; empirical processes; extreme events; inverse regression; supervised learning

References

Aghbalou, A., Portier, F., Sabourin, A. and Zhou, C. (2024). Supplement to “Tail Inverse Regression: dimension reduction for prediction of extremes.” https://doi.org/10.3150/23-BEJ1606SUPP

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

Babichev, D. and Bach, F. (2018). Slice inverse regression with score functions. *Electron. J. Stat.* **12** 1507–1543. MR3804844 https://doi.org/10.1214/18-EJS1428

Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2006). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2108013 https://doi.org/10.1002/0470012382

Bengio, Y., Paiement, J.-F., Vincent, P., Delalleau, O., Roux, N. and Ouimet, M. (2003). Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering. In *NeurIPS Proceedings* **16**. Cambridge: MIT Press.

Bousebata, M., Enjolras, G. and Girard, S. (2023). Extreme partial least-squares. *J. Multivariate Anal.* **194** Paper No. 105101, 21 pp. MR4524584 https://doi.org/10.1016/j.jmva.2022.105101

Bryc, W. (2012). *The Normal Distribution: Characterizations with Applications*. *Lecture Notes in Statistics* **100**. New York: Springer. MR1335228 https://doi.org/10.1007/978-1-4612-2560-7

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electron. J. Stat.* **9** 383–418. MR3323204 https://doi.org/10.1214/15-EJS1002

Chiapino, M. and Sabourin, A. (2016). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *NFMCP Workshop Proceedings* 132–147. Berlin: Springer.

Chiapino, M., Sabourin, A. and Segers, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes* **22** 193–222. MR3949044 https://doi.org/10.1007/s10687-018-0339-3

1350-7265

Chiapino, M., Clémençon, S., Feuillard, V. and Sabourin, A. (2020). A multivariate extreme value theory approach to anomaly clustering and visualization. *Comput. Statist.* **35** 607–628. MR4102945 https://doi.org/10.1007/s00180-019-00913-y

Chojnacki, W. and Brooks, M.J. (2009). A note on the locally linear embedding algorithm. *Int. J. Pattern Recognit. Artif. Intell.* **23** 1739–1752.

Christou, E. (2020). Central quantile subspace. *Stat. Comput.* **30** 677–695. MR4065226 https://doi.org/10.1007/s11222-019-09915-8

Constantinou, P. and Dawid, A.P. (2017). Extended conditional independence and applications in causal inference. *Ann. Statist.* **45** 2618–2653. MR3737904 https://doi.org/10.1214/16-AOS1537

Cook, R.D. (2009). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1645673 https://doi.org/10.1002/9780470316931

Cook, R.D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895 https://doi.org/10.1214/aos/1021379861

Cook, R.D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *J. Amer. Statist. Assoc.* **100** 410–428. MR2160547 https://doi.org/10.1198/016214504000001501

Cook, R.D. and Weisberg, S. (1991). Sliced inverse regression for dimension reduction: Comment. *J. Amer. Statist. Assoc.* **86** 328–332.

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

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85. MR0815960

Dalalyan, A.S., Juditsky, A. and Spokoiny, V. (2008). A new algorithm for estimating the effective dimension-reduction subspace. *J. Mach. Learn. Res.* **9** 1648–1678. MR2438819

Dawid, A.P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31. MR0535541

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3

Delecroix, M., Hristache, M. and Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference* **136** 730–769. MR2181975 https://doi.org/10.1016/j.jspi.2004.09.006

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

Eaton, M.L. (1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20** 272–276. MR0866075 https://doi.org/10.1016/0047-259X(86)90083-7

Einmahl, J.H.J. and Mason, D.M. (1988). Strong limit theorems for weighted quantile processes. *Ann. Probab.* **16** 1623–1643. MR0958207

Elgammal, A. and Lee, C.-S. (2004). Inferring 3D body pose from silhouettes using activity manifold learning. In *CVPR'04 Proceedings* 681–688. Los Alamitos: IEEE Computer Society.

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

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

Fama, E.F. and French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **33** 3–56.

Fama, E.F. and French, K.R. (2015). A five-factor asset pricing model. *J. Financ. Econ.* **116** 1–22.

Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860. MR2093613 https://doi.org/10.3150/bj/1099579158

Fukumizu, K., Bach, F.R. and Jordan, M.I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *J. Mach. Learn. Res.* **5** 73–99. MR2247974 https://doi.org/10.1162/153244303768966111

Fukumizu, K., Bach, F.R. and Jordan, M.I. (2009). Kernel dimension reduction in regression. *Ann. Statist.* **37** 1871–1905. MR2533474 https://doi.org/10.1214/08-AOS637

Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation. *Extremes* **21** 57–95. MR3764611 https://doi.org/10.1007/s10687-017-0300-x

Goix, N., Sabourin, A. and Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *AISTATS Proceedings* **51** 75–83. PMLR.

Goix, N., Sabourin, A. and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *J. Multivariate Anal.* **161** 12–31. MR3698112 https://doi.org/10.1016/j.jmva.2017.06.010

Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. MR1232523 https://doi.org/10.1214/aos/1176349155

Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. MR1134488

Hitz, A. and Evans, R. (2016). One-component regular variation and graphical modeling of extremes. *J. Appl. Probab.* **53** 733–746. MR3570091 https://doi.org/10.1017/jpr.2016.37

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

Hristache, M., Juditsky, A., Polzehl, J. and Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29** 1537–1566. MR1891738 https://doi.org/10.1214/aos/1015345954

Janßen, A. and Wan, P. (2020). k-means clustering of extremes. *Electron. J. Stat.* **14** 1211–1233. MR4071364 https://doi.org/10.1214/20-EJS1689

Jenatton, R., Audibert, J.-Y. and Bach, F. (2011). Structured variable selection with sparsity-inducing norms. *J. Mach. Learn. Res.* **12** 2777–2824. MR2854347

Jessen, A.H. and Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math. (Beograd) (N.S.)* **80** 171–192. MR2281913 https://doi.org/10.2298/PIM0694171J

Jiang, Y., Cooley, D. and Wehner, M.F. (2020). Principal component analysis for extremes and application to US precipitation. *J. Climate* **33** 6441–6451.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

Li, L. and Yin, X. (2008). Sliced inverse regression with regularizations. *Biometrics* **64** 124–131, 323. MR2422826 https://doi.org/10.1111/j.1541-0420.2007.00836.x

Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. (2013). On the statistical consistency of algorithms for binary classification under class imbalance. In *ICML Proceedings* 603–611. PMLR.

Meyer, N. and Wintenberger, O. (2021). Sparse regular variation. *Adv. in Appl. Probab.* **53** 1115–1148. MR4342579 https://doi.org/10.1017/apr.2021.14

Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

Portier, F. (2016). An empirical process view of inverse regression. *Scand. J. Stat.* **43** 827–844. MR3543325 https://doi.org/10.1111/sjos.12209

Portier, F. and Delyon, B. (2013). Optimal transformation: A new approach for covering the central subspace. *J. Multivariate Anal.* **115** 84–107. MR3004547 https://doi.org/10.1016/j.jmva.2012.09.001

Portier, F. and Delyon, B. (2014). Bootstrap testing of the rank of a matrix via least-squared constrained estimation. *J. Amer. Statist. Assoc.* **109** 160–172. MR3180554 https://doi.org/10.1080/01621459.2013.847841

Resnick, S. (2003). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5** 303–336. MR2002121 https://doi.org/10.1023/A:1025148622954

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

Roweis, S.T. and Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** 2323–2326. https://doi.org/10.1126/science.290.5500.2323

Simpson, E.S., Wadsworth, J.L. and Tawn, J.A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika* **107** 513–532. MR4138974 https://doi.org/10.1093/biomet/asaa018

Tenenbaum, J.B., de Silva, V. and Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

Thompson, B. (1984). *Canonical Correlation Analysis: Uses and Interpretation*. *Quantitative Applications in the Social Sciences* **47**. Thousand Oaks: Sage.

Van Der Vaart, A.W. and Wellner, J.A. (2013). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Berlin: Springer.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G. and Koudas, N. (2002). Non-linear dimensionality reduction techniques for classification and visualization. In *ACM SIGKDD Proceedings* 645–651.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965)* 391–420. New York: Academic Press. MR0220397

Wu, H.-M. (2008). Kernel sliced inverse regression with applications to classification. *J. Comput. Graph. Statist.* **17** 590–610. MR2528238 https://doi.org/10.1198/106186008X345161

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35** 2654–2690. MR2382662 https://doi.org/10.1214/009053607000000352

Xu, Z., Dan, C., Khim, J. and Ravikumar, P. (2020). Class-weighted classification: Trade-offs and robust approaches. In *ICML Proceedings* 10544–10554. PMLR.

Yeh, Y.-R., Huang, S.-Y. and Lee, Y.-J. (2008). Nonlinear dimension reduction with kernel sliced inverse regression. *IEEE Trans. Knowl. Data Eng.* **21** 1590–1603.

Zhu, L.-P., Zhu, L.-X. and Feng, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105** 1455–1466. MR2796563 https://doi.org/10.1198/jasa.2010.tm09666

Zwald, L. and Blanchard, G. (2005). On the convergence of eigenspaces in kernel principal component analysis. In *NeurIPS Proceedings* **18**.

Bernoulli **30**(1), 2024, 534–553
https://doi.org/10.3150/23-BEJ1607

Non-asymptotic bounds for the ℓ_∞ estimator in linear regression with uniform noise

YUFEI YI[a] and MATEY NEYKOV[b]

Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, USA, [a]*yy544@fb.com,* [b]*mneykov@stat.cmu.edu*

The Chebyshev or ℓ_∞ estimator is an unconventional alternative to the ordinary least squares in solving linear regressions. It is defined as the minimizer of the ℓ_∞ objective function

$$\widehat{\boldsymbol{\beta}} := \operatorname*{argmin}_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\|_\infty.$$

The asymptotic distribution of the Chebyshev estimator under fixed number of covariates was recently studied (Knight (2020)), yet finite-sample guarantees and generalizations to high-dimensional settings remain open. In this paper, we develop non-asymptotic upper bounds on the estimation error $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|$ for a Chebyshev estimator $\widehat{\boldsymbol{\beta}}$, in a regression setting with uniformly distributed noise $\varepsilon_i \sim U([-a,a])$ where a is either known or unknown. With relatively mild assumptions on the (random) design matrix $\mathbf{X}$, we can bound the error rate by C_p/n with high probability, for some constant C_p depending on the dimension p and the law of the design. Furthermore, we illustrate that there exist designs for which the Chebyshev estimator is (nearly) minimax optimal. On the other hand we also argue that there exist designs for which this estimator behaves sub-optimally in terms of the constant C_p's dependence on p. Finally, we show that "Chebyshev's LASSO" has advantages over the regular LASSO in high dimensional situations, provided that the noise is uniform. Specifically, we argue that it achieves a much faster rate of estimation under certain assumptions on the growth rate of the sparsity level and the ambient dimension with respect to the sample size.

Keywords: Chebyshev estimator; Chebyshev's LASSO; Linear Model; Uniform distribution

References

Akçay, H., Hjalmarsson, H. and Ljung, L. (1996). On the choice of norms in system identification. *IEEE Trans. Automat. Control* **41** 1367–1372. MR1409487 https://doi.org/10.1109/9.536512

Alecu, A., Munteanu, A., Cornelis, J.P.H. and Schelkens, P. (2006). Wavelet-based scalable L-infinity-oriented compression. *IEEE Trans. Image Process.* **15** 2499–2512. https://doi.org/10.1109/tip.2006.877416

Appa, G. and Smith, C. (1973). On L_1 and Chebyshev estimation. *Math. Program.* **5** 73–87. MR0329615 https://doi.org/10.1007/BF01580112

Armstrong, R.D. and Kung, D.S. (1980). A dual method for discrete Chebychev curve fitting. *Math. Program.* **19** 186–199. MR0583278 https://doi.org/10.1007/BF01581640

Beck, A. and Eldar, Y.C. (2007). Regularization in regression with bounded noise: A Chebyshev center approach. *SIAM J. Matrix Anal. Appl.* **29** 606–625. MR2318366 https://doi.org/10.1137/060656784

Berenguer-Rico, V., Johansen, S. and Nielsen, B. (2019). Models where the Least Trimmed Squares and Least Median of Squares estimators are maximum likelihood. Available at SSRN 3455870.

Bertsch, G.F., Sabbey, B. and Uusnäkki, M. (2005). Fitting theories of nuclear binding energies. *Phys. Rev. C* **71** 054311.

Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

Brenner, M.J. (2002). Aeroservoelastic model uncertainty bound estimation from flight data. *J. Guid. Control Dyn.* **25** 748–754.

Castillo, E., Castillo, C., Hadi, A.S. and Sarabia, J.M. (2009). Combined regression models. *Comput. Statist.* **24** 37–66. MR2486211 https://doi.org/10.1007/s00180-007-0105-5

Cline, A.K. (1972). Rate of convergence of Lawson's algorithm. *Math. Comp.* **26** 167–176. MR0298872 https://doi.org/10.2307/2004726

Du, J., Cao, S., Hunt, J.H. and Huo, X. (2019). Optimal Shape Control via L_∞ Loss for Composite Fuselage Assembly. ArXiv Preprint. Available at arXiv:1911.03592.

Fournier, N. and Printems, J. (2010). Absolute continuity for some one-dimensional processes. *Bernoulli* **16** 343–360. MR2668905 https://doi.org/10.3150/09-BEJ215

Guédon, O., Krahmer, F., Kümmerle, C., Mendelson, S. and Rauhut, H. (2022). On the geometry of polytopes generated by heavy-tailed random vectors. *Commun. Contemp. Math.* **24** Paper No. 2150056. MR4400194 https://doi.org/10.1142/S0219199721500565

Hand, M. and Sposito, V. (1980). Using the least squares estimator in Chebyshev estimation: Using the least squares estimator in Chebyshev estimation. *Comm. Statist. Simulation Comput.* **9** 43–49.

Ibragimov, I.A. and Has'Minskii, R.Z. (2013). *Statistical Estimation: Asymptotic Theory* **16**. Berlin: Springer.

James, F. (1983a). Fitting tracks in wire chambers using the Chebyshev norm instead of least squares. *Nucl. Instrum. Methods Phys. Res.* **211** 145–152.

James, F. (1983b). Probability, statistics, and associated computing techniques. In *Techniques and Concepts of High-Energy Physics II* 189–231. Springer.

Jaschke, S.R. (1997). Arbitrage bounds for the term structure of interest rates. *Finance Stoch.* **2** 29–40.

Jaschke, S. and Küchler, U. (2001). Coherent risk measures and good-deal bounds. *Finance Stoch.* **5** 181–200. MR1841716 https://doi.org/10.1007/PL00013530

Jurečková, J. and Picek, J. (2009). Minimum risk equivariant estimator in linear regression model. *Statist. Decisions* **27** 37–54. MR2597425 https://doi.org/10.1524/stnd.2009.1018

Knight, K. (2020). On the asymptotic distribution of the L_∞ estimator in linear regression. Technical Report, Mimeo. Available at http://www.utstat.utoronto.ca/keith/home.html.

Koltchinskii, V. and Mendelson, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **23** 12991–13008. MR3431642 https://doi.org/10.1093/imrn/rnv096

Lawson, C.L. (1961). Contribution to the theory of linear least maximum approximation. Ph. D. dissertation, Univ. Calif. MR2939207

Mendelson, S. (2010). Empirical processes with a bounded ψ_1 diameter. *Geom. Funct. Anal.* **20** 988–1027. MR2729283 https://doi.org/10.1007/s00039-010-0084-5

Milanese, M. and Belforte, G. (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators. *IEEE Trans. Automat. Control* **27** 408–414. MR0680105 https://doi.org/10.1109/TAC.1982.1102926

Mourtada, J. (2022). Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *Ann. Statist.* **50** 2157–2178. MR4474486 https://doi.org/10.1214/22-aos2181

Neykov, M. (2023). On the minimax rate of the Gaussian sequence model under bounded convex constraints. *IEEE Trans. Inf. Theory* **69** 1244–1260. MR4564653

Petrov, V.V. (2007). On lower bounds for tail probabilities. *J. Statist. Plann. Inference* **137** 2703–2705. MR2326120 https://doi.org/10.1016/j.jspi.2006.02.015

Qi, C. (2015). Theoretical uncertainties of the Duflo–Zuker shell-model mass formulae. *JPhysG* **42** 045104.

Rademacher, L.A. (2007). Approximating the centroid is hard. In *Proceedings of the Twenty-Third Annual Symposium on Computational Geometry* 302–305.

Robbins, H. and Zhang, C.-H. (1986). Maximum likelihood estimation in regression with uniform errors. In *Adaptive Statistical Procedures and Related Topics (Upton, N.Y., 1985). Institute of Mathematical Statistics Lecture Notes—Monograph Series* **8** 365–385. Hayward, CA: IMS. MR0898260 https://doi.org/10.1214/lnms/1215540313

Rudelson, M. and Vershynin, R. (2010). Non-asymptotic theory of random matrices: Extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III* 1576–1602. New Delhi: Hindustan Book Agency. MR2827856

Schechtman, E. and Schechtman, G. (1986). Estimating the parameters in regression with uniformly distributed errors. *J. Stat. Comput. Simul.* **26** 221–241. MR0969730 https://doi.org/10.1080/00949658608810965

Sielken, R.L. Jr. and Hartley, H.O. (1973). Two linear programming algorithms for unbiased estimation of linear models. *J. Amer. Statist. Assoc.* **68** 639–641. MR0359190

Sklar, M.G. and Armstrong, R.D. (1982). Least absolute value and Chebychev estimation utilizing least squares results. *Math. Program.* **24** 346–352. MR0676950 https://doi.org/10.1007/BF01585115

Srivastava, N. and Vershynin, R. (2013). Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann. Probab.* **41** 3081–3111. MR3127875 https://doi.org/10.1214/12-AOP760

Tong, Y.L. (2012). *The Multivariate Normal Distribution*. *Springer Series in Statistics*. New York: Springer. MR1029032 https://doi.org/10.1007/978-1-4613-9655-0

Tse, D.N.C., Dahleh, M.A. and Tsitsiklis, J.N. (1993). Optimal asymptotic identification under bounded disturbances. *IEEE Trans. Automat. Control* **38** 1176–1190. MR1235248 https://doi.org/10.1109/9.233151

Vershynin, R. (2011). Spectral norm of products of random and deterministic matrices. *Probab. Theory Related Fields* **150** 471–509. MR2824864 https://doi.org/10.1007/s00440-010-0281-z

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.) 210–268. Cambridge: Cambridge Univ. Press. MR2963170

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

Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

Yaskov, P. (2014). Lower bounds on the smallest eigenvalue of a sample covariance matrix. *Electron. Commun. Probab.* **19** no. 83. MR3291620 https://doi.org/10.1214/ECP.v19-3807

Yi, Y. and Neykov, M. (2024). Supplement to "Non-asymptotic bounds for the ℓ_∞ estimator in linear regression with uniform noise." https://doi.org/10.3150/23-BEJ1607SUPP

Yu, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. New York: Springer. MR1462963

Zolghadri, A. and Henry, D. (2004). Minimax statistical models for air pollution time series. Application to ozone time series data measured in Bordeaux. *Environ. Monit. Assess.* **98** 275–294. https://doi.org/10.1023/b:emas.0000038191.42255.7a

Bernoulli **30**(1), 2024, 554–580
https://doi.org/10.3150/23-BEJ1608

Central limit theorems for semi-discrete Wasserstein distances

EUSTASIO DEL BARRIO[1,a], ALBERTO GONZÁLEZ SANZ[2,b] and JEAN-MICHEL LOUBES[2,c]

[1]*IMUVA, Universidad de Valladolid, Spain,* [a]*tasio@eio.uva.es*
[2]*IMT, Université de Toulouse, France,* [b]*alberto.gonzalez.sanz@uva.es,* [c]*loubes@math.univ-toulouse.fr*

We prove a Central Limit Theorem for the empirical optimal transport cost, $\sqrt{\frac{nm}{n+m}}\{\mathcal{T}_c(P_n,Q_m)-\mathcal{T}_c(P,Q)\}$, in the semi-discrete case, i.e when the distribution P is supported in N points, but without assumptions on Q. We show that the asymptotic distribution is the sup of a centered Gaussian process, which is Gaussian under some additional conditions on the probability Q and on the cost. Such results imply the central limit theorem for the p-Wassertein distance, for $p \geq 1$. This means that, for fixed N, the curse of dimensionality is avoided. To better understand the influence of such N, we provide bounds of $E|\mathcal{W}_p^p(P,Q_m)-\mathcal{W}_p^p(P,Q)|$ depending on m and N. Finally, the semi-discrete framework provides a control on the second derivative of the dual formulation, which yields the first central limit theorem for the optimal transport potentials and Laguerre cells. The results are supported by simulations that help to visualize the given limits and bounds. We analyse also the cases where classical bootstrap works.

Keywords: Central limit theorem; Laguerre cells; optimal transport; optimal transport potentials; semi-discrete optimal transport

References

Acosta, G. and Durán, R.G. (2004). An optimal Poincaré inequality in L^1 for convex domains. *Proc. Amer. Math. Soc.* **132** 195–202. MR2021262 https://doi.org/10.1090/S0002-9939-03-07004-7

Bansil, M. and Kitagawa, J. (2022). Quantitative stability in the geometry of semi-discrete optimal transport. *Int. Math. Res. Not. IMRN* **10** 7354–7389. MR4418710 https://doi.org/10.1093/imrn/rnaa355

Bing, X., Bunea, F. and Niles-Weed, J. (2022). The Sketched Wasserstein Distance for mixture distributions. https://doi.org/10.48550/ARXIV.2206.12768

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

Cárcamo, J., Cuevas, A. and Rodríguez, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* **26** 2143–2175. MR4091104 https://doi.org/10.3150/19-BEJ1188

Cordero-Erausquin, D. and Figalli, A. (2019). Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst.* **39** 7101–7112. MR4026183 https://doi.org/10.3934/dcds.2019297

de Lara, L., González Sanz, A., Asher, N. and Loubes, J.-M. (2021). Transport-based counterfactual models. Pre-Print.

del Barrio, E., González Sanz, A. and Hallin, M. (2022). Nonparametric Multiple-Output Center-Outward Quantile Regression. https://doi.org/10.48550/ARXIV.2204.11756

del Barrio, E., González Sanz, A. and Loubes, J.-M. (2024). Supplement to “Central limit theorems for semi-discrete Wasserstein distances.” https://doi.org/10.3150/23-BEJ1608SUPP

del Barrio, E., González-Sanz, A. and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671. MR4147635 https://doi.org/10.1016/j.jmva.2020.104671

1350-7265 © 2024 ISI/BS

del Barrio, E., Gordaliza, P. and Loubes, J.-M. (2019). A central limit theorem for L_p transportation cost on the real line with application to fairness assessment in machine learning. *Inf. Inference* **8** 817–849. MR4045479 https://doi.org/10.1093/imaiai/iaz016

del Barrio, E. and Loubes, J.-M. (2019a). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275

del Barrio, E., Sanz, A.G. and Loubes, J.-M. (2021). Central limit theorems for general transportation costs. *Ann. Henri Poincaré*. To appear.

del Barrio, E., Sanz, A.G., Loubes, J.-M. and Niles-Weed, J. (2022). An improved central limit theorem and fast convergence rates for entropic transportation costs. ArXiv. Available at arXiv:2204.09105.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. https://doi.org/10.1093/restud/rdy049

Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton, NJ: Princeton Univ. Press. MR3586373 https://doi.org/10.1515/9781400883592

Gallouët, T.O. and Mérigot, Q. (2018). A Lagrangian scheme à la Brenier for the incompressible Euler equations. *Found. Comput. Math.* **18** 835–865. MR3833643 https://doi.org/10.1007/s10208-017-9355-y

Gangbo, W. and McCann, R.J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931 https://doi.org/10.1007/BF02392620

Gayraud, N.T., Rakotomamonjy, A. and Clerc, M. (2017). Optimal transport applied to transfer learning for P300 detection. In *BCI 2017-7th Graz Brain-Computer Interface Conference* 6.

Goldfeld, Z., Kato, K., Nietert, S. and Rioux, G. (2022a). Limit distribution theory for smooth p-Wasserstein distances. https://doi.org/10.48550/ARXIV.2203.00159

Goldfeld, Z., Kato, K., Rioux, G. and Sadhu, R. (2022b). Statistical inference with regularized optimal transport. ArXiv Preprint.

Goldfeld, Z., Kato, K., Rioux, G. and Sadhu, R. (2022c). Limit Theorems for Entropic Optimal Transport Maps and the Sinkhorn Divergence. https://doi.org/10.48550/ARXIV.2207.08683

González Sanz, A., Loubes, J.-M. and Niles-Weed, J. (2022). Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences. https://doi.org/10.48550/ARXIV.2207.07427

González-Delgado, J., González Sanz, A., Cortés, J. and Neuvial, P. (2021). Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology.

Gordaliza, P., del Barrio, E., Fabrice, G. and Loubes, J.-M. (2019). Obtaining fairness using optimal transport theory. In *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proc. Mach. Learn. Res.* **97** 2357–2365. Proc. Mach. Learn. Res.

Graf, S. and Luschgy, H. (2000). *Foundations of Quantization for Probability Distributions*. *Lecture Notes in Math.* **1730**. Berlin: Springer. MR1764176 https://doi.org/10.1007/BFb0103945

Hallin, M., del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

Hartmann, V. and Schuhmacher, D. (2020). Semi-discrete optimal transport: A solution procedure for the unsquared Euclidean distance case. *Math. Methods Oper. Res.* **92** 133–163. MR4152920 https://doi.org/10.1007/s00186-020-00703-z

Hundrieser, S., Klatt, M., Staudt, T. and Munk, A. (2022). A unifying approach to distributional limits for empirical optimal transport. ArXiv Preprint.

Hundrieser, S., Staudt, T. and Munk, A. (2022). Empirical optimal transport between different measures adapts to lower complexity. ArXiv Preprint.

Jiang, R., Pacchiano, A., Stepleton, T., Jiang, H. and Chiappa, S. (2020). Wasserstein fair classification. In *Uncertainty in Artificial Intelligence* 862–872. Proc. Mach. Learn. Res.

Kitagawa, J., Mérigot, Q. and Thibert, B. (2019). Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc. (JEMS)* **21** 2603–2651. MR3985609 https://doi.org/10.4171/JEMS/889

Lévy, B., Mohayaee, R. and von Hausegger, S. (2020). A fast semi-discrete optimal transport algorithm for a unique reconstruction of the early universe. Pre-Print.

Manole, T., Balakrishnan, S. and Wasserman, L. (2022). Minimax confidence intervals for the sliced Wasserstein distance. *Electron. J. Stat.* **16** 2252–2345. MR4402565 https://doi.org/10.1214/22-ejs2001

Mena, G. and Niles-Weed, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems* **32**. Curran Associates.

Meyron, J. (2019). Initialization procedures for discrete and semi-discrete optimal transport. *Comput. Aided Design* **115** 13–22. MR3954050 https://doi.org/10.1016/j.cad.2019.05.037

Milman, E. (2007). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43.

Pollard, D. (1982). A central limit theorem for k-means clustering. *Ann. Probab.* **10** 919–926. MR0672292

Pooladian, A.-A., Divol, V. and Niles-Weed, J. (2023). Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. https://doi.org/10.48550/ARXIV.2301.11302

Rathmair, M. (2019). On how Poincaré inequalities imply weighted ones. *Monatsh. Math.* **188** 753–763. MR3928771 https://doi.org/10.1007/s00605-019-01266-w

Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Mathematical Series, No. 28*. Princeton, NJ: Princeton Univ. Press. MR0274683

Rockafellar, R.T. and Wets, R.J.-B. (1998). *Variational Analysis*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Berlin: Springer. MR1491362 https://doi.org/10.1007/978-3-642-02431-3

Sadhu, R., Goldfeld, Z. and Kato, K. (2021). Limit Distribution Theory for the Smooth 1-Wasserstein Distance with Applications. https://doi.org/10.48550/ARXIV.2107.13494

Segers, J. (2022). Graphical and uniform consistency of estimated optimal transport plans. https://doi.org/10.48550/ARXIV.2208.02508

Shen, J., Qu, Y., Zhang, W. and Yu, Y. (2018). Wasserstein distance guided representation learning for domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Sommerfeld, M. and Munk, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238. MR3744719 https://doi.org/10.1111/rssb.12236

Tameling, C., Sommerfeld, M. and Munk, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.* **29** 2744–2781. MR4019874 https://doi.org/10.1214/19-AAP1463

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

Verdinelli, I. and Wasserman, L. (2019). Hybrid Wasserstein distance and fast distribution clustering. *Electron. J. Stat.* **13** 5088–5119. MR4041703 https://doi.org/10.1214/19-EJS1639

Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: Amer. Math. Soc. MR1964483 https://doi.org/10.1090/gsm/058

Villani, C. (2008). *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Berlin: Springer.

Vitale, R.A. (1985). L_p metrics for compact, convex sets. *J. Approx. Theory* **45** 280–287. MR0812757 https://doi.org/10.1016/0021-9045(85)90051-6

Weed, J. and Bach, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

Bernoulli **30**(1), 2024, 581–609
https://doi.org/10.3150/23-BEJ1609

Explicit bounds for spectral theory of geometrically ergodic Markov kernels and applications

LOÏC HERVÉ[a] and JAMES LEDOUX[b]

Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000, France, [a]*Loic.Herve@insa-rennes.fr,* [b]*James.Ledoux@insa-rennes.fr*

In this paper, we deal with a Markov chain on a measurable state space $(\mathbb{X},\mathcal{X})$ which has a transition kernel P admitting an aperiodic small-set S and satisfying the standard geometric-drift condition. Under these assumptions, there exists $\alpha_0 \in (0,1]$ such that $PV^{\alpha_0} \le \delta^{\alpha_0} V^{\alpha_0} + \nu(V^{\alpha_0})1_S$. Hence P is V^{α_0}-geometrically ergodic and its "second eigenvalue" ϱ_{α_0} provides the best rate of convergence. Setting $R := P - \nu(\cdot)1_S$ and $\Gamma := \{\lambda \in \mathbb{C},\ \delta^{\alpha_0} < |\lambda| < 1\}$, ϱ_{α_0} is shown to satisfy, either $\varrho_{\alpha_0} = \max\left\{|\lambda| : \lambda \in \Gamma,\ \sum_{k=1}^{+\infty} \lambda^{-k}\, \nu(R^{k-1}1_S) = 1\right\}$ if this set is not empty, or $\varrho_{\alpha_0} \le \delta^{\alpha_0}$. Actually the set is finite in the first case and is composed by the spectral values of P in Γ. The second case occurs when P has no spectral value in Γ. Moreover, a bound of the operator-norm of $(zI - P)^{-1}$ allows us to derive an explicit formula for the multiplicative constant in the rate of convergence, which can be evaluated provided that any information of the "second eigenvalue" is available. Such numerical computation is carried out for a classical family of reflected random walks. Moreover we obtain a simple and explicit bound of the operator-norm of $(I - P + \pi(\cdot)1_{\mathbb{X}})^{-1}$ involved in the definition of the so-called fundamental solution to Poisson's equation. This allows us to specify the location of the eigenvalues of P and, then, to obtain a general bound on ϱ_{α_0}. The reversible case is also discussed. In particular, the bound of ϱ_{α_0} obtained for positive reversible Markov kernels is the expected one, and numerical illustrations are proposed for the Metropolis-Hastings algorithm and for the Gaussian autoregressive Markov chain. The bound for the operator-norm of $(I - P + \pi(\cdot)1_{\mathbb{X}})^{-1}$ is derived from an estimate, only depending on δ^{α_0}, of the operator-norm of $(I - R)^{-1}$ which provides another way to get a solution to Poisson's equation. This estimate is also shown to be of greatest interest to generalize the error bounds obtained for perturbed discrete and atomic Markov chains in Liu and Li (*Adv. in Appl. Probab.* **50** (2018) 645–669) to the case of general geometrically ergodic Markov chains. These error estimates are the simplest that can be expected in this context. All the estimates in this work are expressed in the standard V^{α_0}-weighted operator norm.

Keywords: Drift conditions; invariant probability measure; perturbed Markov kernels; Poisson's equation; rate of convergence; second eigenvalue; small set

References

Atchadé, Y.F. and Perron, F. (2007). On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics* **41** 77–84. MR2303970 https://doi.org/10.1080/10485250601033214

Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987 https://doi.org/10.1214/105051604000000710

Delyon, B. (2017). Convergence rate of the powers of an operator. Applications to stochastic systems. *Bernoulli* **23** 2129–2180. MR3648028 https://doi.org/10.3150/15-BEJ778

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov Chains*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

Ferré, D., Hervé, L. and Ledoux, J. (2013). Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab.* **50** 184–194. MR3076780 https://doi.org/10.1239/jap/1363784432

Galtchouk, L. and Pergamenshchikov, S. (2014). Geometric ergodicity for classes of homogeneous Markov chains. *Stochastic Process. Appl.* **124** 3362–3391. MR3231623 https://doi.org/10.1016/j.spa.2014.05.002

Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

Hennion, H. (1993). Sur un théorème spectral et son application aux noyaux lipchitziens. *Proc. Amer. Math. Soc.* **118** 627–634. MR1129880 https://doi.org/10.2307/2160348

Hennion, H. (2006). Quasi-compactness and absolutely continuous kernels, applications to Markov chains. Available at https://arxiv.org/abs/math/0606680.

Hennion, H. (2007b). Quasi-compactness and absolutely continuous kernels. *Probab. Theory Related Fields* **139** 451–471. MR2322704 https://doi.org/10.1007/s00440-006-0048-8

Hennion, H. and Hervé, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness*. *Lecture Notes in Math.* **1766**. Berlin: Springer. MR1862393 https://doi.org/10.1007/b87874

Hervé, L. and Ledoux, J. (2014a). Approximating Markov chains and V-geometric ergodicity via weak perturbation theory. *Stochastic Process. Appl.* **124** 613–638. MR3131307 https://doi.org/10.1016/j.spa.2013.09.003

Hervé, L. and Ledoux, J. (2014b). Spectral analysis of Markov kernels and application to the convergence rate of discrete random walks. *Adv. in Appl. Probab.* **46** 1036–1058. MR3290428 https://doi.org/10.1239/aap/1418396242

Hervé, L. and Ledoux, J. (2020). V-geometrical ergodicity of Markov kernels via finite-rank approximations. *Electron. Commun. Probab.* **25** Paper No. 23. MR4089730 https://doi.org/10.1214/20-ecp303

Hervé, L. and Ledoux, J. (2022a). Quantitative approximation of the invariant distribution of a Markov chain. A new approach. Available at https://hal.archives-ouvertes.fr/hal-03605636.

Hervé, L. and Ledoux, J. (2022b). Explicit bounds for spectral theory of geometrically ergodic Markov kernels and applications. Available at https://hal.archives-ouvertes.fr/hal-03819315.

Hervé, L. and Ledoux, J. (2023). Robustness of iterated function systems of Lipschitz maps. *J. Appl. Probab.* **60**. To appear. Available at https://hal.archives-ouvertes.fr/hal-03423198/.

Kartashov, N.V. (1986). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space, Parts I and II. *Theory Probab. Appl.* **30** 247–259, 505–515.

Kemeny, J.G. (1981). Generalization of a fundamental matrix. *Linear Algebra Appl.* **38** 193–206. MR0636037 https://doi.org/10.1016/0024-3795(81)90020-3

Kontoyiannis, I. and Meyn, S.P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13** 304–362. MR1952001 https://doi.org/10.1214/aoap/1042765670

Kontoyiannis, I. and Meyn, S.P. (2005). Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.* **10** 61–123. MR2120240 https://doi.org/10.1214/EJP.v10-231

Liu, Y. and Li, W. (2018). Error bounds for augmented truncation approximations of Markov chains via the perturbation method. *Adv. in Appl. Probab.* **50** 645–669. MR3832888 https://doi.org/10.1017/apr.2018.28

Medina-Aguayo, F., Rudolf, D. and Schweizer, N. (2020). Perturbation bounds for Monte Carlo within Metropolis via restricted approximations. *Stochastic Process. Appl.* **130** 2200–2227. MR4074698 https://doi.org/10.1016/j.spa.2019.06.015

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. London: Springer London, Ltd. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

Meyn, S.P. and Tweedie, R.L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770

Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2509253 https://doi.org/10.1017/CBO9780511626630

Qin, Q. and Hobert, J.P. (2021). On the limitations of single-step drift and minorization in Markov chain convergence analysis. *Ann. Appl. Probab.* **31** 1633–1659. MR4312841 https://doi.org/10.1214/20-aap1628

Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981

Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565 https://doi.org/10.1214/154957804100000024

Rudolf, D. and Schweizer, N. (2018). Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli* **24** 2610–2639. MR3779696 https://doi.org/10.3150/17-BEJ938

Shardlow, T. and Stuart, A.M. (2000). A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.* **37** 1120–1137. MR1756418 https://doi.org/10.1137/S0036142998337235

Tweedie, R.L. (1998). Truncation approximations of invariant measures for Markov chains. *J. Appl. Probab.* **35** 517–536. MR1659488 https://doi.org/10.1239/jap/1032265201

Wu, L. (2004). Essential spectral radius for Markov semigroups. I. Discrete time case. *Probab. Theory Related Fields* **128** 255–321. MR2031227 https://doi.org/10.1007/s00440-003-0304-0

Bernoulli **30**(1), 2024, 610–635
https://doi.org/10.3150/23-BEJ1610

Minimax estimation of low-rank quantum states and their linear functionals

SAMRIDDHA LAHIRY[1,a] and MICHAEL NUSSBAUM[2,b]

[1]*Department of Statistics, Science Center, Harvard University, Cambridge, MA 02138, USA,*
[a]*slahiry@fas.harvard.edu*
[2]*Department of Mathematics, Malott Hall, Cornell University, Ithaca, NY 14853, USA,*
[b]*nussbaum@math.cornell.edu*

In classical statistics, a well known paradigm consists in establishing asymptotic equivalence between an experiment of i.i.d. observations and a Gaussian shift experiment, with the aim of obtaining optimal estimators in the former complicated model from the latter simpler model. In particular, a statistical experiment consisting of n i.i.d. observations from d-dimensional multinomial distributions can be well approximated by an experiment consisting of $d-1$ dimensional Gaussian distributions. In a quantum version of the result, it has been shown that a collection of n qudits (d-dimensional quantum states) of full rank can be well approximated by a quantum system containing a classical part, which is a $d-1$ dimensional Gaussian distribution, and a quantum part containing an ensemble of $d(d-1)/2$ shifted thermal states. In this paper, we obtain a generalization of this result when the qudits are not of full rank. We show that when the rank of the qudits is r, then the limiting experiment consists of an $r-1$ dimensional Gaussian distribution and an ensemble of both shifted pure and shifted thermal states. For estimation purposes, we establish an asymptotic minimax result in the limiting Gaussian model. Analogous results are then obtained for estimation of a low rank qudit from an ensemble of identically prepared, independent quantum systems, using the local asymptotic equivalence result. We also consider the problem of estimation of a linear functional of the quantum state. We construct an estimator for the functional, analyze the risk and use quantum local asymptotic equivalence to show that our estimator is also optimal in the minimax sense.

Keywords: Functional estimation; low rank states; quantum local asymptotic normality; quantum minimax estimation

References

[1] Acharya, A. and Guţă, M. (2017). Statistical analysis of compressive low rank tomography with random measurements. *J. Phys. A* **50** 195301, 21 pp. MR3635730 https://doi.org/10.1088/1751-8121/aa682e
[2] Acharya, A., Kypraios, T. and Guţă, M. (2016). Statistically efficient tomography of low rank states with incomplete measurements. *New J. Phys.* **18** 043018. https://doi.org/10.1088/1367-2630/18/4/043018
[3] Acharya, A., Kypraios, T. and Guţă, M. (2019). A comparative study of estimation methods in quantum tomography. *J. Phys. A* **52** 234001, 36 pp. MR3953819 https://doi.org/10.1088/1751-8121/ab1958
[4] Acharya, J., Issa, I., Shende, N.V. and Wagner, A.B. (2020). Estimating quantum entropy. *IEEE J. Sel. Areas Inf. Theory* **1** 454–468. https://doi.org/10.1109/JSAIT.2020.3015235
[5] Alquier, P., Butucea, C., Hebiri, M., Meziani, K. and Morimae, T. (2013). Rank-penalized estimation of a quantum system. *Phys. Rev. A* **88** 032113. https://doi.org/10.1103/PhysRevA.88.032113
[6] Audenaert, K.M.R., Calsamiglia, J., Muñoz Tapia, R., Bagan, E., Masanes, L., Acin, A. and Verstraete, F. (2007). Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501. https://doi.org/10.1103/PhysRevLett.98.160501
[7] Audenaert, K.M.R., Nussbaum, M., Szkoła, A. and Verstraete, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** 251–283. MR2377635 https://doi.org/10.1007/s00220-008-0417-5
[8] Barndorff-Nielsen, O.E., Gill, R.D. and Jupp, P.E. (2003). On quantum statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 775–816. MR2017871 https://doi.org/10.1111/1467-9868.00415

1350-7265

[9] Braunstein, S.L. and Caves, C.M. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72** 3439–3443. MR1274631 https://doi.org/10.1103/PhysRevLett.72.3439
[10] Brown, L.D. and Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958 https://doi.org/10.1214/aos/1032181159
[11] Butucea, C., Guţă, M. and Kypraios, T. (2015). Spectral thresholding quantum tomography for low rank states. *New J. Phys.* **17** 113050. https://doi.org/10.1088/1367-2630/17/11/113050
[12] Butucea, C., Guţă, M. and Nussbaum, M. (2018). Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise. *Ann. Statist.* **46** 3676–3706. MR3852665 https://doi.org/10.1214/17-AOS1672
[13] Butucea, C., Guţă, M. and Nussbaum, M. (2018). Supplementary material to “Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise.” Ann. Statist. **46**. https://doi.org/10.1214/17-AOS1672SUPP
[14] Cai, T., Kim, D., Wang, Y., Yuan, M. and Zhou, H.H. (2016). Optimal large-scale quantum state tomography with Pauli measurements. *Ann. Statist.* **44** 682–712. MR3476614 https://doi.org/10.1214/15-AOS1382
[15] Carpentier, A., Eisert, J., Gross, D. and Nickl, R. (2019). Uncertainty quantification for matrix compressed sensing and quantum tomography problems. In *High Dimensional Probability VIII – The Oaxaca Volume*. *Progress in Probability* **74** 385–430. Cham: Birkhäuser/Springer. MR4181374 https://doi.org/10.1007/978-3-030-26391-1_18
[16] Carter, A.V. (2002). Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.* **30** 708–730. MR1922539 https://doi.org/10.1214/aos/1028674839
[17] Flammia, S.T., Gross, D., Liu, Y.-K. and Eisert, J. (2012). Quantum tomography via compressed sensing: Error bounds, sample complexity, and efficient estimators. *New J. Phys.* **14** 095022. https://doi.org/10.1088/1367-2630/14/9/095022
[18] Fulton, W. and Harris, J. (1991). *Representation Theory: A First Course, Readings in Mathematics*. *Graduate Texts in Mathematics* **129**. New York: Springer. MR1153249 https://doi.org/10.1007/978-1-4612-0979-9
[19] Gill, R.D. and Guţă, M.I. (2013). On asymptotic quantum statistical inference. In *From Probability to Statistics and Back: High-Dimensional Models and Processes*. *Inst. Math. Stat. (IMS) Collect.* **9** 105–127. Beachwood, OH: IMS. MR3186752 https://doi.org/10.1214/12-IMSCOLL909
[20] Gross, D., Liu, Y.-K., Flammia, S.T., Becker, S. and Eisert, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401. https://doi.org/10.1103/PhysRevLett.105.150401
[21] Guţă, M. and Butucea, C. (2010). Quantum U-statistics. *J. Math. Phys.* **51** 102202, 24 pp. MR2761295 https://doi.org/10.1063/1.3476776
[22] Guţă, M., Janssens, B. and Kahn, J. (2008). Optimal estimation of qubit states with continuous time measurements. *Comm. Math. Phys.* **277** 127–160. MR2357428 https://doi.org/10.1007/s00220-007-0357-5
[23] Guţă, M. and Jenčová, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. MR2346393 https://doi.org/10.1007/s00220-007-0340-1
[24] Guţă, M. and Kahn, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A (3)* **73** 052108, 15 pp. MR2229156 https://doi.org/10.1103/PhysRevA.73.052108
[25] Guţă, M., Kahn, J., Kueng, R. and Tropp, J.A. (2020). Fast state tomography with optimal error bounds. *J. Phys. A* **53** 204001, 28 pp. MR4093475 https://doi.org/10.1088/1751-8121/ab8111
[26] Hayashi, M. (2017). *Quantum Information Theory: Mathematical Foundation*, 2nd ed. *Graduate Texts in Physics*. Berlin: Springer. MR3558531 https://doi.org/10.1007/978-3-662-49725-8
[27] Helstrom, C.W. (1976). *Quantum Detection and Estimation Theory*. New York: Academic Press.
[28] Holevo, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. *Quaderni/Monographs* **1**. Pisa: Edizioni della Normale. MR2797301 https://doi.org/10.1007/978-88-7642-378-9
[29] Holevo, A.S. (1973). Optimal quantum measurements. *Theoret. Math. Phys.* **17** 1172–1177. https://doi.org/10.1007/BF01037594
[30] Holevo, A.S. (1978). *Investigations in the General Theory of Statistical Decisions*. Providence, RI: Amer. Math. Soc. MR0652516
[31] Holevo, A.S. (2001). *Statistical Structure of Quantum Theory*. *Lecture Notes in Physics. Monographs* **67**. Berlin: Springer. MR1889193 https://doi.org/10.1007/3-540-44998-1
[32] Ibragimov, I.A. and Has’minskiĭ, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. *Applications of Mathematics* **16**. New York–Berlin: Springer. MR0620321

[33] Kahn, J. and Guţă, M. (2008). Local asymptotic normality and optimal estimation for d-dimensional quantum systems. In *Quantum Stochastics and Information: Statistics, Filtering and Control* 300–322. Hackensack, NJ: World Sci. Publ. MR2562438 https://doi.org/10.1142/9789812832962_0014
[34] Kahn, J. and Guţă, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764 https://doi.org/10.1007/s00220-009-0787-3
[35] Koltchinskii, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. MR3012397 https://doi.org/10.1214/11-AOS926
[36] Koltchinskii, V. and Xia, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798
[37] Koševnik, J.A. and Levit, B.J. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753. https://doi.org/10.1137/1121087
[38] Lahiry, S. and Nussbaum, M. (2024). Supplement to "Minimax estimation of low-rank quantum states and their linear functionals." https://doi.org/10.3150/23-BEJ1610SUPP
[39] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. New York: Springer. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7
[40] Li, K. (2016). Discriminating quantum states: The multiple Chernoff distance. *Ann. Statist.* **44** 1661–1679. MR3519936 https://doi.org/10.1214/16-AOS1436
[41] Liu, Y.-K. (2011). Universal low-rank matrix recovery from Pauli measurements. *Adv. Neural Inf. Process. Syst.* **24**.
[42] 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
[43] Nussbaum, M. and Szkoła, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. MR2502660 https://doi.org/10.1214/08-AOS593
[44] Ogawa, T. and Nagaoka, H. (2000). Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inf. Theory* **46** 2428–2433. MR1806811 https://doi.org/10.1109/18.887855
[45] Parthasarathy, K.R. (1992). *An Introduction to Quantum Stochastic Calculus*. *Modern Birkhäuser Classics*. Basel: Birkhäuser/Springer Basel AG. MR3012668
[46] Petz, D. (2008). *Quantum Information Theory and Quantum Statistics*. *Theoretical and Mathematical Physics*. Berlin: Springer. MR2363070
[47] Shiryaev, A.N. and Spokoiny, V.G. (2000). *Statistical Experiments and Decisions: Asymptotic Theory*. *Advanced Series on Statistical Science & Applied Probability* **8**. River Edge, NJ: World Scientific Co., Inc. MR1791434 https://doi.org/10.1142/9789812779243
[48] Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. *De Gruyter Studies in Mathematics* **7**. Berlin: de Gruyter. MR0812467 https://doi.org/10.1515/9783110850826
[49] Tsang, M., Albarelli, F. and Datta, A. (2020). Quantum semiparametric estimation. *Phys. Rev. X* **10** 031023. https://doi.org/10.1103/PhysRevX.10.031023
[50] Wang, Y. (2011). Quantum Monte Carlo simulation. *Ann. Appl. Stat.* **5** 669–683. MR2840170 https://doi.org/10.1214/10-AOAS406
[51] Wang, Y. (2012). Quantum computation and quantum information. *Statist. Sci.* **27** 373–394. MR3012432 https://doi.org/10.1214/11-STS378
[52] Wang, Y. and Song, X. (2020). Quantum science and quantum technology. *Statist. Sci.* **35** 51–74. MR4071358 https://doi.org/10.1214/19-STS745
[53] Wang, Y., Wu, S. and Zou, J. (2016). Quantum annealing with Markov chain Monte Carlo simulations and D-wave quantum computers. *Statist. Sci.* **31** 362–398. MR3552740 https://doi.org/10.1214/16-STS560
[54] Yamagata, K., Fujiwara, A. and Gill, R.D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** 2197–2217. MR3127863 https://doi.org/10.1214/13-AOS1147

Bernoulli **30**(1), 2024, 636–665
https://doi.org/10.3150/23-BEJ1611

Linear and nonlinear signal detection and estimation in high-dimensional nonparametric regression under weak sparsity

KIN YAP CHEUNG[a], STEPHEN M.S. LEE and XIAOYA XU[c]

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong,
[a]*u3507181@connect.hku.hk,* [b]*smslee@hku.hk,* [c]*xuxiaoya@connect.hku.hk*

The partially linear model provides an effective tool to combat the curse of dimensionality in nonparametric regression. Its applicability is, however, compromised by the need for correct distinction between linear and nonlinear components. Existing solutions are restricted to low dimensions or regression functions endowed with special structures. This paper considers a general nonparametric regression framework under which signal strength is embedded in a continuous spectrum scaled by asymptotic orders. Under a weak sparsity condition which allows for the presence of many weak, non-detectable, signals, a novel penalised regression procedure is proposed for detection and estimation of strong linear and nonlinear signals under high dimensions. The procedure applies bandwidth regularisation and SCAD penalisation to select nonlinear and linear signals, respectively, under a partially linear model setting. Theoretical results are established for its consistency in detecting strong signals and its error rate in estimating the regression function. Numerical examples are presented to illustrate its performance.

Keywords: High dimensions; local linear regression; partially linear model; SCAD; variable selection; weak sparsity

References

Avramidis, P. (2005). Two-step cross-validation selection method for partially linear models. *Statist. Sinica* **15** 1033–1048. MR2234412

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

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inf. Theory* **57** 5467–5484. MR2849368 https://doi.org/10.1109/TIT.2011.2158486

Giordano, F., Lahiri, S.N. and Parrella, M.L. (2020). GRID: A variable selection and structure discovery method for high dimensional nonparametric regression. *Ann. Statist.* **48** 1848–1874. MR4124346 https://doi.org/10.1214/19-AOS1846

Hamilton, S.A. and Truong, Y.K. (1997). Local linear estimation in partly linear models. *J. Multivariate Anal.* **60** 1–19. MR1441456 https://doi.org/10.1006/jmva.1996.1642

He, X. and Wang, J. (2020). Discovering model structure for partially linear models. *Ann. Inst. Statist. Math.* **72** 45–63. MR4052650 https://doi.org/10.1007/s10463-018-0682-9

Horowitz, J.L. and Huang, J. (2013). Penalized estimation of high-dimensional models under a generalized sparsity condition. *Statist. Sinica* **23** 725–748. MR3086653

Huang, J., Wei, F. and Ma, S. (2012). Semiparametric regression pursuit. *Statist. Sinica* **22** 1403–1426. MR3027093

Lafferty, J. and Wasserman, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *Ann. Statist.* **36** 28–63. MR2387963 https://doi.org/10.1214/009053607000000811

Lian, H., Liang, H. and Ruppert, D. (2015). Separation of covariates into nonparametric and parametric parts in high-dimensional partially linear additive models. *Statist. Sinica* **25** 591–607. MR3379090

Liang, H. (2006). Estimation in partially linear models and numerical comparisons. *Comput. Statist. Data Anal.* **50** 675–687. MR2207001 https://doi.org/10.1016/j.csda.2004.10.007

Liu, H., Xu, X. and Li, J.J. (2020). A bootstrap lasso + partial ridge method to construct confidence intervals for parameters in high-dimensional sparse linear models. *Statist. Sinica* **30** 1333–1355. MR4257536 https://doi.org/10.5705/ss.202018.0131
Lou, Y., Bien, J., Caruana, R. and Gehrke, J. (2016). Sparse partially linear additive models. *J. Comput. Graph. Statist.* **25** 1026–1040. MR3572032 https://doi.org/10.1080/10618600.2015.1089775
Müller, S. and Vial, C. (2009). Partially linear model selection by the bootstrap. *Aust. N. Z. J. Stat.* **51** 183–200. MR2531986 https://doi.org/10.1111/j.1467-842X.2009.00540.x
Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979 https://doi.org/10.1214/aos/1176325632
Samarov, A., Spokoiny, V. and Vial, C. (2005). Component identification and estimation in nonlinear high-dimensional regression models by structural adaptation. *J. Amer. Statist. Assoc.* **100** 429–445. MR2160548 https://doi.org/10.1198/016214504000001529
Shao, J. and Deng, X. (2012). Estimation in high-dimensional linear models with deterministic design matrices. *Ann. Statist.* **40** 812–831. MR2933667 https://doi.org/10.1214/12-AOS982
Shi, P. and Qu, A. (2017). Weak signal identification and inference in penalized model selection. *Ann. Statist.* **45** 1214–1253. MR3662453 https://doi.org/10.1214/16-AOS1482
Tan, Z. and Zhang, C.-H. (2019). Doubly penalized estimation in additive regression with high-dimensional data. *Ann. Statist.* **47** 2567–2600. MR3988766 https://doi.org/10.1214/18-AOS1757
White, K.R., Stefanski, L.A. and Wu, Y. (2017). Variable selection in kernel regression using measurement error selection likelihoods. *J. Amer. Statist. Assoc.* **112** 1587–1597. MR3750883 https://doi.org/10.1080/01621459.2016.1222287
Xia, Y. (2008). A multiple-index model and dimension reduction. *J. Amer. Statist. Assoc.* **103** 1631–1640. MR2504209 https://doi.org/10.1198/016214508000000805
Xia, Y., Tong, H., Li, W.K. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. MR1924297 https://doi.org/10.1111/1467-9868.03411
Yang, L., Fang, Y., Wang, J. and Shao, Y. (2017). Variable selection for partially linear models via learning gradients. *Electron. J. Stat.* **11** 2907–2930. MR3694572 https://doi.org/10.1214/17-EJS1300
Zhang, H.H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106** 1099–1112. MR2894767 https://doi.org/10.1198/jasa.2011.tm10281
Zhao, T., Liu, H. and Zhang, T. (2018). Pathwise coordinate optimization for sparse learning: Algorithm and theory. *Ann. Statist.* **46** 180–218. MR3766950 https://doi.org/10.1214/17-AOS1547

Bernoulli **30**(1), 2024, 666–682
https://doi.org/10.3150/23-BEJ1612

Normality of smooth statistics for planar determinantal point processes

ANTTI HAIMI[1,a] and JOSÉ LUIS ROMERO[1,2,b]

[1] *Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090, Vienna, Austria,*
[a] *antti.haimi@univie.ac.at*
[2] *Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14 A-1040, Vienna, Austria,*
[b] *jose.luis.romero@univie.ac.at*

We consider smooth linear statistics of determinantal point processes on the complex plane, and their large scale asymptotics. We prove asymptotic normality in the finite variance case, where Soshnikov's theorem is not applicable. The setting is similar to that of Rider and Virág [Electron. J. Probab., 12, no. 45, 1238–1257, (2007)] for the complex plane, but replaces analyticity conditions by the assumption that the correlation kernel is reproducing. Our proof is a streamlined version of that of Ameur, Hedenmalm and Makarov [Duke Math J., 159, 31–81, (2011)] for eigenvalues of normal random matrices. In our case, the reproducing property is brought to bear to compensate for the lack of analyticity and radial symmetries.

Keywords: Asymptotic normality; determinantal point process; linear statistics; Weyl-Heisenberg DPP

References

[1] Abreu, L.D., Pereira, J.M., Romero, J.L. and Torquato, S. (2017). The Weyl-Heisenberg ensemble: Hyperuniformity and higher Landau levels. *J. Stat. Mech. Theory Exp.* **4** 043103. MR3652713 https://doi.org/10.1088/1742-5468/aa68a7

[2] Ameur, Y., Hedenmalm, H. and Makarov, N. (2010). Berezin transform in polynomial Bergman spaces. *Comm. Pure Appl. Math.* **63** 1533–1584. MR2742007 https://doi.org/10.1002/cpa.20329

[3] Ameur, Y., Hedenmalm, H. and Makarov, N. (2011). Fluctuations of eigenvalues of random normal matrices. *Duke Math. J.* **159** 31–81. MR2817648 https://doi.org/10.1215/00127094-1384782

[4] Ameur, Y., Hedenmalm, H. and Makarov, N. (2015). Random normal matrices and Ward identities. *Ann. Probab.* **43** 1157–1201. MR3342661 https://doi.org/10.1214/13-AOP885

[5] Berman, R.J. (2018). Determinantal point processes and fermions on polarized complex manifolds: Bulk universality. In *Algebraic and Analytic Microlocal Analysis*. *Springer Proc. Math. Stat.* **269** 341–393. Cham: Springer. MR3903320 https://doi.org/10.1007/978-3-030-01588-6_5

[6] Borodin, A. (2011). Determinantal point processes. In *The Oxford Handbook of Random Matrix Theory* 231–249. Oxford: Oxford Univ. Press. MR2932631

[7] Costin, O. and Lebowitz, J.L. (1995). Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75** 69–72. MR3155254 https://doi.org/10.1103/PhysRevLett.75.69

[8] Dinh, T.-C., Ghosh, S., Tran, H.-S. and Tran, M.-H. Quantitative Marcinkiewicz's theorem and central limit theorems: Applications to spin systems and point processes. ArXiv Preprint. Available at arXiv:2107.08469v2.

[9] Fenzl, M. and Lambert, G. (2022). Precise deviations for disk counting statistics of invariant determinantal processes. *Int. Math. Res. Not. IMRN* **10** 7420–7494. MR4418712 https://doi.org/10.1093/imrn/rnaa341

[10] Folland, G.B. (1989). *Harmonic Analysis in Phase Space*. *Annals of Mathematics Studies* **122**. Princeton, NJ: Princeton Univ. Press. MR0983366 https://doi.org/10.1515/9781400882427

[11] Ginibre, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.* **6** 440–449. MR0173726 https://doi.org/10.1063/1.1704292

[12] Haimi, A. (2014). Bulk asymptotics for polyanalytic correlation kernels. *J. Funct. Anal.* **266** 3083–3133. MR3158718 https://doi.org/10.1016/j.jfa.2013.11.021
[13] Haimi, A. and Hedenmalm, H. (2013). The polyanalytic Ginibre ensembles. *J. Stat. Phys.* **153** 10–47. MR3100813 https://doi.org/10.1007/s10955-013-0813-x
[14] Haimi, A. and Wennman, A. (2019). A central limit theorem for fluctuations in polyanalytic Ginibre ensembles. *Int. Math. Res. Not. IMRN* **5** 1350–1372. MR3920350 https://doi.org/10.1093/imrn/rnx147
[15] Hough, J.B., Krishnapur, M., Peres, Y. and Virág, B. (2009). *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. *University Lecture Series* **51**. Providence, RI: Amer. Math. Soc. MR2552864 https://doi.org/10.1090/ulect/051
[16] Johansson, K. (2006). Random matrices and determinantal processes. In *Mathematical Statistical Physics* 1–55. Amsterdam: Elsevier. MR2581882 https://doi.org/10.1016/S0924-8099(06)80038-7
[17] Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab.* **7** 83–122. MR0380979 https://doi.org/10.2307/1425855
[18] Matsui, T., Katori, M. and Shirai, T. (2021). Local number variances and hyperuniformity of the Heisenberg family of determinantal point processes. *J. Phys. A* **54** Paper No. 165201. MR4256659
[19] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus*. *Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. From Stein's method to universality. MR2962301 https://doi.org/10.1017/CBO9781139084659
[20] Rider, B. and Virág, B. (2007). Complex determinantal processes and H^1 noise. *Electron. J. Probab.* **12** 1238–1257. MR2346510 https://doi.org/10.1214/EJP.v12-446
[21] Rider, B. and Virág, B. (2007). The noise in the circular law and the Gaussian free field. *Int. Math. Res. Not. IMRN* **2** Art. ID rnm006. MR2361453 https://doi.org/10.1093/imrn/rnm006
[22] Shirai, T. (2015). Ginibre-type point processes and their asymptotic behavior. *J. Math. Soc. Japan* **67** 763–787. MR3340195 https://doi.org/10.2969/jmsj/06720763
[23] Simon, B. (2005). *Trace Ideals and Their Applications*, 2nd ed. *Mathematical Surveys and Monographs* **120**. Providence, RI: Amer. Math. Soc. MR2154153 https://doi.org/10.1090/surv/120
[24] Soshnikov, A. (2000). Determinantal random point fields. *Uspekhi Mat. Nauk* **55** 107–160. MR1799012 https://doi.org/10.1070/rm2000v055n05ABEH000321
[25] Soshnikov, A. (2002). Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** 171–187. MR1894104 https://doi.org/10.1214/aop/1020107764
[26] Vasilevski, N.L. (2000). Poly-Fock spaces. In *Differential Operators and Related Topics, Vol. I (Odessa, 1997)*. *Oper. Theory Adv. Appl.* **117** 371–386. Basel: Birkhäuser. MR1764974

Bernoulli **30**(1), 2024, 683–711
https://doi.org/10.3150/23-BEJ1613

Dimension-agnostic inference using cross U-statistics

ILMUN KIM[1,a] and AADITYA RAMDAS[2,b]

[1]*Departments of Statistics & Data Science, Applied Statistics, Yonsei University, Seoul, South Korea,*
[a]*ilmun@yonsei.ac.kr*
[2]*Departments of Statistics & Data Science, Machine Learning, Carnegie Mellon University, Pittsburgh, USA,*
[b]*aramdas@cmu.edu*

Classical asymptotic theory for statistical inference usually involves calibrating a statistic by fixing the dimension d while letting the sample size n increase to infinity. Recently, much effort has been dedicated towards understanding how these methods behave in high-dimensional settings, where d and n both increase to infinity together. This often leads to different inference procedures, depending on the assumptions about the dimensionality, leaving the practitioner in a bind: given a dataset with 100 samples in 20 dimensions, should they calibrate by assuming $n \gg d$, or $d/n \approx 0.2$? This paper considers the goal of *dimension-agnostic inference*—developing methods whose validity does not depend on any assumption on d versus n. We introduce an approach that uses variational representations of existing test statistics along with sample splitting and self-normalization to produce a refined test statistic with a Gaussian limiting distribution, regardless of how d scales with n. The resulting statistic can be viewed as a careful modification of degenerate U-statistics, dropping diagonal blocks and retaining off-diagonal blocks. We exemplify our technique for some classical problems including one-sample mean and covariance testing, and show that our tests have minimax rate-optimal power against appropriate local alternatives. In most settings, our cross U-statistic matches the high-dimensional power of the corresponding (degenerate) U-statistic up to a $\sqrt{2}$ factor.

Keywords: Sample splitting; studentization; degenerate U-statistics; high-dimensional limits; minimax optimality

References

[1] Ahmad, I.A. (1993). Modification of some goodness-of-fit statistics to yield asymptotically normal null distributions. *Biometrika* **80** 466–472. MR1243521 https://doi.org/10.1093/biomet/80.2.466
[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience,. MR1990662
[3] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305
[4] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648
[5] Bean, D., Bickel, P.J., El Karoui, N. and Yu, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.
[6] Bentkus, V. and Götze, F. (1996). The Berry-Esseen bound for Student's statistic. *Ann. Probab.* **24** 491–503. MR1387647 https://doi.org/10.1214/aop/1042644728
[7] Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467 https://doi.org/10.1016/j.spl.2005.04.051
[8] Blanchard, G., Carpentier, A. and Gutzeit, M. (2018). Minimax Euclidean separation rates for testing convex hypotheses in $\mathbb{R}^d$. *Electron. J. Stat.* **12** 3713–3735. MR3873534 https://doi.org/10.1214/18-ejs1472
[9] Blom, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63** 573–580. MR0474582 https://doi.org/10.1093/biomet/63.3.573
[10] Cai, T.T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

[11] 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
[12] 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
[13] 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
[14] 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
[15] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450
[16] Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. *Proceedings of Machine Learning Research* **48** 2606–2615.
[17] Cox, D.R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika* **62** 441–444. MR0378189 https://doi.org/10.1093/biomet/62.2.441
[18] Deb, N., Bhattacharya, B.B. and Sen, B. (2021). Efficiency lower bounds for distribution-free hotelling-type two-sample tests based on optimal transport. ArXiv preprint. Available at arXiv:2104.01986.
[19] Deb, N. and Sen, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* **118** 192–207. MR4571116 https://doi.org/10.1080/01621459.2021.1923508
[20] Decrouez, G. and Hall, P. (2014). Split sample methods for constructing confidence intervals for binomial and Poisson parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 949–975. MR3271174 https://doi.org/10.1111/rssb.12051
[21] Donoho, D.L. and Feldman, M.J. (2022). Optimal Eigenvalue Shrinkage in the Semicircle Limit. ArXiv preprint. Available at arXiv:2210.04488.
[22] Dunn, R., Ramdas, A., Balakrishnan, S. and Wasserman, L. (2022). Gaussian universal likelihood ratio testing. *Biometrika*.
[23] Efron, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc.* **64** 1278–1302. MR0251826
[24] El Karoui, N. and Purdom, E. (2018). Can we trust the bootstrap in high-dimensions? The case of linear models. *J. Mach. Learn. Res.* **19** Paper No. 5. MR3862412
[25] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716
[26] 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.
[27] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096 https://doi.org/10.1016/0047-259X(84)90044-7
[28] 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
[29] Hallin, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.
[30] Hallin, M., del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996
[31] Hao, N. and Zhang, H.H. (2014). Interaction screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **109** 1285–1301. MR3265697 https://doi.org/10.1080/01621459.2014.881741
[32] 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
[33] Huang, Y. (2015). Projection test for high-dimensional mean vectors with optimal direction. PhD thesis, Pennsylvania State Univ.
[34] 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

[35] Ingster, Y.I. and Suslina, I.A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. New York: Springer. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[36] Janková, J., Shah, R.D., Bühlmann, P. and Samworth, R.J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 773–795. MR4112784 https://doi.org/10.1111/rssb.12371

[37] 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

[38] Katsevich, E. and Ramdas, A. (2022). On the power of conditional independence testing under model-X. *Electron. J. Stat.* **16** 6348–6394. MR4517344 https://doi.org/10.1214/22-ejs2085

[39] Kellner, J. and Celisse, A. (2019). A one-sample test for normality with kernel methods. *Bernoulli* **25** 1816–1837. MR3961232 https://doi.org/10.3150/18-BEJ1037

[40] Kim, I. (2020). Multinomial goodness-of-fit based on U-statistics: High-dimensional asymptotic and minimax optimality. *J. Statist. Plann. Inference* **205** 74–91. MR4011624 https://doi.org/10.1016/j.jspi.2019.06.005

[41] Kim, I., Balakrishnan, S. and Wasserman, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

[42] Kim, I., Balakrishnan, S. and Wasserman, L. (2022). Minimax optimality of permutation tests. *Ann. Statist.* **50** 225–251. MR4382015 https://doi.org/10.1214/21-aos2103

[43] Kim, I. and Ramdas, A. (2024). Supplement to "Dimension-agnostic inference using cross U-statistics." https://doi.org/10.3150/23-BEJ1613SUPP

[44] Kim, I., Ramdas, A., Singh, A. and Wasserman, L. (2021). Classification accuracy as a proxy for two-sample testing. *Ann. Statist.* **49** 411–434. MR4206684 https://doi.org/10.1214/20-AOS1962

[45] Kübler, J.M., Jitkrittum, W., Schölkopf, B. and Muandet, K. (2022). A witness two-sample test. In *International Conference on Artificial Intelligence and Statistics* 1403–1419. PMLR.

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

[47] Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. *Holden-Day Series in Probability and Statistics*. San Francisco, CA: Holden-Day. MR0395032

[48] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. MR2135927

[49] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J. and Wasserman, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

[50] Li, T. and Yuan, M. (2019). On the Optimality of Gaussian Kernel Based Nonparametric Tests against Smooth Alternatives. ArXiv preprint. Available at arXiv:1909.03302.

[51] Liu, Q., Lee, J. and Jordan, M. (2016). A kernelized Stein discrepancy for goodness-of-fit tests. *Proceedings of Machine Learning Research* **48** 276–284.

[52] Liu, W., Yu, X., Zhong, W. and Li, R. (2022). Projection test for mean vector in high dimensions. *J. Amer. Statist. Assoc.* 1–13.

[53] Makigusa, N. and Naito, K. (2020). Asymptotic normality of a consistent estimator of maximum mean discrepancy in Hilbert space. *Statist. Probab. Lett.* **156** 108596. MR3998243 https://doi.org/10.1016/j.spl.2019.108596

[54] Mentch, L. and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.* **17** Paper No. 26. MR3491120

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

[56] Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli* **22** 1745–1769. MR3474832 https://doi.org/10.3150/15-BEJ710

[57] Papa, G., Clémençon, S. and Bellet, A. (2015). SGD algorithms based on incomplete U-statistics: Large-scale minimization of empirical risk. In *Advances in Neural Information Processing Systems* 1027–1035.

[58] Pinelis, I. (2015). Exact bounds on the closeness between the Student and standard normal distributions. *ESAIM Probab. Stat.* **19** 24–27. MR3374867 https://doi.org/10.1051/ps/2014014

[59] Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690 https://doi.org/10.1214/aos/1176346793

[60] Portnoy, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499 https://doi.org/10.1214/aos/1176349744

[61] 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. ArXiv preprint. Available at arXiv:1508.00655.

[62] Rinaldo, A., Wasserman, L. and G'Sell, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784

[63] Robins, J.M., Li, L., Tchetgen, E.T. and van der Vaart, A. (2016). Asymptotic normality of quadratic estimators. *Stochastic Process. Appl.* **126** 3733–3759. MR3565475 https://doi.org/10.1016/j.spa.2016.04.005

[64] Romano, J.P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108. MR2156821 https://doi.org/10.1198/016214504000000539

[65] Serfling, R.J. (2009). *Approximation Theorems of Mathematical Statistics* **162**. John Wiley & Sons.

[66] Shekhar, S., Kim, I. and Ramdas, A. (2022). A permutation-free kernel two-sample test. *Neural Inf. Process. Syst.*

[67] Shekhar, S., Kim, I. and Ramdas, A. (2022). A Permutation-Free Kernel Independence Test. ArXiv preprint. Available at arXiv:2212.09108.

[68] Shi, H., Drton, M. and Han, F. (2022). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. MR4399094 https://doi.org/10.1080/01621459.2020.1782223

[69] Srivastava, M.S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970 https://doi.org/10.1016/j.jmva.2006.11.002

[70] Sur, P. and Candès, E.J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

[71] Székely, G.J. and Rizzo, M.L. (2005). A new test for multivariate normality. *J. Multivariate Anal.* **93** 58–80. MR2119764 https://doi.org/10.1016/j.jmva.2003.12.002

[72] Tansey, W., Veitch, V., Zhang, H., Rabadan, R. and Blei, D.M. (2022). The holdout randomization test for feature selection in black box models. *J. Comput. Graph. Statist.* **31** 151–162. MR4387217 https://doi.org/10.1080/10618600.2021.1923520

[73] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge: Cambridge Univ. Press. MR3837109 https://doi.org/10.1017/9781108231596

[74] Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

[75] Wang, R. and Xu, X. (2019). A feasible high dimensional randomization test for the mean vector. *J. Statist. Plann. Inference* **199** 160–178. MR3857820 https://doi.org/10.1016/j.jspi.2018.06.003

[76] Wang, X., Pan, W., Hu, W., Tian, Y. and Zhang, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734. MR3449068 https://doi.org/10.1080/01621459.2014.993081

[77] Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

[78] Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.

[79] 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

[80] Zhong, P.-S. and Chen, S.X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719 https://doi.org/10.1198/jasa.2011.tm10284

Bernoulli **30**(1), 2024, 712–742
https://doi.org/10.3150/23-BEJ1614

Central limit theorems for high dimensional dependent data

JINYUAN CHANG[1,2,a], XIAOHUI CHEN[3,4,c] and MINGCONG WU[1,b]

[1]*Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China,* [a]*changjinyuan@swufe.edu.cn,* [b]*wumingcong@smail.swufe.edu.cn*
[2]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[3]*Department of Mathematics, University of Southern California, Los Angeles, CA, USA,* [c]*xiaohuic@usc.edu*
[4]*Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

Motivated by statistical inference problems in high-dimensional time series data analysis, we first derive non-asymptotic error bounds for Gaussian approximations of sums of high-dimensional dependent random vectors on hyper-rectangles, simple convex sets and sparsely convex sets. We investigate the quantitative effect of temporal dependence on the rates of convergence to a Gaussian random vector over three different dependency frameworks (α-mixing, m-dependent, and physical dependence measure). In particular, we establish new error bounds under the α-mixing framework and derive faster rate over existing results under the physical dependence measure. To implement the proposed results in practical statistical inference problems, we also derive a data-driven parametric bootstrap procedure based on a kernel-type estimator for the long-run covariance matrices. The unified Gaussian and parametric bootstrap approximation results can be used to test mean vectors with combined ℓ^2 and ℓ^∞ type statistics, do change point detection, and construct confidence regions for covariance and precision matrices, all for time series data.

Keywords: Central limit theorem; dependent data; Gaussian approximation; high-dimensional statistical inference; parametric bootstrap

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513 https://doi.org/10.2307/2938229

Baldi, P. and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650. MR1048950

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

Berkes, I., Liu, W. and Wu, W.B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

Bhattacharya, R.N. and Holmes, S. (2010). An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem. Available at arXiv:1003.4254.

Bhattacharya, R.N. and Rao, R.R. (2010). *Normal Approximation and Asymptotic Expansions*. *Classics in Applied Mathematics* **64**. Philadelphia, PA: SIAM. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Process. Appl.* **121** 2331–2360. MR2822779 https://doi.org/10.1016/j.spa.2011.06.001

Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions. Vol. 2*. Heber City, UT: Kendrick Press. MR2325295

1350-7265

Cai, T.T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

Chang, J., Chen, X. and Wu, M. (2024). Supplement to "Central limit theorems for high dimensional dependent data." https://doi.org/10.3150/23-BEJ1614SUPP

Chang, J., Hu, Q., Liu, C. and Tang, C.Y. (2023). Optimal covariance matrix estimation for high-dimensional noise in high-frequency data. *J. Econometrics*. To appear. https://doi.org/10.1016/j.jeconom.2022.06.010

Chang, J., Jiang, Q. and Shao, X. (2023). Testing the martingale difference hypothesis in high dimension. *J. Econometrics*. To appear. https://doi.org/10.1016/j.jeconom.2022.09.001

Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *J. Econometrics* **206** 57–82. MR3840783 https://doi.org/10.1016/j.jeconom.2018.03.020

Chang, J., Tang, C.Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* **41** 2123–2148. MR3127860 https://doi.org/10.1214/13-AOS1139

Chang, J., Yao, Q. and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika* **104** 111–127. MR3626482 https://doi.org/10.1093/biomet/asw066

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017a). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2017b). 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

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, X. and Kato, K. (2019). Randomized incomplete U-statistics in high dimensions. *Ann. Statist.* **47** 3127–3156. MR4025737 https://doi.org/10.1214/18-AOS1773

Chen, X. and Kato, K. (2020). Jackknife multiplier bootstrap: Finite sample approximations to the U-process supremum with applications. *Probab. Theory Related Fields* **176** 1097–1163. MR4087490 https://doi.org/10.1007/s00440-019-00936-y

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, L.H.Y. and Shao, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183 https://doi.org/10.1214/009117904000000450

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

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. MR4009488 https://doi.org/10.1093/restud/rdy065

Chernozhuokov, V., Chetverikov, D. and Koike, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* To appear.

Chernozhuokov, V., Chetverikov, D., Kato, K. and Koike, Y. (2022a). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

Das, D. and Lahiri, S. (2021). Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** 6991–7009. MR4315595 https://doi.org/10.1090/tran/8459

Davydov, J.A. (1968). The convergence of distributions which are generated by stationary random processes. *Theory Probab. Appl.* **13** 691–696.

den Haan, W.J. and Levin, A.T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust Inference*. *Handbook of Statist.* **15** 299–342. Amsterdam: North-Holland. MR1492717 https://doi.org/10.1016/S0169-7161(97)15014-3

Deng, H. (2020). Slightly conservative bootstrap for maxima of sums. Available at arXiv:2007.15877.

Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

Doukhan, P., Massart, P. and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **30** 63–82. MR1262892

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. *Springer Series in Statistics*. New York: Springer. MR1964455 https://doi.org/10.1007/b97702

Fang, X. and Koike, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

Hafner, C.M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *J. Multivariate Anal.* **100** 2044–2054. MR2543085 https://doi.org/10.1016/j.jmva.2009.03.011

Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15** 773–780. MR0026771

Hörmann, S. (2009). Berry-Esseen bounds for econometric time series. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** 377–397. MR2557877

Jirak, M. (2016). Berry-Esseen theorems under weak dependence. *Ann. Probab.* **44** 2024–2063. MR3502600 https://doi.org/10.1214/15-AOP1017

Kiefer, N.M., Vogelsang, T.J. and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica* **68** 695–714. MR1769382 https://doi.org/10.1111/1468-0262.00128

Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7

Koike, Y. (2023). High-dimensional central limit theorems for homogeneous sums. *J. Theoret. Probab.* To appear. https://doi.org/10.1007/s10959-022-01156-2

Kuchibhotla, A.K. and Rinaldo, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Available at arXiv:2009.13673.

Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*. *Springer Series in Statistics*. New York: Springer. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

Lopes, M.E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

Lopes, M.E., Lin, Z. and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229. MR4102694 https://doi.org/10.1214/19-AOS1844

Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. Oxford University Press, New York: The Clarendon Press. MR1353441

Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. MR2472010 https://doi.org/10.1007/s10959-008-0155-9

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. *Probability Theory and Stochastic Modelling* **80**. Berlin: Springer. MR3642873 https://doi.org/10.1007/978-3-662-54323-8

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

Saulis, L. and Statulevičius, V.A. (1991). *Limit Theorems for Large Deviations*. Netherlands: Springer.

Song, Y., Chen, X. and Kato, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electron. J. Stat.* **13** 4794–4848. MR4038726 https://doi.org/10.1214/19-EJS1643

Sunklodas, Ĭ. (1984). The rate of convergence in the central limit theorem for strongly mixing random variables. *Lith. Math. J.* **24** 182–190.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

Wong, K.C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690 https://doi.org/10.1214/19-AOS1840

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

Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 https://doi.org/10.1214/009117907000000060

Wu, W.B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 https://doi.org/10.1239/jap/1082999076

Wu, W.-B. and Wu, Y.N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186 https://doi.org/10.1214/16-EJS1108

Yu, M. and Chen, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

Yu, M. and Chen, X. (2022). A robust bootstrap change point test for high-dimensional location parameter. *Electron. J. Stat.* **16** 1096–1152. MR4377134 https://doi.org/10.1214/21-ejs1915

Zhang, X. (2015). Testing high dimensional mean under sparsity. Available at arXiv:1509.08444.

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697 https://doi.org/10.3150/17-BEJ939

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

Bernoulli **30**(1), 2024, 743–769
https://doi.org/10.3150/23-BEJ1615

A central limit theorem for the Benjamini-Hochberg false discovery proportion under a factor model

DAN M. KLUGER[a] and ART B. OWEN[b]

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

The Benjamini-Hochberg (BH) procedure remains widely popular despite having limited theoretical guarantees in the commonly encountered scenario of correlated test statistics. Of particular concern is the possibility that the method could exhibit bursty behavior, meaning that it might typically yield no false discoveries while occasionally yielding both a large number of false discoveries and a false discovery proportion (FDP) that far exceeds its own well controlled mean. In this paper, we investigate which test statistic correlation structures lead to bursty behavior and which ones lead to well controlled FDPs. To this end, we develop a central limit theorem for the FDP in a multiple testing setup where the test statistic correlations can be either short-range or long-range as well as either weak or strong. The theorem and our simulations from a data-driven factor model suggest that the BH procedure exhibits severe burstiness when the test statistics have many strong, long-range correlations, but does not otherwise.

Keywords: Empirical cumulative distribution function; functional central limit theorem; functional delta method; multiple hypothesis testing; Simes line

References

Andrews, D.W.K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int. Stat. Rev.* **62** 119–132.

Azriel, D. and Schwartzman, A. (2015). The empirical distribution of a large number of correlated normal variables. *J. Amer. Statist. Assoc.* **110** 1217–1228. MR3420696 https://doi.org/10.1080/01621459.2014.958156

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 Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* **25** 60–83.

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

Chi, Z. (2007). On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **35** 1409–1431. MR2351091 https://doi.org/10.1214/009053607000000037

Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.* **81** 111–115. MR2740072 https://doi.org/10.1016/j.spl.2010.09.025

Delattre, S. and Roquain, E. (2015). New procedures controlling the false discovery proportion via Romano-Wolf's heuristic. *Ann. Statist.* **43** 1141–1177. MR3346700 https://doi.org/10.1214/14-AOS1302

Delattre, S. and Roquain, E. (2016). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *Bernoulli* **22** 302–324. MR3449784 https://doi.org/10.3150/14-BEJ659

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist. Assoc.* **102** 93–103. MR2293302 https://doi.org/10.1198/016214506000001211

Fan, J., Han, X. and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887 https://doi.org/10.1080/01621459.2012.720478

Fan, J., Ke, Y., Sun, Q. and Zhou, W.-X. (2019). FarmTest: Factor-adjusted robust multiple testing with approximate false discovery control. *J. Amer. Statist. Assoc.* **114** 1880–1893. MR4047307 https://doi.org/10.1080/01621459.2018.1527700

Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Stat. Methods Appl.* **15** 43–73. MR2281214 https://doi.org/10.1007/s10260-006-0002-z

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Stat.* **34** 275–297. MR2346640 https://doi.org/10.1111/j.1467-9469.2006.00530.x

Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist.* **35** 1432–1455. MR2351092 https://doi.org/10.1214/009053607000000046

Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biom. J.* **43** 985–1005. MR1878272 https://doi.org/10.1002/1521-4036(200112)43:8<985::AID-BIMJ985>3.0.CO;2-4

Finner, H. and Roters, M. (2002). Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.* **30** 220–238. MR1892662 https://doi.org/10.1214/aos/1015362191

Fithian, W. and Lei, L. (2022). Conditional calibration for false discovery rate control under dependence. *Ann. Statist.* **50** 3091–3118. MR4524490 https://doi.org/10.1214/21-aos2137

Friguet, C., Kloareg, M. and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist. Assoc.* **104** 1406–1415. MR2750571 https://doi.org/10.1198/jasa.2009.tm08332

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

Gerard, D. and Stephens, M. (2020). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics* **21** 15–32. MR4043843 https://doi.org/10.1093/biostatistics/kxy029

Gontscharuk, V. and Finner, H. (2013). Asymptotic FDR control under weak dependence: A counterexample. *Statist. Probab. Lett.* **83** 1888–1893. MR3069893 https://doi.org/10.1016/j.spl.2013.04.025

Izmirlian, G. (2020). Strong consistency and asymptotic normality for quantities related to the Benjamini-Hochberg false discovery rate procedure. *Statist. Probab. Lett.* **160** 108713, 10. MR4061847 https://doi.org/10.1016/j.spl.2020.108713

Kim, K.I. and van de Wiel, M.A. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinform.* **9** 114. https://doi.org/10.1186/1471-2105-9-114

Kluger, D.M. and Owen, A.B. (2024). Supplement to "A central limit theorem for the Benjamini-Hochberg false discovery proportion under a factor model." https://doi.org/10.3150/23-BEJ1615SUPP

Korn, E.L., Troendle, J.F., McShane, L.M. and Simon, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371 https://doi.org/10.1016/S0378-3758(03)00211-8

Kotelnikova, E., Shkrob, M.A., Pyatnitskiy, M.A., Ferlini, A. and Daraselia, N. (2012). Novel approach to meta-analysis of microarray datasets reveals muscle remodeling-related drug targets and biomarkers in Duchenne muscular dystrophy. *PLoS Comput. Biol.* **8** 1–10. https://doi.org/10.1371/journal.pcbi.1002365

Lindner, A.M. (2009). Stationarity, mixing, distributional properties and moments of GARCH(p, q)–processes. In *Handbook of Financial Time Series* 43–69. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71297-8\protect \T1\textunderscore 2

Lucas, J.E., Kung, H.N. and Chi, J.T.A. (2010). Latent factor analysis to discover pathway-associated putative segmental aneuploidies in human cancers. *PLoS Comput. Biol.* **6** e100920:1–15.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Process. Appl.* **29** 309–315. MR0958507 https://doi.org/10.1016/0304-4149(88)90045-2

Neumann, M.H. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM Probab. Stat.* **17** 120–134. MR3021312 https://doi.org/10.1051/ps/2011144

Neuvial, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2** 1065–1110. MR2460858 https://doi.org/10.1214/08-EJS207

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

Owen, A.B. and Wang, J. (2016). Bi-cross-validation for factor analysis. *Statist. Sci.* **31** 119–139. MR3458596 https://doi.org/10.1214/15-STS539

Paolella, M.S. (2019). *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR3839332

Romano, J.P. and Shaikh, A.M. (2006). On stepdown control of the false discovery proportion. In *Optimality*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **49** 33–50. Beachwood, OH: IMS. MR2337829 https://doi.org/10.1214/074921706000000383

Romano, J.P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. MR2351090 https://doi.org/10.1214/009053606000001622

Schwartzman, A. and Lin, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika* **98** 199–214. MR2804220 https://doi.org/10.1093/biomet/asq075

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

Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100** 9440–9445. MR1994856 https://doi.org/10.1073/pnas.1530509100

Sun, Y., Zhang, N.R. and Owen, A.B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *Ann. Appl. Stat.* **6** 1664–1688. MR3058679 https://doi.org/10.1214/12-AOAS561

van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Wang, J., Zhao, Q., Hastie, T. and Owen, A.B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Statist.* **45** 1863–1894. MR3718155 https://doi.org/10.1214/16-AOS1511

Wang, J., Gui, L., Su, W.J., Sabatti, C. and Owen, A.B. (2022). Detecting multiple replicating signals using adaptive filtering procedures. *Ann. Statist.* **50** 1890–1909. MR4474476 https://doi.org/10.1214/21-aos2139

Zhang, C., Fan, J. and Yu, T. (2011). Multiple testing via FDR_L for large-scale imaging data. *Ann. Statist.* **39** 613–642. MR2797858 https://doi.org/10.1214/10-AOS848

Bernoulli **30**(1), 2024, 770–796
https://doi.org/10.3150/23-BEJ1616

Limiting distributions of graph-based test statistics on sparse and dense graphs

YEJIONG ZHU[a] and HAO CHEN[b]

University of California, Davis, USA, [a]*yjzhu@ucdavis.edu,* [b]*hxchen@ucdavis.edu*

Two-sample tests utilizing a similarity graph on observations are useful for high-dimensional and non-Euclidean data due to their flexibility and good performance under a wide range of alternatives. Existing works mainly focused on sparse graphs, such as graphs with the number of edges in the order of the number of observations, and their asymptotic results imposed strong conditions on the graph that can easily be violated by commonly constructed graphs they suggested. Moreover, the graph-based tests have better performance with denser graphs under many settings. In this work, we establish the theoretical ground for graph-based tests with graphs ranging from those recommended in current literature to much denser ones.

Keywords: Dense graphs; graph-based methods; nonparametric two-sample tests; Stein's method

References

Beckmann, N.D., Comella, P.H., Cheng, E., Lepow, L., Beckmann, A.G., Tyler, S.R., Mouskas, K., Simons, N.W., Hoffman, G.E., Francoeur, N.J. et al. (2021). Downregulation of exhausted cytotoxic T cells in gene expression networks of multisystem inflammatory syndrome in children. *Nat. Commun.* **12** 1–15.

Bickel, P.J. (1968). A distribution free version of the Smirnov two sample test in the *p*-variate case. *Ann. Math. Stat.* **40** 1–23. MR0256519 https://doi.org/10.1214/aoms/1177697800

Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S. et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci.* **107** 4734–4739.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10** 186–198. https://doi.org/10.1038/nrn2575

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, 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, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

Chen, H. and Zhang, N.R. (2013). Graph-based tests for two-sample comparisons of categorical data. *Statist. Sinica* **23** 1479–1503. MR3222245

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

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

Daniels, H.E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33** 129–135. MR0010941 https://doi.org/10.2307/2334112

Feigenson, K.A., Gara, M.A., Roché, M.W. and Silverstein, S.M. (2014). Is disorganization a feature of schizophrenia or a modifying influence: Evidence of covariation of perceptual and cognitive organization in a non-patient sample. *Psychiatry Res.* **217** 1–8.

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

Friedman, J.H. and Rafsky, L.C. (1983). Graph-theoretic measures of multivariate association and prediction. *Ann. Statist.* **11** 377–391. MR0696054

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

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

Henze, N. and Penrose, M.D. (1999). On the multivariate runs test. *Ann. Statist.* **27** 290–298. MR1701112 https://doi.org/10.1214/aos/1018031112

Hoeffding, W. (1951). A combinatorial central limit theorem. *Ann. Math. Stat.* **22** 558–566. MR0044058 https://doi.org/10.1214/aoms/1177729545

Li, H., Aue, A., Paul, D., Peng, J. and Wang, P. (2020). An adaptable generalization of Hotelling's T^2 test in high dimension. *Ann. Statist.* **48** 1815–1847. MR4124345 https://doi.org/10.1214/19-AOS1869

Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18** 50–60. MR0022058 https://doi.org/10.1214/aoms/1177730491

Network, C.G.A. et al. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490** 61.

Pham, D.T., Möcks, J. and Sroka, L. (1989). Asymptotic normality of double-indexed linear permutation statistics. *Ann. Inst. Statist. Math.* **41** 415–427. MR1032590 https://doi.org/10.1007/BF00050659

Robins, G. and Salowe, J.S. (1994). On the maximum degree of minimum spanning trees. In *Proceedings of the Tenth Annual Symposium on Computational Geometry* 250–258.

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

Smirnov, N.V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou* **2** 3–14.

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

Weiss, L. (1960). Two-sample tests for multivariate distributions. *Ann. Math. Stat.* **31** 159–164. MR0119305 https://doi.org/10.1214/aoms/1177705995

Zhang, J. and Chen, H. (2022). Graph-based two-sample tests for data with repeated observations. *Statist. Sinica* **32** 391–415. MR4359638 https://doi.org/10.5705/ss.202019.0116

Zhang, J.-T., Guo, J., Zhou, B. and Cheng, M.-Y. (2020). A simple two-sample test in high dimensions based on L^2-norm. *J. Amer. Statist. Assoc.* **115** 1011–1027. MR4107696 https://doi.org/10.1080/01621459.2019.1604366

Zhu, Y. and Chen, H. (2024). Supplement to "Limiting distributions of graph-based test statistics on sparse and dense graphs." https://doi.org/10.3150/23-BEJ1616SUPP

Bernoulli **30**(1), 2024, 797–824
https://doi.org/10.3150/23-BEJ1617

On estimators of the mean of infinite dimensional data in finite populations

ANURAG DEY[a] and PROBAL CHAUDHURI[b]

Statistics and Mathematics unit, Indian Statistical Institute, Kolkata, India, [a]*deyanuragsaltlake64@gmail.com,* [b]*probalchaudhuri@gmail.com*

The Horvitz-Thompson (HT), the Rao-Hartley-Cochran (RHC) and the generalized regression (GREG) estimators of the finite population mean are considered, when the observations are from an infinite dimensional space. We compare these estimators based on their asymptotic distributions under some commonly used sampling designs and some superpopulations satisfying linear regression models. We show that the GREG estimator is asymptotically at least as efficient as any of the other two estimators under different sampling designs considered in this paper. Further, we show that the use of some well known sampling designs utilizing auxiliary information may have an adverse effect on the performance of the GREG estimator, when the degree of heteroscedasticity present in linear regression models is not very large. On the other hand, the use of those sampling designs improves the performance of this estimator, when the degree of heteroscedasticity present in linear regression models is large. We develop methods for determining the degree of heteroscedasticity, which in turn determines the choice of appropriate sampling design to be used with the GREG estimator. We also investigate the consistency of the covariance operators of the above estimators. We carry out some numerical studies using real and synthetic data, and our theoretical results are supported by the results obtained from those numerical studies.

Keywords: Asymptotic normality; consistency of estimators; covariance operator; heteroscedasticity; high entropy sampling design; inclusion probability; relative efficiency; separable Hilbert space

References

Berger, Y.G. (1998). Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693 https://doi.org/10.1016/S0378-3758(97)00107-9

Boistard, H., Lopuhaä, H.P. and Ruiz-Gazen, A. (2017). Functional central limit theorems for single-stage sampling designs. *Ann. Statist.* **45** 1728–1758. MR3670194 https://doi.org/10.1214/16-AOS1507

Bondesson, L., Traat, I. and Lundqvist, A. (2006). Pareto sampling versus Sampford and conditional Poisson sampling. *Scand. J. Stat.* **33** 699–720. MR2300911 https://doi.org/10.1111/j.1467-9469.2006.00497.x

Breusch, T.S. and Pagan, A.R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47** 1287–1294. MR0545960 https://doi.org/10.2307/1911963

Cardot, H., Degras, D. and Josserand, E. (2013). Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data. *Bernoulli* **19** 2067–2097. MR3129044 https://doi.org/10.3150/12-BEJ443

Cardot, H., Goga, C. and Lardin, P. (2013). Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electron. J. Stat.* **7** 562–596. MR3035266 https://doi.org/10.1214/13-EJS779

Cardot, H., Goga, C. and Lardin, P. (2014). Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Scand. J. Stat.* **41** 516–534. MR3207184 https://doi.org/10.1111/sjos.12048

Cardot, H. and Josserand, E. (2011). Horvitz-Thompson estimators for functional data: Asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika* **98** 107–118. MR2804213 https://doi.org/10.1093/biomet/asq070

1350-7265

Cardot, H., Dessertaine, A., Goga, C., Josserand, É. and Lardin, P. (2013). Comparison of different sample designs and construction of confidence bands to estimate the mean of functional data: An illustration on electricity consumption. *Surv. Methodol.* **39** 283–301.

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

Chaudhuri, A. (2014). *Modern Survey Sampling*. Boca Raton, FL: CRC Press. MR3234566

Chaudhuri, A., Dihidar, K. and Bose, M. (2006). On the feasibility of basing Horvitz and Thompson's estimator on a sample by Rao, Hartley, and Cochran's scheme. *Comm. Statist. Theory Methods* **35** 2239–2244. MR2338928 https://doi.org/10.1080/03610920600853464

Chen, J. and Sitter, R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statist. Sinica* **9** 385–406. MR1707846

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0474575

Conti, P.L. (2014). On the estimation of the distribution function of a finite population under high entropy sampling designs, with applications. *Sankhya B* **76** 234–259. MR3302272 https://doi.org/10.1007/s13571-014-0083-x

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

Dey, A. and Chaudhuri, P. (2024). Supplement to "On estimators of the mean of infinite dimensional data in finite populations." https://doi.org/10.3150/23-BEJ1617SUPP

Fuller, W.A. (2011). *Sampling Statistics*. New York: Wiley.

Glejser, H. (1969). A new test for heteroskedasticity. *J. Amer. Statist. Assoc.* **64** 316–323.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.* **35** 1491–1523. MR0178555 https://doi.org/10.1214/aoms/1177700375

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

Kundu, S., Majumdar, S. and Mukherjee, K. (2000). Central limit theorems revisited. *Statist. Probab. Lett.* **47** 265–275. MR1747487 https://doi.org/10.1016/S0167-7152(99)00164-9

Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.* **33** 133–140.

Midzuno, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Ann. Inst. Stat. Math., Tokyo* **3** 99–107. MR0050840 https://doi.org/10.1007/bf02949779

Murthy, M.N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society. MR0474578

Rao, J.N.K. (2003). *Small Area Estimation*. *Wiley Series in Survey Methodology*. Hoboken, NJ: Wiley. MR1953089 https://doi.org/10.1002/0471722189

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser. B* **24** 482–491. MR0148196

Robinson, P.M. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya, Ser. B* **45** 240–248. MR0748468

Särndal, C.-E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika* **67** 639–650. MR0601102 https://doi.org/10.2307/2335134

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. *Springer Series in Statistics*. New York: Springer. MR1140409 https://doi.org/10.1007/978-1-4612-4378-6

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agricultural Statist.* **5** 119–127. MR0068179

Wang, J.C. and Opsomer, J.D. (2011). On asymptotic normality and variance estimation for nondifferentiable survey estimators. *Biometrika* **98** 91–106. MR2804212 https://doi.org/10.1093/biomet/asq077

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027 https://doi.org/10.2307/1912934

Bernoulli **30**(1), 2024, 825–850
https://doi.org/10.3150/23-BEJ1618

High dimensional Bernoulli distributions: Algebraic representation and applications

ROBERTO FONTANA[a] and PATRIZIA SEMERARO[b]

Department of Mathematical Sciences, Politecnico di Torino, corso Duca Degli Abruzzi 24, 10129 Torino, Italy, [a]*roberto.fontana@polito.it,* [b]*patrizia.semeraro@polito.it*

The main contribution of this paper is to find a representation of the class $\mathcal{F}_d(p)$ of multivariate Bernoulli distributions with the same mean p that allows us to find its generators analytically in any dimension. We map $\mathcal{F}_d(p)$ to an ideal of points and we prove that the class $\mathcal{F}_d(p)$ can be generated from a finite set of simple polynomials. We present two applications. Firstly, we show that polynomial generators help to find extremal points of the convex polytope $\mathcal{F}_d(p)$ in high dimensions. Secondly, we solve the problem of determining the lower bounds in the convex order for sums of multivariate Bernoulli distributions with given margins, but with an unspecified dependence structure.

Keywords: Convex order; extremal points; ideal of points; multidimensional Bernoulli distribution

References

[1] 4ti2team (2018). 4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.

[2] Boros, E. and Prékopa, A. (1989). Closed form two-sided bounds for probabilities that at least r and exactly r out of n events occur. *Math. Oper. Res.* **14** 317–342. MR0997039 https://doi.org/10.1287/moor.14.2.317

[3] Chaganty, N.R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93** 197–206. MR2277750 https://doi.org/10.1093/biomet/93.1.197

[4] Cheung, K.C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance Math. Econom.* **55** 180–190. MR3179810 https://doi.org/10.1016/j.insmatheco.2014.01.001

[5] Cox, D.A., Little, J. and O'Shea, D. (2015). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed. *Undergraduate Texts in Mathematics*. Cham: Springer. MR3330490 https://doi.org/10.1007/978-3-319-16721-3

[6] Dai, B., Ding, S. and Wahba, G. (2013). Multivariate Bernoulli distribution. *Bernoulli* **19** 1465–1483. MR3102559 https://doi.org/10.3150/12-BEJSP10

[7] Denuit, M. and Vermandele, C. (1998). Optimal reinsurance and stop-loss order. *Insurance Math. Econom.* **22** 229–233. MR1650288 https://doi.org/10.1016/S0167-6687(97)00039-5

[8] Dhaene, J. and Denuit, M. (1999). The safest dependence structure among risks. *Insurance Math. Econom.* **25** 11–21. MR1718527 https://doi.org/10.1016/S0167-6687(99)00009-8

[9] Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156 https://doi.org/10.1214/aos/1030563990

[10] Emrich, L.J. and Piedmonte, M.R. (1991). A method for generating high-dimensional multivariate binary variates. *Amer. Statist.* **45** 302–304.

[11] Fontana, R., Luciano, E. and Semeraro, P. (2021). Model risk in credit risk. *Math. Finance* **31** 176–202. MR4205881 https://doi.org/10.1111/mafi.12285

[12] Fontana, R. and Semeraro, P. (2018). Representation of multivariate Bernoulli distributions with a given set of specified moments. *J. Multivariate Anal.* **168** 290–303. MR3858366 https://doi.org/10.1016/j.jmva.2018.08.003

1350-7265

[13] Fontana, R. and Semeraro, P. (2022). Computational and analytical bounds for multivariate Bernoulli distributions. *J. Stat. Theory Pract.* **16** Paper No. 6, 12. MR4366952 https://doi.org/10.1007/s42519-021-00231-x
[14] Fontana, R. and Semeraro, P. (2023). Exchangeable Bernoulli distributions: High dimensional simulation, estimation, and testing. *J. Statist. Plann. Inference* **225** 52–70. MR4520494 https://doi.org/10.1016/j.jspi.2022.11.005
[15] Fontana, R. and Semeraro, P. (2024). Supplement to "High dimensional Bernoulli distributions: algebraic representation and applications." https://doi.org/10.3150/23-BEJ1618SUPP
[16] Frostig, E. (2001). Comparison of portfolios which depend on multivariate Bernoulli random variables with fixed marginals. *Insurance Math. Econom.* **29** 319–331. MR1874627 https://doi.org/10.1016/S0167-6687(01)00092-0
[17] Haynes, M.E., Sabo, R.T. and Chaganty, N.R. (2016). Simulating dependent binary variables through multinomial sampling. *J. Stat. Comput. Simul.* **86** 510–523. MR3421742 https://doi.org/10.1080/00949655.2015.1020313
[18] Hofert, M., Memartoluie, A., Saunders, D. and Wirjanto, T. (2017). Improved algorithms for computing worst Value-at-Risk. *Stat. Risk Model.* **34** 13–31. MR3652295 https://doi.org/10.1515/strm-2015-0028
[19] Kaas, R., Dhaene, J. and Goovaerts, M.J. (2000). Upper and lower bounds for sums of random variables. *Insurance Math. Econom.* **27** 151–168. MR1801600 https://doi.org/10.1016/S0167-6687(00)00060-3
[20] Kang, S.-H. and Jung, S.-H. (2001). Generating correlated binary variables with complete specification of the joint distribution. *Biom. J.* **43** 263–269. MR1855743 https://doi.org/10.1002/1521-4036(200106)43:3<263::AID-BIMJ263>3.0.CO;2-5
[21] Kvam, P.H. (1996). Maximum likelihood estimation and the multivariate Bernoulli distribution: An application to reliability. In *Lifetime Data: Models in Reliability and Survival Analysis* 187–194. Springer.
[22] Lee, W., Cheung, K.C. and Ahn, J.Y. (2017). Multivariate countermonotonicity and the minimal copulas. *J. Comput. Appl. Math.* **317** 589–602. MR3606098 https://doi.org/10.1016/j.cam.2016.12.032
[23] Marchetti, G.M. and Wermuth, N. (2016). Palindromic Bernoulli distributions. *Electron. J. Stat.* **10** 2435–2460. MR3545465 https://doi.org/10.1214/16-EJS1175
[24] McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised ed. *Princeton Series in Finance*. Princeton, NJ: Princeton Univ. Press. MR3445371
[25] Padmanabhan, D. and Natarajan, K. (2021). Tree bounds for sums of Bernoulli random variables: A linear optimization approach. *INFORMS J. Optim.* **3** 23–45. MR4237329 https://doi.org/10.1287/ijoo.2019.0038
[26] Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statist. Sci.* **30** 485–517. MR3432838 https://doi.org/10.1214/15-STS525
[27] Qaqish, B.F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* **90** 455–463. MR1986660 https://doi.org/10.1093/biomet/90.2.455
[28] Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton, NJ: Princeton Univ. Press. MR0274683
[29] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer. MR2265633 https://doi.org/10.1007/978-0-387-34675-5
[30] Shults, J. (2017). Simulating longer vectors of correlated binary random variables via multinomial sampling. *Comput. Statist. Data Anal.* **114** 1–11. MR3660834 https://doi.org/10.1016/j.csda.2017.04.002
[31] Terzer, M. (2009). Large scale methods to enumerate extreme rays and elementary modes. Ph.D. thesis, ETH Zurich.