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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 53, No. 5—October 2025

THE ANNALS OF STATISTICS Vol. 53, No. 5, pp. 1833–2302 October 2025

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 53, Number 5, October 2025. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, MD 21769, USA.

The Annals of Statistics
2025, Vol. 53, No. 5, 1833–1856
https://doi.org/10.1214/25-AOS2517

LOW-DEGREE HARDNESS OF DETECTION FOR CORRELATED ERDŐS–RÉNYI GRAPHS

BY JIAN DING[1,a], HANG DU[2,c] AND ZHANGSONG LI[1,b]

[1]*School of Mathematical Science, Peking University,* [a]*dingjian@math.pku.edu.cn,* [b]*ramblerlzs@pku.edu.cn*
[2]*Department of Mathematics, Massachusetts Institute of Technology,* [c]*hangdu@mit.edu*

Given two Erdős–Rényi graphs with n vertices whose edges are correlated through a latent vertex correspondence, we study complexity lower bounds for the associated correlation detection problem for the class of low-degree polynomial algorithms. We prove that no degree-$O(\rho^{-1})$ polynomial algorithm can succeed for detection under an appropriate definition of success, where ρ is the edge correlation. Furthermore, in the sparse regime where the edge density $q = n^{-1+o(1)}$, we prove that no degree-d polynomial algorithm can succeed for detection under an appropriate definition of success, as long as $d = \exp(o(\frac{\log n}{\log nq} \wedge \sqrt{\log n}))$ and the correlation $\rho < \sqrt{\alpha}$ where $\alpha \approx 0.338$ is the Otter's constant. Our result suggests that several state-of-the-art algorithms on correlation detection and exact matching recovery may be essentially the best possible.

REFERENCES

[1] APOSTOL, T. M. (1976). *Introduction to Analytic Number Theory*. *Undergraduate Texts in Mathematics*. Springer, New York. MR0434929

[2] BANDEIRA, A. S., ALAOUI, A. E., HOPKINS, S., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. In *Advances in Neural Information Processing Systems* (*NIPS*) **35** 33831–33844. Curran Associates, Red Hook.

[3] BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **151** Art. No. 78, 29. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181

[4] BARAK, B., CHOU, C. N., LEI, Z., SCHRAMM, T. and SHENG, Y. (2019). (Nearly) efficient algorithms for the graph matching problem on correlated random graphs. In *Advances in Neural Information Processing Systems* (*NIPS*) **32** 9190–9198. Curran Associates, Red Hook.

[5] BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput*. **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

[6] BERG, A., BERG, T. and MALIK, J. (2005). Shape matching and object recognition using low distortion correspondences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR*) **1** 26–33. IEEE.

[7] BOZORG, M., SALEHKALEYBAR, S. and HASHEMI, M. (2019). Seedless graph matching via tail of degree distribution for correlated Erdős-Rényi graphs. Preprint. Available at arXiv:1907.06334.

[8] BRESLER, G. and HUANG, B. (2022). The algorithmic phase transition of random k-SAT for low degree polynomials. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 298–309. IEEE Computer Soc., Los Alamitos, CA. MR4399691

[9] COUR, T., SRINIVASAN, P. and SHI, J. (2006). *Balanced Graph Matching*. *Advances in Neural Information Processing Systems* (*NIPS*) **19**. MIT Press, Cambridge.

[10] CULLINA, D. and KIYAVASH, N. (2016). Improved achievability and converse bounds for Erdős-Rényi graph matching. In *Proceedings of the* 2016 *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* 63–72. ACM, New York.

MSC2020 subject classifications. Primary 62M20; secondary 68Q87.
Key words and phrases. Correlated Erdős–Rényi random graphs, low-degree polynomials, computational complexity.

[11] Cullina, D. and Kiyavash, N. (2017). Exact alignment recovery for correlated Erdős-Rényi graphs. Preprint. Available at arXiv:1711.06783.

[12] Cullina, D., Kiyavash, N., Mittal, P. and Poor, H. V. (2020). Partial recovery of Erdős–Rényi graph alignment via k-core alignment. In *Proceedings of the* 2020 *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* 99–100. ACM, New York.

[13] Dai, O. E., Cullina, D., Kiyavash, N. and Grossglauser, M. (2019). Analysis of a canonical labeling algorithm for the alignment of correlated Erdős-Rényi graphs. In *Proc. ACM Meas. Anal. Comput. Syst.* **3**.

[14] Dhawan, A., Mao, C. and Wein, A. S. (2025). Detection of dense subhypergraphs by low-degree polynomials. *Random Structures Algorithms* **66** Paper No. e21279, 16. MR4853082 https://doi.org/10.1002/rsa.21279

[15] Ding, J. and Du, H. (2023). Detection threshold for correlated Erdős-Rényi graphs via densest subgraph. *IEEE Trans. Inf. Theory* **69** 5289–5298. MR4624732 https://doi.org/10.1109/tit.2023.3265009

[16] Ding, J. and Du, H. (2023). Matching recovery threshold for correlated random graphs. *Ann. Statist.* **51** 1718–1743. MR4658574 https://doi.org/10.1214/23-aos2305

[17] Ding, J., Du, H. and Li, Z. (2025). Supplement to "Low-degree hardness of detection for correlated Erdős–Rényi graphs." https://doi.org/10.1214/25-AOS2517SUPP

[18] Ding, J. and Li, Z. (2025). A polynomial time iterative algorithm for matching Gaussian matrices with non-vanishing correlation. *Found. Comput. Math.* **25** 1287–1344. https://doi.org/10.1007/s10208-024-09662-x

[19] Ding, J. and Li, Z. (2023). A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. Preprint. Available at arXiv:2306.00266.

[20] Ding, J., Ma, Z., Wu, Y. and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probab. Theory Related Fields* **179** 29–115. MR4221654 https://doi.org/10.1007/s00440-020-00997-4

[21] Ding, Y., Kunisky, D., Wein, A. S. and Bandeira, A. S. (2024). Subexponential-time algorithms for sparse PCA. *Found. Comput. Math.* **24** 865–914. MR4760356 https://doi.org/10.1007/s10208-023-09603-0

[22] Fan, Z., Mao, C., Wu, Y. and Xu, J. (2023). Spectral graph matching and regularized quadratic relaxations I Algorithm and Gaussian analysis. *Found. Comput. Math.* **23** 1511–1565. MR4649430 https://doi.org/10.1007/s10208-022-09570-y

[23] Fan, Z., Mao, C., Wu, Y. and Xu, J. (2023). Spectral graph matching and regularized quadratic relaxations II. *Found. Comput. Math.* **23** 1567–1617. MR4649431 https://doi.org/10.1007/s10208-022-09575-7

[24] Feizi, S., Quon, G., Medard, M., Kellis, M. and Jadbabaie, A. (2016). Spectral alignment of networks. Preprint. Available at arXiv:1602.04181.

[25] Gamarnik, D., Jagannath, A. and Wein, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Computer Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

[26] Ganassali, L., Lelarge, M. and Massoulié, L. (2024). Correlation detection in trees for planted graph alignment. *Ann. Appl. Probab.* **34** 2799–2843. MR4757492 https://doi.org/10.1214/23-aap2020

[27] Ganassali, L. and Massoulié, L. (2020). From tree matching to sparse graph alignment. In *Proceedings of the Conference on Learning Theory* (*COLT*) **33** 1633–1665. PMLR.

[28] Ganassali, L., Massoulié, L. and Semerjian, G. (2024). Statistical limits of correlation detection in trees. *Ann. Appl. Probab.* **34** 3701–3734. MR4783028 https://doi.org/10.1214/23-aap2048

[29] Haghighi, A., Ng, A. and Manning, C. (2005). Robust textual inference via graph matching. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (*EMNLP*) 387–394. ACL.

[30] Hall, G. and Massoulié, L. (2023). Partial recovery in the graph alignment problem. *Oper. Res.* **71** 259–272. MR4585566

[31] Hopkins, S. (2018). *Statistical Inference and the Sum of Squares Method*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Cornell University. MR3864930

[32] Hopkins, S. B., Kothari, P. K., Potechin, A., Raghavendra, P., Schramm, T. and Steurer, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Computer Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

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

[34] Kazemi, E., Hassani, S. H. and Grossglauser, M. (2015). Growing a graph matching from a handful of seeds. *Proc. VLDB Endow.* **8** 1010–1021.

[35] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation. Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

[36] LYZINSKI, V., FISHKIND, D. E. and PRIEBE, C. E. (2014). Seeded graph matching for correlated Erdős-Rényi graphs. *J. Mach. Learn. Res.* **15** 3513–3540. MR3291408

[37] MAO, C., RUDELSON, M. and TIKHOMIROV, K. (2021). Random graph matching with improved noise robustness. In *Proceedings of the Conference on Learning Theory (COLT)* **35** 3296–3329. PMLR.

[38] MAO, C., RUDELSON, M. and TIKHOMIROV, K. (2023). Exact matching of random graphs with constant correlation. *Probab. Theory Related Fields* **186** 327–389. MR4586222 https://doi.org/10.1007/s00440-022-01184-3

[39] MAO, C., WU, Y., XU, J. and YU, S. H. (2023). Random graph matching at Otter's threshold via counting chandeliers. In *STOC'23—Proceedings of the 55th Annual ACM Symposium on Theory of Computing* 1345–1356. ACM, New York. MR4617472 https://doi.org/10.1145/3564246.3585156

[40] MAO, C., WU, Y., XU, J. and YU, S. H. (2024). Testing network correlation efficiently via counting trees. *Ann. Statist.* **52** 2483–2505. MR4841674 https://doi.org/10.1214/23-AOS2261

[41] MONTANARI, A. and WEIN, A. S. (2025). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *Probab. Theory Related Fields* **191** 181–233. MR4869255 https://doi.org/10.1007/s00440-024-01322-z

[42] MOSSEL, E. and XU, J. (2020). Seeded graph matching via large neighborhood statistics. *Random Structures Algorithms* **57** 570–611. MR4144076 https://doi.org/10.1002/rsa.20934

[43] NARAYANAN, A. and SHMATIKOV, V. (2008). Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy* 111–125. IEEE.

[44] NARAYANAN, A. and SHMATIKOV, V. (2009). De-anonymizing social networks. In *IEEE Symposium on Security and Privacy* 173–187. IEEE.

[45] OTTER, R. (1948). The number of trees. *Ann. of Math.* (2) **49** 583–599. MR0025715 https://doi.org/10.2307/1969046

[46] PEDARSANI, P. and GROSSGLAUSER, M. (2011). On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1235–1243. ACM, New York.

[47] PICCIOLI, G., SEMERJIAN, G., SICURO, G. and ZDEBOROVÁ, L. (2022). Aligning random graphs with a sub-tree similarity message-passing algorithm. *J. Stat. Mech. Theory Exp.* **6** Paper No. 063401, 44. MR4483763 https://doi.org/10.1088/1742-5468/ac70d2

[48] SCHRAMM, T. and WEIN, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann. Statist.* **50** 1833–1858. MR4441142 https://doi.org/10.1214/22-aos2179

[49] SHIRANI, F., GARG, S. and ERKIP, E. (2017). Seeded graph matching: Efficient algorithms and theoretical guarantees. In *Asilomar Conference on Signals, Systems, and Computers* **51** 253–257.

[50] SINGH, R., XU, J. and BERGER, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA* **105** 12763–12768.

[51] VOGELSTEIN, J. T., CONROY, J. M., LYZINSKI, V., PODRAZIK, L. J., KRATZER, S. G., HARLEY, E. T., FISHKIND, D. E., VOGELSTEIN, R. J. and PRIEBE, C. E. (2015). Fast approximate quadratic programming for graph matching. *PLoS ONE* **10** 1–17.

[52] WEIN, A. S. (2021). Optimal low-degree hardness of maximum independent set. *Math. Stat. Learn.* **4** 221–251. MR4383734

[53] WU, Y., XU, J. and YU, S. H. (2022). Settling the sharp reconstruction thresholds of random graph matching. *IEEE Trans. Inf. Theory* **68** 5391–5417. MR4476403 https://doi.org/10.1109/tit.2022.3169005

[54] WU, Y., XU, J. and YU, S. H. (2023). Testing correlation of unlabeled random graphs. *Ann. Appl. Probab.* **33** 2519–2558. MR4612649 https://doi.org/10.1214/22-aap1786

[55] YARTSEVA, L. and GROSSGLAUSER, M. (2013). On the performance of percolation graph matching. In *Proceedings of the ACM Conference on Online Social Networks* **1** 119–130.

The Annals of Statistics
2025, Vol. 53, No. 5, 1857–1885
https://doi.org/10.1214/25-AOS2521

THE FUNCTIONAL GRAPHICAL LASSO

BY KARTIK G. WAGHMARE[1,a], TOMAS MASAK[2,b] AND VICTOR M. PANARETOS[3,c]

[1]*Seminar for Statistics, Eidgenössische Technische Hochschule Zürich,* [a]*kartik.waghmare@math.ethz.ch*
[2]*Institute for Statistics and Mathematics, Vienna University of Economics and Business,* [b]*tomas.masak@wu.ac.at*
[3]*Institute of Mathematics, École Polytechnique Fédérale de Lausanne,* [c]*victor.panaretos@epfl.ch*

We consider the problem of recovering conditional independence relationships between p jointly distributed Hilbertian random elements given n realizations thereof. We operate in the sparse high-dimensional regime, where $n \ll p$ and no element is related to more than $d \ll p$ other elements. In this context, we propose an infinite-dimensional generalization of the graphical lasso. We prove model selection consistency under natural assumptions and extend many classical results to infinite dimensions. In particular, we do not make additional structural assumptions. The plug-in nature of our method makes it applicable to heterogeneous data measured under any observational regime, whether sparse or dense, and indifferent to serial dependence between samples. In addition, it does not require dimensionality reduction by truncation. Importantly, our method can be understood as naturally arising from a coherent maximum likelihood philosophy.

REFERENCES

ASTON, J. A. D., PIGOLI, D. and TAVAKOLI, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.* **45** 1431–1461. MR3670184 https://doi.org/10.1214/16-AOS1495

BAKER, C. R. (1973). Joint measures and cross-covariance operators. *Trans. Amer. Math. Soc.* **186** 273–289. MR0336795 https://doi.org/10.2307/1996566

BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391 https://doi.org/10.1090/surv/062

BOYD, S., PARIKH, N., CHU, E., PELEATO, B., ECKSTEIN, J. et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

CHEN, X. and YANG, Y. (2021). Hanson-Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data. *Bernoulli* **27** 586–614. MR4177382 https://doi.org/10.3150/20-BEJ1251

CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24** 1571–1596. MR3308652

DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. MR3164871 https://doi.org/10.1111/rssb.12033

DE PRADO, M. L. (2018). *Advances in Financial Machine Learning*. Wiley, New York.

DONINI, M., MONTEIRO, J. M., PONTIL, M., HAHN, T., FALLGATTER, A. J., SHAWE-TAYLOR, J. and MOURÃO-MIRANDA, J. (2019). Combining heterogeneous data sources for neuroimaging based diagnosis: Re-weighting and selecting what is important. *NeuroImage* **195** 215–231. https://doi.org/10.1016/j.neuroimage.2019.01.053

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

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

GIULIANO ANTONINI, R. (1997). Sub-Gaussian random variables in Hilbert spaces. *Rend. Semin. Mat. Univ. Padova* **98** 89–99. MR1492970

GOHBERG, I., GOLDBERG, S. and KRUPNIK, N. (2000). *Traces and Determinants of Linear Operators. Operator Theory: Advances and Applications* **116**. Birkhäuser, Basel. MR1744872 https://doi.org/10.1007/978-3-0348-8401-3

MSC2020 subject classifications. Primary 62R10, 62A09; secondary 62H22, 62J07.
Key words and phrases. Graphical models, functional data analysis, correlation operator.

GUO, S., LI, D., QIAO, X. and WANG, Y. (2025). From sparse to dense functional data in high dimensions: Revisiting phase transitions from a non-asymptotic perspective. *J. Mach. Learn. Res.* **26** Paper No. [15], 40. MR4870571

HANKE, M. (2017). *A Taste of Inverse Problems: Basic Theory and Examples*. SIAM, Philadelphia. MR3696342 https://doi.org/10.1137/1.9781611974942.ch1

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

KNEIP, A. and LIEBL, D. (2020). On the optimal reconstruction of partially observed functional data. *Ann. Statist.* **48** 1692–1717. MR4124340 https://doi.org/10.1214/19-AOS1864

KOVÁCS, S., RUCKSTUHL, T., OBRIST, H. and BÜHLMANN, P. (2021). Graphical elastic net and target matrices: Fast algorithms and software for sparse precision matrix estimation. Preprint. Available at arXiv:2101.02148.

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

LEE, K.-Y., JI, D., LI, L., CONSTABLE, T. and ZHAO, H. (2023a). Conditional functional graphical models. *J. Amer. Statist. Assoc.* **118** 257–271. MR4571120 https://doi.org/10.1080/01621459.2021.1924178

LEE, K.-Y., LI, L., LI, B. and ZHAO, H. (2023b). Nonparametric functional graphical modeling through functional additive regression operator. *J. Amer. Statist. Assoc.* **118** 1718–1732. MR4646601 https://doi.org/10.1080/01621459.2021.2006667

LI, B. and SOLEA, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Amer. Statist. Assoc.* **113** 1637–1655. MR3902235 https://doi.org/10.1080/01621459.2017.1356726

LOÈVE, M. (2017). *Probab. Theory Related Fields*. Dover, New York.

MASAK, T. and PANARETOS, V. M. (2023). Random surface covariance estimation by shifted partial tracing. *J. Amer. Statist. Assoc.* **118** 2562–2574. MR4681604 https://doi.org/10.1080/01621459.2022.2061982

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

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523 https://doi.org/10.1111/j.1467-9868.2010.00740.x

QIAO, X., GUO, S. and JAMES, G. M. (2019). Functional graphical models. *J. Amer. Statist. Assoc.* **114** 211–222. MR3941249 https://doi.org/10.1080/01621459.2017.1390466

QIAO, X., QIAN, C., JAMES, G. M. and GUO, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika* **107** 415–431. MR4108937 https://doi.org/10.1093/biomet/asz072

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

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

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

SIMON, B. (1977). Notes on infinite determinants of Hilbert space operators. *Adv. Math.* **24** 244–273. MR0482328 https://doi.org/10.1016/0001-8708(77)90057-3

SOLEA, E. and DETTE, H. (2022). Nonparametric and high-dimensional functional graphical models. *Electron. J. Stat.* **16** 6175–6231. MR4515718 https://doi.org/10.1214/22-ejs2087

STONE, M. (1987). *Coordinate-Free Multivariable Statistics*. *Oxford Statistical Science Series* **2**. The Clarendon Press, New York. MR0916472

STROOCK, D. W. (2023). *Gaussian Measures in Finite and Infinite Dimensions*. *Universitext*. Springer, Cham. MR4573029 https://doi.org/10.1007/978-3-031-23122-3

TSAI, K., ZHAO, B., KOYEJO, S. and KOLAR, M. (2024). Latent multimodal functional graphical model estimation. *J. Amer. Statist. Assoc.* **119** 2217–2229. MR4797935 https://doi.org/10.1080/01621459.2023.2252142

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

WAGHMARE, K. G. and PANARETOS, V. M. (2025). Continuously indexed graphical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **87** 211–231. MR4862096 https://doi.org/10.1093/jrsssb/qkae086

WAGHMARE, K. G., MASAK, T. and PANARETOS, V. M. (2025). Supplement to “The functional graphical lasso.” https://doi.org/10.1214/25-AOS2521SUPP

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

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

ZAPATA, J., OH, S. Y. and PETERSEN, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika* **109** 665–681. MR4472841 https://doi.org/10.1093/biomet/asab046

ZHAO, B., WANG, Y. S. and KOLAR, M. (2019). Direct estimation of differential functional graphical models. *Adv. Neural Inf. Process. Syst.* **32**.

ZHAO, B., ZHAI, P. S., WANG, Y. S. and KOLAR, M. (2024). High-dimensional functional graphical model structure learning via neighborhood selection approach. *Electron. J. Stat.* **18** 1042–1129. MR4718467 https://doi.org/10.1214/24-ejs2219

The Annals of Statistics
2025, Vol. 53, No. 5, 1886–1912
https://doi.org/10.1214/25-AOS2522

DEEP HORSESHOE GAUSSIAN PROCESSES

BY ISMAËL CASTILLO[1,a] AND THIBAULT RANDRIANARISOA[2,b]

[1]*Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université,* [a]*ismael.castillo@upmc.fr*
[2]*Department of Statistical Sciences and Department of Computer and Mathematical Sciences, University of Toronto,* [b]*t.randrianarisoa@utoronto.ca*

Deep Gaussian processes have recently been proposed as natural objects to fit, similarly to deep neural networks, possibly complex features present in modern data samples, such as compositional structures. Adopting a Bayesian nonparametric approach, it is natural to use deep Gaussian processes as prior distributions, and use the corresponding posterior distributions for statistical inference. We introduce the deep Horseshoe Gaussian process Deep–HGP, a new simple prior based on deep Gaussian processes with a squared-exponential kernel, that in particular enables data-driven choices of the key lengthscale parameters. For nonparametric regression with random design, we show that the associated posterior distribution recovers the unknown true regression curve optimally in terms of quadratic loss, up to a logarithmic factor, in an adaptive way. The convergence rates are *simultaneously* adaptive to both the smoothness of the regression function and to its structure in terms of compositions. The dependence of the rates in terms of dimension are explicit, allowing in particular for input spaces of dimension increasing with the number of observations.

REFERENCES

[1] ABRAHAM, K. and DEO, N. (2023). Deep Gaussian process priors for Bayesian inference in nonlinear inverse problems. Preprint. Available at arXiv:2312.14294.

[2] AGAPIOU, S. and CASTILLO, I. (2024). Heavy-tailed Bayesian nonparametric adaptation. *Ann. Statist.* **52** 1433–1459. MR4804815 https://doi.org/10.1214/24-aos2397

[3] BACHOC, F. and LAGNOUX, A. (2025). Posterior contraction rates for constrained deep Gaussian processes in density estimation and classification. *Comm. Statist. Theory Methods* **54** 774–811. MR4841142 https://doi.org/10.1080/03610926.2024.2321185

[4] BAI, J., SONG, Q. and CHENG, G. (2020). Efficient variational inference for sparse deep learning with theoretical guarantee. In *Advances in Neural Information Processing Systems* **33** 466–476.

[5] BHATTACHARYA, A., PATI, D. and DUNSON, D. (2014). Anisotropic function estimation using multi-bandwidth Gaussian processes. *Ann. Statist.* **42** 352–381. MR3189489 https://doi.org/10.1214/13-AOS1192

[6] BHATTACHARYA, A., PATI, D. and YANG, Y. (2019). Bayesian fractional posteriors. *Ann. Statist.* **47** 39–66. MR3909926 https://doi.org/10.1214/18-AOS1712

[7] CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

[8] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. MR2471287 https://doi.org/10.1214/08-EJS273

[9] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. MR2875753 https://doi.org/10.1007/s00440-010-0316-5

[10] CASTILLO, I. and EGELS, P. (2025). Posterior and variational inference for deep neural networks with heavy-tailed weights. *J. Mach. Learn. Res.* **26**. To appear. Available at arXiv:2406.03369.

[11] CASTILLO, I., KERKYACHARIAN, G. and PICARD, D. (2014). Thomas Bayes' walk on manifolds. *Probab. Theory Related Fields* **158** 665–710. MR3176362 https://doi.org/10.1007/s00440-013-0493-0

[12] CASTILLO, I. and RANDRIANARISOA, T. (2025). Supplement to "Deep Horseshoe Gaussian Processes." https://doi.org/10.1214/25-AOS2522SUPP

MSC2020 subject classifications. 62G20.
Key words and phrases. Bayesian nonparametrics, deep Gaussian processes, multibandwidth Gaussian process priors, fractional posteriors, high-dimensional regression.

[13] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[14] CATONI, O. (2004). *Statistical Learning Theory and Stochastic Optimization*. *Lecture Notes in Math.* **1851**. Springer, Berlin. MR2163920 https://doi.org/10.1007/b99352

[15] CHÉRIEF-ABDELLATIF, B. (2020). Convergence rates of variational inference in sparse deep learning. In *Proceedings of the* 37*th International Conference on Machine Learning*, *ICML* 2020. *Proceedings of Machine Learning Research* **119** 1831–1842.

[16] DAMIANOU, A. and LAWRENCE, N. D. (2013). Deep Gaussian processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*. *Proceedings of Machine Learning Research* **31** 207–215.

[17] DE G. MATTHEWS, A. G., HRON, J., ROWLAND, M., TURNER, R. E. and GHAHRAMANI, Z. (2018). Gaussian process behaviour in wide deep neural networks. In *International Conference on Learning Representations*.

[18] DUTORDOIR, V., SALIMBENI, H., HAMBRO, E., MCLEOD, J., LEIBFRIED, F., ARTEMEV, A., VAN DER WILK, M., HENSMAN, J., DEISENROTH, M. P. et al. (2021). GPflux: A library for deep Gaussian processes. Preprint. Available at arXiv:2104.05674.

[19] FINOCCHIO, G. and SCHMIDT-HIEBER, J. (2023). Posterior contraction for deep Gaussian process priors. *J. Mach. Learn. Res.* **24** Paper No. [66], 49 pp. MR4582488

[20] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

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

[22] GIORDANO, M., RAY, K. and SCHMIDT-HIEBER, J. (2022). On the inability of Gaussian process regression to optimally learn compositional functions. In *Advances in Neural Information Processing Systems* **35** 22341–22353.

[23] HAZAN, T. and JAAKKOLA, T. (2015). Steps toward deep kernel methods from infinite neural networks. Preprint. Available at arXiv:1508.05133.

[24] JIANG, S. and TOKDAR, S. T. (2021). Variable selection consistency of Gaussian process regression. *Ann. Statist.* **49** 2491–2505. MR4338372 https://doi.org/10.1214/20-aos2043

[25] 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

[26] KRUIJER, W. and VAN DER VAART, A. (2013). Analyzing posteriors by the information inequality. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes*. *Inst. Math. Stat.* (*IMS*) *Collect.* **9** 227–240. IMS, Beachwood, OH. MR3202636 https://doi.org/10.1214/12-IMSCOLL916

[27] L'HUILLIER, A., TRAVIS, L., CASTILLO, I. and RAY, K. (2023). Semiparametric inference using fractional posteriors. *J. Mach. Learn. Res.* **24** Paper No. [389], 61 pp. MR4720845

[28] MAI, T. T. (2024). Adaptive posterior concentration rates for sparse high-dimensional linear regression with random design and unknown error variance. Preprint. Available at arXiv:2405.19016.

[29] MORIARTY-OSBORNE, C. and TECKENTRUP, A. L. (2025). Convergence rates of non-stationary and deep Gaussian process regression. *Found. Data Sci.* To appear. https://doi.org/10.3934/fods.2025004

[30] NEAL, R. M. (2012). *Bayesian Learning for Neural Networks* **118**. Springer, Berlin.

[31] OHN, I. and LIN, L. (2024). Adaptive variational Bayes: Optimality, computation and applications. *Ann. Statist.* **52** 335–363. MR4718418 https://doi.org/10.1214/23-aos2349

[32] PATI, D., BHATTACHARYA, A. and CHENG, G. (2015). Optimal Bayesian estimation in random covariate design with a rescaled Gaussian process prior. *J. Mach. Learn. Res.* **16** 2837–2851. MR3450525

[33] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[34] ROCKOVÁ, V. and POLSON, N. (2018). Posterior concentration for sparse deep learning. In *Annual Conference on Neural Information Processing Systems* 2018, *NeurIPS* 2018 938–949.

[35] SALIMBENI, H. and DEISENROTH, M. (2017). Doubly stochastic variational inference for deep Gaussian processes. In *Advances in Neural Information Processing Systems* **30**.

[36] SALIMBENI, H., DUTORDOIR, V., HENSMAN, J. and DEISENROTH, M. (2019). Deep Gaussian processes with importance-weighted variational inference. In *Proceedings of the* 36*th International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **97** 5589–5598.

[37] SAUER, A. (2022). deepgp: Deep Gaussian processes using MCMC. R package version 1.1.1.

[38] SAUER, A., COOPER, A. and GRAMACY, R. B. (2023). Vecchia-approximated deep Gaussian processes for computer experiments. *J. Comput. Graph. Statist.* **32** 824–837. MR4641462 https://doi.org/10.1080/10618600.2022.2129662

[39] 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

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

[41] TANG, T., WU, N., CHENG, X. and DUNSON, D. (2024). Adaptive Bayesian regression on data with low intrinsic dimensionality. Preprint. Available at arXiv:2407.09286.

[42] TECKENTRUP, A. L. (2020). Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantificat.* **8** 1310–1337. MR4164077 https://doi.org/10.1137/19M1284816

[43] VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.* **12** 1221–1274. MR3724985 https://doi.org/10.1214/17-BA1065

[44] VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Stat.* **8** 2585–2618. MR3285877 https://doi.org/10.1214/14-EJS962

[45] VAN DER VAART, A. and VAN ZANTEN, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.* **1** 433–448. MR2357712 https://doi.org/10.1214/07-EJS098

[46] VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028

[47] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 https://doi.org/10.1214/009053607000000613

[48] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

[49] YANG, Y., BHATTACHARYA, A. and PATI, D. (2017). Frequentist coverage and sup-norm convergence rate in Gaussian process regression. Preprint. Available at arXiv:1708.04753.

[50] YANG, Y. and DUNSON, D. B. (2016). Bayesian manifold regression. *Ann. Statist.* **44** 876–905. MR3476620 https://doi.org/10.1214/15-AOS1390

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

[52] YU, W., WADE, S., BONDELL, H. D. and AZIZI, L. (2023). Nonstationary Gaussian process discriminant analysis with variable selection for high-dimensional functional data. *J. Comput. Graph. Statist.* **32** 588–600. MR4592932 https://doi.org/10.1080/10618600.2022.2098136

[53] ZHANG, T. (2006). From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Ann. Statist.* **34** 2180–2210. MR2291497 https://doi.org/10.1214/009053606000000704

[54] ZHAO, Z., EMZIR, M. and SÄRKKÄ, S. (2021). Deep state-space Gaussian processes. *Stat. Comput.* **31** Paper No. 75, 26 pp. MR4319943 https://doi.org/10.1007/s11222-021-10050-6

The Annals of Statistics
2025, Vol. 53, No. 5, 1913–1937
https://doi.org/10.1214/25-AOS2523

HIGH-DIMENSIONAL STATISTICAL INFERENCE FOR LINKAGE DISEQUILIBRIUM SCORE REGRESSION AND ITS CROSS-ANCESTRY EXTENSIONS

BY FEI XUE[1,a] AND BINGXIN ZHAO[2,b]

[1]*Department of Statistics, Purdue University,* [a]*feixue@purdue.edu*
[2]*Department of Statistics and Data Science, University of Pennsylvania,* [b]*bxzhao@upenn.edu*

Linkage disequilibrium score regression (LDSC) has emerged as an essential tool for genetic and genomic analyses of complex traits, utilizing high-dimensional data derived from genome-wide association studies (GWAS). LDSC computes the linkage disequilibrium (LD) scores using an external reference panel, and integrates the LD scores with only summary data from the original GWAS. In this paper, we investigate LDSC within a fixed-effect data integration framework, underscoring its ability to merge multisource GWAS data and reference panels. In particular, we take account of the genome-wide dependence among the high-dimensional GWAS summary statistics, along with the block-diagonal dependence pattern in estimated LD scores. Our analysis uncovers several key factors of both the original GWAS and reference panel datasets that determine the performance of LDSC. We show that it is relatively feasible for LDSC-based estimators to achieve asymptotic normality when applied to genome-wide genetic variants (e.g., in genetic variance and covariance estimation), whereas it becomes considerably challenging when we focus on a much smaller subset of genetic variants (e.g., in partitioned heritability analysis). Moreover, by modeling the disparities in LD patterns across different populations, we show that LDSC can be expanded to conduct cross-ancestry analyses using data from genetically distinct global populations. We validate our theoretical findings through extensive numerical evaluations using real genetic data from the UK Biobank study.

REFERENCES

[1] 1000-GENOMES-CONSORTIUM (2015). A global reference for human genetic variation. *Nature* **526** 68–74.
[2] BERISA, T. and PICKRELL, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32** 283–285.
[3] BONNET, A., LÉVY-LEDUC, C., GASSIAT, E., TORO, R. and BOURGERON, T. (2018). Improving heritability estimation by a variable selection approach in sparse high dimensional linear mixed models. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **67** 813–839. MR3832252 https://doi.org/10.1111/rssc.12261
[4] BOYLE, E. A., LI, Y. I. and PRITCHARD, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169** 1177–1186.
[5] BULIK-SULLIVAN, B., FINUCANE, H. K., ANTTILA, V., GUSEV, A., DAY, F. R., LOH, P.-R., DUNCAN, L., PERRY, J. R., PATTERSON, N. et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47** 1236–1241.
[6] BULIK-SULLIVAN, B. K., LOH, P.-R., FINUCANE, H. K., RIPKE, S., YANG, J., PATTERSON, N., DALY, M. J., PRICE, A. L., NEALE, B. M. et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47** 291–295.
[7] BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562** 203–209.
[8] CAI, M., XIAO, J., ZHANG, S., WAN, X., ZHAO, H., CHEN, G. and YANG, C. (2021). A unified framework for cross-population trait prediction by leveraging the genetic correlation of polygenic traits. *Amer. J. Hum. Genet.* **108** 632–655.

MSC2020 subject classifications. Primary 62F12, 62J05; secondary 62P10.
Key words and phrases. Asymptotic normality, consistency, genetic covariance, genetic variance, GWAS summary data, LDSC, UK Biobank.

[9] CAI, T. T. and GUO, Z. (2020). Semisupervised inference for explained variance in high dimensional linear regression and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 391–419. MR4084169

[10] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

[11] DICKER, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269–284. MR3215347 https://doi.org/10.1093/biomet/ast065

[12] FINUCANE, H. K., BULIK-SULLIVAN, B., GUSEV, A., TRYNKA, G., RESHEF, Y., LOH, P.-R., ANTTILA, V., XU, H., ZANG, C. et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47** 1228–1235.

[13] FINUCANE, H. K., RESHEF, Y. A., ANTTILA, V., SLOWIKOWSKI, K., GUSEV, A., BYRNES, A., GAZAL, S., LOH, P.-R., LAREAU, C. et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50** 621–629.

[14] GAZAL, S., FINUCANE, H. K., FURLOTTE, N. A., LOH, P.-R., PALAMARA, P. F., LIU, X., SCHOECH, A., BULIK-SULLIVAN, B., NEALE, B. M. et al. (2017). Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49** 1421–1427.

[15] GE, T., CHEN, C.-Y., NI, Y., FENG, Y.-C. A. and SMOLLER, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10** 1–10.

[16] GIANNAKOPOULOU, O., LIN, K., MENG, X., SU, M.-H., KUO, P.-H., PETERSON, R. E., AWASTHI, S., MOSCATI, A., COLEMAN, J. R. et al. (2021). The genetic architecture of depression in individuals of East Asian ancestry: A genome-wide association study. *JAMA Psychiatr.* **78** 1258–1269.

[17] GUO, Z., WANG, W., CAI, T. T. and LI, H. (2019). Optimal estimation of genetic relatedness in high-dimensional linear models. *J. Amer. Statist. Assoc.* **114** 358–369. MR3941260 https://doi.org/10.1080/01621459.2017.1407774

[18] HAPMAP3-CONSORTIUM (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* **467** 52–58.

[19] HAUBERG, M. E., CREUS-MUNCUNILL, J., BENDL, J., KOZLENKOV, A., ZENG, B., CORWIN, C., CHOWDHURY, S., KRANZ, H., HURD, Y. L. et al. (2020). Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat. Commun.* **11** 5581.

[20] HU, Y., LI, M., LU, Q., WENG, H., WANG, J., ZEKAVAT, S., YU, Z., LI, B., GU, J. et al. (2019). A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.* **51** 568–576.

[21] JIANG, J., JIANG, W., PAUL, D., ZHANG, Y. and ZHAO, H. (2023). High-dimensional asymptotic behavior of inference based on GWAS summary statistic. *Statist. Sinica* **33** 1555–1576. MR4596852

[22] JIANG, J., LI, C., PAUL, D., YANG, C. and ZHAO, H. (2016). On high-dimensional misspecified mixed model analysis in genome-wide association study. *Ann. Statist.* **44** 2127–2160. MR3546446 https://doi.org/10.1214/15-AOS1421

[23] JIANG, L., ZHENG, Z., QI, T., KEMPER, K., WRAY, N., VISSCHER, P. and YANG, J. (2019). A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51** 1749–1755.

[24] LAM, M., CHEN, C.-Y., LI, Z., MARTIN, A. R., BRYOIS, J., MA, X., GASPAR, H., IKEDA, M., BENYAMIN, B. et al. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51** 1670–1678.

[25] LEE, J. J., MCGUE, M., IACONO, W. G. and CHOW, C. C. (2018). The accuracy of LD score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genet. Epidemiol.* **42** 783–795.

[26] LU, Q., LI, B., OU, D., ERLENDSDOTTIR, M., POWLES, R. L., JIANG, T., HU, Y., CHANG, D., JIN, C. et al. (2017). A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. *Amer. J. Hum. Genet.* **101** 939–964.

[27] LUO, Y., LI, X., WANG, X., GAZAL, S., MERCADER, J. M., TEAM, M. R., CONSORTIUM, S. T. D., NEALE, B. M., FLOREZ, J. C. et al. (2021). Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum. Mol. Genet.* **30** 1521–1534.

[28] MA, R. and DICKER, L. H. (2019). The Mahalanobis kernel for heritability estimation in genome-wide association studies: Fixed-effects and random-effects methods. arXiv preprint. Available at arXiv:1901.02936.

[29] MAK, T. S. H., PORSCH, R. M., CHOI, S. W., ZHOU, X. and SHAM, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41** 469–480.

[30] MOMIN, M. M., SHIN, J., LEE, S., TRUONG, B., BENYAMIN, B. and LEE, S. H. (2023). A method for an unbiased estimate of cross-ancestry genetic correlation using individual-level data. *Nat. Commun.* **14** 722.

[31] NATIONAL-ACADEMIES-SCIENCES (2023). Using Population Descriptors in Genetics and Genomics Research: a New Framework for an Evolving Field, Washington (DC): National Academies Press (US). March 14, 2023.

[32] Ni, G., Moser, G., Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K.-H., Holmans, P. A., Lee, P. et al. (2018). Estimation of genetic correlation via linkage disequilibrium score regression and genomic restricted maximum likelihood. *Amer. J. Hum. Genet.* **102** 1185–1194.

[33] Ning, Z., Pawitan, Y. and Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* **52** 859–864.

[34] O'Connor, L. J. (2021). The distribution of common-variant effect sizes. *Nat. Genet.* **53** 1243–1249.

[35] O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N. and Price, A. L. (2019). Extreme polygenicity of complex traits is explained by negative selection. *Amer. J. Hum. Genet.* **105** 456–476.

[36] Palmer, D. S., Zhou, W., Abbott, L., Wigdor, E. M., Baya, N., Churchhouse, C., Seed, C., Poterba, T., King, D. et al. (2023). Analysis of genetic dominance in the UK Biobank. *Science* **379** 1341–1348.

[37] Pasaniuc, B. and Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18** 117–127.

[38] Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C.-Y., Popejoy, A. B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R. J. et al. (2019). Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. *Cell* **179** 589–603.

[39] Power, R. A., Steinberg, S., Bjornsdottir, G., Rietveld, C. A., Abdellaoui, A., Nivard, M. M., Johannesson, M., Galesloot, T. E., Hottenga, J. J. et al. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat. Neurosci.* **18** 953–955.

[40] Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38** 904–909.

[41] Ridge, P. G., Mukherjee, S., Crane, P. K., Kauwe, J. S. and Consortium, A. D. G. (2013). Alzheimer's disease: Analyzing the missing heritability. *PLoS ONE* **8** e79771.

[42] Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K. et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53** 1415–1424.

[43] Schoech, A. P., Jordan, D. M., Loh, P.-R., Gazal, S., O'Connor, L. J., Balick, D. J., Palamara, P. F., Finucane, H. K., Sunyaev, S. R. et al. (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* **10** 790.

[44] Schwartzman, A., Schork, A. J., Zablocki, R. and Thompson, W. K. (2019). A simple, consistent estimator of SNP heritability from genome-wide association studies. *Ann. Appl. Stat.* **13** 2509–2538. MR4037439 https://doi.org/10.1214/19-aoas1291

[45] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality: Complete samples. *Biometrika* **52** 591–611. MR0205384 https://doi.org/10.1093/biomet/52.3-4.591

[46] Shi, H., Burch, K. S., Johnson, R., Freund, M. K., Kichaev, G., Mancuso, N., Manuel, A. M., Dong, N. and Pasaniuc, B. (2020). Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. *Amer. J. Hum. Genet.* **106** 805–817.

[47] Smith, S. P., Darnell, G., Udwin, D., Harpak, A., Ramachandran, S. and Crawford, L. (2023). Accounting for statistical non-additive interactions enables the recovery of missing heritability from GWAS summary statistics. *bioRxiv*.

[48] Song, S., Jiang, W., Zhang, Y., Hou, L. and Zhao, H. (2022). Leveraging LD eigenvalue regression to improve the estimation of SNP heritability and confounding inflation. *Amer. J. Hum. Genet.* **109** 802–811.

[49] Speed, D. and Balding, D. (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51** 277–284.

[50] Tashman, K. C., Cui, R., O'Connor, L. J., Neale, B. M. and Finucane, H. K. (2021). Significance testing for small annotations in stratified LD-score regression. *MedRxiv*.

[51] Timpson, N. J., Greenwood, C. M., Soranzo, N., Lawson, D. J. and Richards, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19** 110–125.

[52] Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T. and Posthuma, D. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* **1** 1–21.

[53] Van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. and Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nat. Rev. Genet.* **20** 567–581.

[54] Verzelen, N. and Gassiat, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli* **24** 3683–3710. MR3788186 https://doi.org/10.3150/17-BEJ975

[55] VISSCHER, P. M., WRAY, N. R., ZHANG, Q., SKLAR, P., MCCARTHY, M. I., BROWN, M. A. and YANG, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *Amer. J. Hum. Genet.* **101** 5–22.

[56] WANG, J. and LI, H. (2022). Estimation of genetic correlation with summary association statistics. *Biometrika* **109** 421–438. MR4430966 https://doi.org/10.1093/biomet/asab030

[57] WHARRIE, S., YANG, Z., RAJ, V., MONTI, R., GUPTA, R., WANG, Y., MARTIN, A., O'CONNOR, L. J., KASKI, S. et al. (2023). HAPNEST: Efficient, large-scale generation and evaluation of synthetic datasets for genotypes and phenotypes. *Bioinformatics* **39** btad535.

[58] XUE, F. and ZHAO, B. (2025). Supplement to "High-dimensional statistical inference for linkage disequilibrium score regression and its cross-ancestry extensions." https://doi.org/10.1214/25-AOS2523SUPPA, https://doi.org/10.1214/25-AOS2523SUPPB

[59] XUE, H., SHEN, X. and PAN, W. (2023). Causal inference in transcriptome-wide association studies with invalid instruments and GWAS summary data. *J. Amer. Statist. Assoc.* **118** 1525–1537. MR4646581 https://doi.org/10.1080/01621459.2023.2183127

[60] YANG, J., LEE, S. H., GODDARD, M. E. and VISSCHER, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Amer. J. Hum. Genet.* **88** 76–82.

[61] YANG, J., ZENG, J., GODDARD, M. E., WRAY, N. R. and VISSCHER, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49** 1304–1310.

[62] YE, T., SHAO, J. and KANG, H. (2021). Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Ann. Statist.* **49** 2079–2100. MR4319242 https://doi.org/10.1214/20-aos2027

[63] ZHANG, M. J., HOU, K., DEY, K. K., SAKAUE, S., JAGADEESH, K. A., WEINAND, K., TAYCHAMEEKIATCHAI, A., RAO, P., PISCO, A. O. et al. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat. Genet.* **54** 1572–1580.

[64] ZHANG, Y., CHENG, Y., JIANG, W., YE, Y., LU, Q. and ZHAO, H. (2021). Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Brief. Bioinform.* **22** bbaa442.

[65] ZHANG, Y., QI, G., PARK, J.-H. and CHATTERJEE, N. (2018). Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50** 1318–1326.

[66] ZHAO, B., ZHENG, S. and ZHU, H. (2024). On blockwise and reference panel-based estimators for genetic data prediction in high dimensions. *Ann. Statist.* **52** 948–965. MR4784065 https://doi.org/10.1214/24-aos2378

[67] ZHAO, B. and ZHU, H. (2022). On genetic correlation estimation with summary statistics from genome-wide association studies. *J. Amer. Statist. Assoc.* **117** 1–11. MR4399063 https://doi.org/10.1080/01621459.2021.1906684

[68] ZHAO, Q., WANG, J., HEMANI, G., BOWDEN, J. and SMALL, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Statist.* **48** 1742–1769. MR4124342 https://doi.org/10.1214/19-AOS1866

[69] ZHOU, W., KANAI, M., WU, K.-H. H., RASHEED, H., TSUO, K., HIRBO, J. B., WANG, Y., BHATTACHARYA, A., ZHAO, H. et al. (2022). Global Biobank Meta-analysis initiative: Powering genetic discovery across human disease. *Cell Genom.* **2** 100192.

[70] ZHOU, X. (2017). A unified framework for variance component estimation with summary statistics in genome-wide association studies. *Ann. Appl. Stat.* **11** 2027–2051. MR3743287 https://doi.org/10.1214/17-AOAS1052

[71] ZHU, H. and ZHOU, X. (2020). Statistical methods for SNP heritability estimation and partition: A review. *Comput. Struct. Biotechnol. J.* **18** 1557–1568.

The Annals of Statistics
2025, Vol. 53, No. 5, 1938–1962
https://doi.org/10.1214/25-AOS2525

STRUCTURED MATRIX LEARNING UNDER ARBITRARY ENTRYWISE DEPENDENCE AND ESTIMATION OF MARKOV TRANSITION KERNEL

BY JINHANG CHAI[a] AND JIANQING FAN[b]

Department of Operations Research and Financial Engineering, Princeton University, [a]*jhchai@princeton.edu,* [b]*jqfan@princeton.edu*

The problem of structured matrix estimation has been studied mostly under strong noise dependence assumptions. This paper considers a general framework of noisy low-rank-plus-sparse matrix recovery, where the noise matrix may come from any joint distribution with arbitrary dependence across entries. We propose an incoherent-constrained least-square estimator and prove its tightness both in the sense of deterministic lower bound and matching minimax risks under various noise distributions. To attain this, we establish a novel result asserting that the difference between two arbitrary low-rank incoherent matrices must spread energy out across its entries; in other words, it cannot be too sparse, which sheds light on the structure of incoherent low-rank matrices and may be of independent interest. We then showcase the applications of our framework to several important statistical machine learning problems. In the problem of estimating a structured Markov transition kernel, the proposed method achieves the minimax optimality and the result can be extended to estimating the conditional mean operator, a crucial component in reinforcement learning. The applications to multitask regression and structured covariance estimation are also presented. We propose an alternating minimization algorithm to approximately solve the potentially hard optimization problem. Numerical results corroborate the effectiveness of our method which typically converges in a few steps.

REFERENCES

AGARWAL, A., KAKADE, S., KRISHNAMURTHY, A. and SUN, W. (2020). Flambe: Structural complexity and representation learning of low rank mdps. *Adv. Neural Inf. Process. Syst.* **33** 20095–20107.

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947 https://doi.org/10.1214/12-AOS1000

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 https://doi.org/10.1109/tit.2022.3159085

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Stat.* **22** 327–351. MR0042664 https://doi.org/10.1214/aoms/1177729580

ANDO, R. K. and ZHANG, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *J. Mach. Learn. Res.* **6** 1817–1853. MR2249873

ARGYRIOU, A., EVGENIOU, T. and PONTIL, M. (2008). Convex multi-task feature learning. *Mach. Learn.* **73** 243–272.

BERESTYCKI, N. (2016). Mixing times of Markov chains: Techniques and examples. *ALEA Lat. Amer. J. Probab. Math. Stat.*.

BERTSEKAS, D. P. and TSITSIKLIS, J. N. (2014). *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA. Originally published by Prentice-Hall, Inc. in 1989. Includes corrections (1997). MR3587745

CAI, J.-F., LI, J. and XIA, D. (2023). Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *J. Amer. Statist. Assoc.* **118** 2588–2604. MR4681606 https://doi.org/10.1080/01621459.2022.2063131

MSC2020 subject classifications. Primary 62F10, 62M05; secondary 62B10.

Key words and phrases. Incoherent-constrained least-squares, deterministic bounds, minimax optimality, Markovian processes, frequency matrices, transition mean.

CANDES, E. and RECHT, B. (2012). Exact matrix completion via convex optimization. *Commun*. *ACM* **55** 111–119.

CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J*. *ACM* **58** Art. 11. MR2811000 https://doi.org/10.1145/1970392.1970395

CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc*. *IEEE* **98** 925–936.

CHAI, J. and FAN, J. (2025). Supplement to "Structured Matrix Learning under Arbitrary Entrywise Dependence and Estimation of Markov Transition Kernel." https://doi.org/10.1214/25-AOS2525SUPPA, https://doi.org/10.1214/25-AOS2525SUPPB

CHAN, S. O., DING, Q. and LI, S. H. (2021). Learning and testing irreducible Markov chains via the k-cover time. In *Algorithmic Learning Theory* 458–480. PMLR.

CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J*. *Optim*. **21** 572–596. MR2817479 https://doi.org/10.1137/090761793

CHEN, Y. (2015). Incoherence-optimal matrix completion. *IEEE Trans*. *Inf*. *Theory* **61** 2909–2923. MR3342311 https://doi.org/10.1109/TIT.2015.2415195

CHEN, Y., CHI, Y., FAN, J., MA, C. et al. (2021a). Spectral methods for data science: A statistical perspective. *Found*. *Trends Mach*. *Learn*. **14** 566–806.

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021b). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *Ann*. *Statist*. **49** 2948–2971. MR4338899 https://doi.org/10.1214/21-aos2066

CHERAPANAMJERI, Y., GUPTA, K. and JAIN, P. (2017). Nearly optimal robust matrix completion. In *International Conference on Machine Learning* 797–805. PMLR.

CHUNG, K. L. (1967). *Markov Chains*. Springer-Verlag, New York.

DAUBECHIES, I., DEVORE, R., FORNASIER, M. and GÜNTÜRK, C. S. (2010). Iteratively reweighted least squares minimization for sparse recovery. *Comm*. *Pure Appl*. *Math*. **63** 1–38. MR2588385 https://doi.org/10.1002/cpa.20303

DUAN, Y. and WANG, K. (2023). Adaptive and robust multi-task learning. *Ann*. *Statist*. **51** 2015–2039. MR4678794 https://doi.org/10.1214/23-aos2319

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J*. *Econometrics* **147** 186–197. MR2472991 https://doi.org/10.1016/j.jeconom.2008.09.017

FAN, J., 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

FAN, J., LI, Q. and WANG, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **79** 247–265. MR3597972 https://doi.org/10.1111/rssb.12166

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

FAN, J., WANG, K., ZHONG, Y. and ZHU, Z. (2021). Robust high-dimensional factor models with applications to statistical machine learning. *Statist*. *Sci*. **36** 303–327. MR4255196 https://doi.org/10.1214/20-sts785

FAN, J., WANG, W. and ZHONG, Y. (2019). Robust covariance estimation for approximate factor models. *J*. *Econometrics* **208** 5–22. MR3906959 https://doi.org/10.1016/j.jeconom.2018.09.003

FAN, J., WANG, W. and ZHU, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Ann*. *Statist*. **49** 1239–1266. MR4298863 https://doi.org/10.1214/20-aos1980

GU, Q., WANG, Z. W. and LIU, H. (2016). Low-rank and sparse structure pursuit via alternating minimization. In *Artificial Intelligence and Statistics* 600–609. PMLR.

GU, Y., FANG, C., BÜHLMANN, P. and FAN, J. (2024). Causality Pursuit from Heterogeneous Environments via Neural Adversarial Invariance Learning. arXiv preprint. Available at arXiv:2405.04715.

HAO, Y., ORLITSKY, A. and PICHAPATI, V. (2018). On learning Markov chains. *Adv*. *Neural Inf*. *Process*. *Syst*. **31**.

IZENMAN, A. J. (1975). Reduced-rank regression for the multivariate linear model. *J*. *Multivariate Anal*. **5** 248–264. MR0373179 https://doi.org/10.1016/0047-259X(75)90042-1

JIANG, B., SUN, Q. and FAN, J. (2018). Bernstein's inequality for general Markov chains. arXiv preprint. Available at arXiv:1805.10721.

JIN, C., YANG, Z., WANG, Z. and JORDAN, M. I. (2020). Provably efficient reinforcement learning with linear function approximation. In *Conference on Learning Theory* 2137–2143. PMLR.

KUMAR, A. and DAUME III, H. (2012). Learning task grouping and overlap in multi-task learning. arXiv preprint. Available at arXiv:1206.6417.

LEVIN, D. A. and PERES, Y. (2017). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR3726904 https://doi.org/10.1090/mbk/107

McCullagh, P. (2019). *Generalized Linear Models*. Routledge, London.

Modi, A., Chen, J., Krishnamurthy, A., Jiang, N. and Agarwal, A. (2024). Model-free representation learning and exploration in low-rank MDPs. *J. Mach. Learn. Res.* **25** Paper No. [6]. MR4723856

Norris, J. R. (1998). *Markov Chains*. *Cambridge Series in Statistical and Probabilistic Mathematics* **2**. Cambridge Univ. Press, Cambridge. MR1600720

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York. MR0273810

Rohde, A. and Tsybakov, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

Wang, B. and Fan, J. (2025). Robust matrix completion with heavy-tailed noise. *J. Amer. Statist. Assoc.* **120** 922–934. MR4917267 https://doi.org/10.1080/01621459.2024.2375037

Wolfer, G. and Kontorovich, A. (2019). Minimax learning of ergodic Markov chains. In *Algorithmic Learning Theory* 2019. *Proc. Mach. Learn. Res.* (*PMLR*) **98** [903–929], 27. Proceedings of Machine Learning Research PMLR. MR3932874

Wolfer, G. and Kontorovich, A. (2021). Statistical estimation of ergodic Markov chain kernel over discrete state space. *Bernoulli* **27** 532–553. MR4177379 https://doi.org/10.3150/20-BEJ1248

Wright, J., Ganesh, A., Rao, S., Peng, Y. and Ma, Y. (2009). Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Adv. Neural Inf. Process. Syst.* **22**.

Yi, X., Park, D., Chen, Y. and Caramanis, C. (2016). Fast algorithms for robust PCA via gradient descent. *Adv. Neural Inf. Process. Syst.* **29**.

Yuan, M., Ekici, A., Lu, Z. and Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 329–346. MR2323756 https://doi.org/10.1111/j.1467-9868.2007.00591.x

Zhang, A. and Wang, M. (2020). Spectral state compression of Markov processes. *IEEE Trans. Inf. Theory* **66** 3202–3231. MR4089777 https://doi.org/10.1109/TIT.2019.2956737

The Annals of Statistics
2025, Vol. 53, No. 5, 1963–1986
https://doi.org/10.1214/25-AOS2526

ONLINE STATISTICAL INFERENCE IN DECISION-MAKING WITH MATRIX CONTEXT

BY QIYU HAN[a], WILL WEI SUN[b] AND YICHEN ZHANG[c]

Daniels School of Business, Purdue University, [a]*han541@purdue.edu,* [b]*sun244@purdue.edu,* [c]*yichen@purdue.edu*

The study of online decision-making problems that leverage contextual information has drawn notable attention due to their significant applications in fields ranging from healthcare to autonomous systems. In modern applications, contextual information can be rich and is often represented as a matrix. Moreover, while existing online decision algorithms mainly focus on reward maximization, less attention has been devoted to statistical inference. To address these gaps, in this work, we consider an online decision-making problem with a matrix context where the true model parameters have a low-rank structure. We propose a *fully online* procedure to conduct statistical inference with adaptively collected data. The low-rank structure of the model parameter and the adaptive nature of the data collection process make this difficult: standard low-rank estimators are biased and cannot be obtained in a sequential manner while existing inference approaches in sequential decision-making algorithms fail to account for the low-rankness and are also biased. To overcome these challenges, we introduce a new online debiasing procedure to simultaneously handle both sources of bias. Our inference framework encompasses both parameter inference and optimal policy value inference. In theory, we establish the asymptotic normality of the proposed online debiased estimators and prove the validity of the constructed confidence intervals for both inference tasks. Our inference results are built upon a newly developed low-rank stochastic gradient descent estimator and its convergence result, which are also of independent interest.

REFERENCES

AGARWAL, A., KAKADE, S. M., LEE, J. D. and MAHAJAN, G. (2021). On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *J. Mach. Learn. Res.* **22** 98. MR4279749

AGRAWAL, S. and GOYAL, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*.

AKROUT, M., FARAHMAND, A.-M., JARMAIN, T. and ABID, L. (2019). Improving skin condition classification with a visual symptom checker trained using reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77. MR1954855 https://doi.org/10.1137/S0097539701398375

BIAN, Z., SHI, C., QI, Z. and WANG, L. (2025). Off-policy evaluation in doubly inhomogeneous environments. *J. Amer. Statist. Assoc.* **120** 1102–1114. MR4917281 https://doi.org/10.1080/01621459.2024.2395593

BIBAUT, A., DIMAKOPOULOU, M., KALLUS, N., CHAMBAZ, A. and VAN DER LAAN, M. (2021). Post-contextual-bandit inference. *Adv. Neural Inf. Process. Syst.* **34** 28548–28559.

BOUTILIER, C., HSU, C.-W., KVETON, B., MLADENOV, M., SZEPESVARI, C. and ZAHEER, M. (2020). Differentiable meta-learning of bandit policies. *Adv. Neural Inf. Process. Syst.* **33** 2122–2134.

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

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 Vol-*

MSC2020 subject classifications. Primary 62L20, 90B50; secondary 62E20.

Key words and phrases. Online inference, online decision-making, low-rank matrix, reinforcement learning, stochastic gradient descent.

ume. *Progress in Probability* **74** 385–430. Birkhäuser/Springer, Cham. MR4181374 https://doi.org/10.1007/978-3-030-26391-1_18

CARPENTIER, A. and KIM, A. K. H. (2018). An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution. *Statist*. *Sinica* **28** 1371–1393. MR3821009

CHEN, E. Y., XIA, D., CAI, C. and FAN, J. (2024a). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **86** 793–823. MR4792086 https://doi.org/10.1093/jrsssb/qkae001

CHEN, H., LU, W. and SONG, R. (2021a). Statistical inference for online decision making via stochastic gradient descent. *J*. *Amer*. *Statist*. *Assoc*. **116** 708–719. MR4270018 https://doi.org/10.1080/01621459.2020.1826325

CHEN, H., LU, W. and SONG, R. (2021b). Statistical inference for online decision making: In a contextual bandit setting. *J*. *Amer*. *Statist*. *Assoc*. **116** 240–255. MR4227691 https://doi.org/10.1080/01621459.2020.1770098

CHEN, X., LAI, Z., LI, H. and ZHANG, Y. (2022). Online statistical inference for contextual bandits via stochastic gradient descent. arXiv preprint. Available at arXiv:2212.14883.

CHEN, X., LAI, Z., LI, H. and ZHANG, Y. (2024b). Online statistical inference for stochastic optimization via Kiefer-Wolfowitz methods. *J*. *Amer*. *Statist*. *Assoc*. **119** 2972–2982. MR4833930 https://doi.org/10.1080/01621459.2023.2296703

CHEN, X., LEE, J. D., TONG, X. T. and ZHANG, Y. (2020). Statistical inference for model parameters in stochastic gradient descent. *Ann*. *Statist*. **48** 251–273. MR4065161 https://doi.org/10.1214/18-AOS1801

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

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

DELIU, N., WILLIAMS, J. J. and CHAKRABORTY, B. (2024). Reinforcement learning in modern biostatistics: Constructing optimal adaptive interventions. *Int*. *Stat*. *Rev*..

DESHPANDE, Y., JAVANMARD, A. and MEHRABI, M. (2023). Online debiasing for adaptively collected high-dimensional data with applications to time series analysis. *J*. *Amer*. *Statist*. *Assoc*. **118** 1126–1139. MR4595482 https://doi.org/10.1080/01621459.2021.1979011

FANG, E. X., WANG, Z. and WANG, L. (2023). Fairness-oriented learning for optimal individualized treatment rules. *J*. *Amer*. *Statist*. *Assoc*. **118** 1733–1746. MR4646602 https://doi.org/10.1080/01621459.2021.2008402

FANG, Y., XU, J. and YANG, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *J*. *Mach*. *Learn*. *Res*. **19** 78. MR3899780

HADAD, V., HIRSHBERG, D. A., ZHAN, R., WAGER, S. and ATHEY, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **118** e2014602118. MR4294058 https://doi.org/10.1073/pnas.2014602118

HAN, Q., SUN, W. W and ZHANG, Y. (2025). Supplement to "Online Statistical Inference in Decision-Making with Matrix Context." https://doi.org/10.1214/25-AOS2526SUPP

ISTEPANIAN, R., LAXMINARAYAN, S. and PATTICHIS, C. S. (2007). *M-Health*: *Emerging Mobile Health Systems*. Springer, Berlin.

JIN, C., KAKADE, S. M. and NETRAPALLI, P. (2016). Provable efficient online matrix completion via non-convex stochastic gradient descent. *Adv*. *Neural Inf*. *Process*. *Syst*.

KHAMARU, K., DESHPANDE, Y., MACKEY, L. and WAINWRIGHT, M. J. (2021). Near-optimal inference in adaptive linear regression. arXiv preprint. Available at arXiv:2107.02266.

KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J*. *Mach*. *Learn*. *Res*. **16** 1757–1792. MR3417798

KOSOROK, M. R. and LABER, E. B. (2019). Precision medicine. *Annu*. *Rev*. *Stat*. *Appl*. **6** 263–286. MR3939521 https://doi.org/10.1146/annurev-statistics-030718-105251

LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

LI, L., CHU, W., LANGFORD, J. and SCHAPIRE, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *International Conference on World Wide Web* 661–670.

LI, L., LU, Y. and ZHOU, D. (2017). Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*.

LI, Y., XIE, H., LIN, Y. and LUI, J. C. (2021). Unifying offline causal inference and online bandit learning for data driven decision. In *Proceedings of the Web Conference* 2021 2291–2303.

LIU, W., TU, J., ZHANG, Y. and CHEN, X. (2023). Online estimation and inference for robust policy evaluation in reinforcement learning. arXiv preprint. Available at arXiv:2310.02581.

MEI, J., XIAO, C., SZEPESVARI, C. and SCHUURMANS, D. (2020). On the global convergence rates of softmax policy gradient methods. In *International Conference on Machine Learning* 6820–6829. PMLR.

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann*. *Statist*. **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

POLDRACK, R. A., MUMFORD, J. A. and NICHOLS, T. E. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge Univ. Press, Cambridge. MR2839490 https://doi.org/10.1017/CBO9780511895029

POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim*. **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

QI, Z., PANG, J.-S. and LIU, Y. (2023). On robustness of individualized decision rules. *J. Amer. Statist. Assoc*. **118** 2143–2157. MR4646632 https://doi.org/10.1080/01621459.2022.2038180

RAMPRASAD, P., LI, Y., YANG, Z., WANG, Z., SUN, W. W. and CHENG, G. (2023). Online bootstrap inference for policy evaluation in reinforcement learning. *J. Amer. Statist. Assoc*. **118** 2901–2914. MR4681629 https://doi.org/10.1080/01621459.2022.2096620

SHEN, Y., CAI, H. and SONG, R. (2024). Doubly robust interval estimation for optimal policy evaluation in online learning. *J. Amer. Statist. Assoc*. **119** 2811–2821. MR4833917 https://doi.org/10.1080/01621459.2023.2279289

SHI, C., LUO, S., ZHU, H. and SONG, R. (2021a). An online sequential test for qualitative treatment effects. *J. Mach. Learn. Res*. **22** 286. MR4353065

SHI, C., SONG, R., LU, W. and LI, R. (2021b). Statistical inference for high-dimensional models via recursive online-score estimation. *J. Amer. Statist. Assoc*. **116** 1307–1318. MR4309274 https://doi.org/10.1080/01621459.2019.1710154

SHI, C., WANG, X., LUO, S., ZHU, H., YE, J. and SONG, R. (2023). Dynamic causal effects evaluation in A/B testing with a reinforcement learning framework. *J. Amer. Statist. Assoc*. **118** 2059–2071. MR4646626 https://doi.org/10.1080/01621459.2022.2027776

SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **84** 765–793. MR4460575

SHI, C., ZHU, J., YE, S., LUO, S., ZHU, H. and SONG, R. (2024). Off-policy confidence interval estimation with confounded Markov decision process. *J. Amer. Statist. Assoc*. **119** 273–284. MR4713891 https://doi.org/10.1080/01621459.2022.2110878

SIMCHI-LEVI, D. and WANG, C. (2023). Multi-armed bandit experimental design: Online decision-making and adaptive inference. In *International Conference on Artificial Intelligence and Statistics*.

TANG, K., LIU, W., ZHANG, Y. and CHEN, X. (2023). Acceleration of stochastic gradient descent with momentum by averaging: Finite-sample rates and asymptotic normality. arXiv preprint. Available at arXiv:2305.17665.

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

XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **83** 58–77. MR4220984 https://doi.org/10.1111/rssb.12400

ZHAN, R., HADAD, V., HIRSHBERG, D. A. and ATHEY, S. (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Proceedings of the* 27*th ACM SIGKDD Conference*.

ZHANG, K., JANSON, L. and MURPHY, S. (2021). Statistical inference with M-estimators on adaptively collected data. *Adv. Neural Inf. Process. Syst*.

ZHANG, K. W., JANSON, L. and MURPHY, S. A. (2022). Statistical inference after adaptive sampling in non-Markovian environments. arXiv preprint. Available at arXiv:2202.07098.

ZHU, W., CHEN, X. and WU, W. B. (2023). Online covariance matrix estimation in stochastic gradient descent. *J. Amer. Statist. Assoc*. **118** 393–404. MR4571129 https://doi.org/10.1080/01621459.2021.1933498

ZHU, Z., LI, X., WANG, M. and ZHANG, A. (2022). Learning Markov models via low-rank optimization. *Oper. Res*. **70** 2384–2398. MR4484413

The Annals of Statistics
2025, Vol. 53, No. 5, 1987–2011
https://doi.org/10.1214/25-AOS2527

ESTIMATION AND INFERENCE IN DISTRIBUTIONAL REINFORCEMENT LEARNING

BY LIANGYU ZHANG[1,a], YANG PENG[2,b], JIADONG LIANG[2,c], WENHAO YANG[2,d] AND ZHIHUA ZHANG[3,e]

[1]*School of Statistics and Data Science, Shanghai University of Finance and Economics,* [a]*zhangliangyu@sufe.edu.cn*
[2]*School of Mathematical Sciences, Peking University,* [b]*pengyang@pku.edu.cn,* [c]*jdliang@pku.edu.cn,* [d]*yangwenhaosms@pku.edu.cn*
[3]*School of Mathematical Sciences, School of Computer Science, Peking University,* [e]*zhzhang@math.pku.edu.cn*

In this paper, we study distributional reinforcement learning from the perspective of statistical efficiency. We investigate distributional policy evaluation, aiming to estimate the complete return distribution (denoted η^π) attained by a given policy π. We use the certainty equivalence method to construct our estimator $\hat{\eta}_n^\pi$, based on a generative model. In this circumstance, we need a dataset of size $\widetilde{O}(|\mathcal{S}||\mathcal{A}|\varepsilon^{-2p}(1-\gamma)^{-(2p+2)})$ to guarantee the supremum p-Wasserstein metric between $\hat{\eta}_n^\pi$ and η^π less than ε with high probability. This implies the distributional policy evaluation problem can be solved with sample efficiency. Also, we show that under different mild assumptions a dataset of size $\widetilde{O}(|\mathcal{S}||\mathcal{A}|\varepsilon^{-2}(1-\gamma)^{-4})$ suffices to ensure the supremum Kolmogorov-Smirnov metric and supremum total variation metric between $\hat{\eta}_n^\pi$ and η^π is below ε with high probability. Furthermore, we investigate the asymptotic behavior of $\hat{\eta}_n^\pi$. We demonstrate that the "empirical process" $\sqrt{n}(\hat{\eta}_n^\pi - \eta^\pi)$ converges weakly to a Gaussian process in the space of bounded functionals on a Lipschitz function class $\ell^\infty(\mathcal{F}_{W_1})$, also in the space of bounded functionals on an indicator function class $\ell^\infty(\mathcal{F}_{\mathrm{KS}})$ and a bounded measurable function class $\ell^\infty(\mathcal{F}_{\mathrm{TV}})$ when some mild conditions hold. Our findings give rise to a unified approach to statistical inference of a wide class of statistical functionals of η^π.

REFERENCES

[1] ALQUIER, P., FRIEL, N., EVERITT, R. and BOLAND, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Stat. Comput.* **26** 29–47. MR3439357 https://doi.org/10.1007/s11222-014-9521-x
[2] BARDENET, R., DOUCET, A. and HOLMES, C. (2014). Towards scaling up Markov chain Monte Carlo: An adaptive subsampling approach. In *International Conference on Machine Learning* 405–413. PMLR.
[3] BELLEMARE, M. G., DABNEY, W. and MUNOS, R. (2017). A distributional perspective on reinforcement learning. In *International Conference on Machine Learning* 449–458. PMLR.
[4] BELLEMARE, M. G., DABNEY, W. and ROWLAND, M. (2023). *Distributional Reinforcement Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. http://www.distributional-rl.org. MR4649766
[5] CHAE, M. and WALKER, S. G. (2020). Wasserstein upper bounds of the total variation for smooth densities. *Statist. Probab. Lett.* **163** 108771, 6. MR4083852 https://doi.org/10.1016/j.spl.2020.108771
[6] CHANDAK, Y., NIEKUM, S., DA SILVA, B., LEARNED-MILLER, E., BRUNSKILL, E. and THOMAS, P. S. (2021). Universal off-policy evaluation. *Adv. Neural Inf. Process. Syst.* **34** 27475–27490.
[7] CHEN, J. and JIANG, N. (2019). Information-theoretic considerations in batch reinforcement learning. In *International Conference on Machine Learning* 1042–1051. PMLR.
[8] CLEMENTS, W. R., VAN DELFT, B., ROBAGLIA, B.-M., SLAOUI, R. B. and TOTH, S. (2019). Estimating risk and uncertainty in deep reinforcement learning. arXiv preprint. Available at arXiv:1905.09638.

MSC2020 subject classifications. Primary 62C05, 62G05, 62G15; secondary 90C40, 68Q32.
Key words and phrases. Distributional reinforcement learning, probability metrics, sample complexity bounds, weak convergence.

[9] DABNEY, W., OSTROVSKI, G., SILVER, D. and MUNOS, R. (2018). Implicit quantile networks for distributional reinforcement learning. In *International Conference on Machine Learning* 1096–1105. PMLR.

[10] DABNEY, W., ROWLAND, M., BELLEMARE, M. and MUNOS, R. (2018). Distributional reinforcement learning with quantile regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[11] DOAN, T., MAZOURE, B. and LYLE, C. (2018). Gan q-learning. arXiv preprint. Available at arXiv:1805.04874.

[12] FAWZI, A., BALOG, M., HUANG, A., HUBERT, T., ROMERA-PAREDES, B., BAREKATAIN, M., NOVIKOV, A., RUIZ, F. J., SCHRITTWIESER, J. et al. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610** 47–53.

[13] 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

[14] FREIRICH, D., SHIMKIN, T., MEIR, R. and TAMAR, A. (2019). Distributional multivariate policy evaluation and exploration with the Bellman gan. In *International Conference on Machine Learning* 1983–1992. PMLR.

[15] GHYSELS, E., SANTA-CLARA, P. and VALKANOV, R. (2005). There is a risk-return trade-off after all. *J. Financ. Econ.* **76** 509–548.

[16] HAO, B., JI, X., DUAN, Y., LU, H., SZEPESVARI, C. and WANG, M. (2021). Bootstrapping fitted q-evaluation for off-policy inference. In *International Conference on Machine Learning* 4074–4084. PMLR.

[17] HUA, Y., LI, R., ZHAO, Z., CHEN, X. and ZHANG, H. (2019). GAN-powered deep distributional reinforcement learning for resource management in network slicing. *IEEE J. Sel. Areas Commun.* **38** 334–349.

[18] HUANG, A., LEQI, L., LIPTON, Z. and AZIZZADENESHELI, K. (2022). Off-policy risk assessment for Markov decision processes. In *International Conference on Artificial Intelligence and Statistics* 5022–5050. PMLR.

[19] JIANG, N. and LI, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning* 652–661. PMLR.

[20] JOHNDROW, J. E. and MATTINGLY, J. C. (2017). Error bounds for approximations of Markov chains used in Bayesian sampling. arXiv preprint. Available at arXiv:1711.05382.

[21] KARTASHOV, N. V. (1986). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Theory Probab. Appl.* **30** 247–259.

[22] KOBER, J., BAGNELL, J. A. and PETERS, J. (2013). Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **32** 1238–1274.

[23] LAVORI, P. W. and DAWSON, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clin. Trials* **1** 9–20.

[24] LI, G., WEI, Y., CHI, Y. and CHEN, Y. (2024). Breaking the sample size barrier in model-based reinforcement learning with a generative model. *Oper. Res.* **72** 203–221. MR4705834

[25] LI, X., LIANG, J. and ZHANG, Z. (2023). Online statistical inference for nonlinear stochastic approximation with Markovian data. arXiv preprint. Available at arXiv:2302.07690.

[26] LI, X., YANG, W., LIANG, J., ZHANG, Z. and JORDAN, M. I. (2023). A statistical analysis of Polyak-ruppert averaged q-learning. In *International Conference on Artificial Intelligence and Statistics* 2207–2261. PMLR.

[27] LIANG, H. and LUO, Z. (2024). Regret bounds for risk-sensitive reinforcement learning with Lipschitz dynamic risk measures. In *Proceedings of the* 27*th International Conference on Artificial Intelligence and Statistics* (S. Dasgupta, S. Mandt and Y. Li, eds.). *Proceedings of Machine Learning Research* **238** 1774–1782. PMLR.

[28] LIM, S. H. and MALIK, I. (2022). Distributional reinforcement learning for risk-sensitive policies. In *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.) **35** 30977–30989. Curran Associates, Red Hook.

[29] LOCKWOOD, O. and SI, M. (2022). A review of uncertainty for deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* **18** 155–162.

[30] MA, X., XIA, L., ZHOU, Z., YANG, J. and ZHAO, Q. (2020). Dsac: Distributional soft actor critic for risk-sensitive reinforcement learning. arXiv preprint. Available at arXiv:2004.14547.

[31] MITROPHANOV, A. Y. (2005). Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Probab.* **42** 1003–1014. MR2203818 https://doi.org/10.1239/jap/1134587812

[32] MORIMURA, T., SUGIYAMA, M., KASHIMA, H., HACHIYA, H. and TANAKA, T. (2010). Nonparametric return distribution approximation for reinforcement learning. In *Proceedings of the* 27*th International Conference on Machine Learning* (*ICML*-10) 799–806.

[33] NAEEM, F., SEIFOLLAHI, S., ZHOU, Z. and TARIQ, M. (2020). A generative adversarial network enabled deep distributional reinforcement learning for transmission scheduling in Internet of vehicles. *IEEE Trans. Intell. Transp. Syst.* **22** 4550–4559.

[34] OPENAI (2023). GPT-4 Technical Report.
[35] OUYANG, L., WU, J., JIANG, X., ALMEIDA, D., WAINWRIGHT, C., MISHKIN, P., ZHANG, C., AGARWAL, S., SLAMA, K. et al. (2022). Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **35** 27730–27744.
[36] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182
[37] ROWLAND, M., BELLEMARE, M., DABNEY, W., MUNOS, R. and TEH, Y. W. (2018). An analysis of categorical distributional reinforcement learning. In *International Conference on Artificial Intelligence and Statistics* 29–37. PMLR.
[38] ROWLAND, M., MUNOS, R., AZAR, M. G., TANG, Y., OSTROVSKI, G., HARUTYUNYAN, A., TUYLS, K., BELLEMARE, M. G. and DABNEY, W. (2024). An analysis of quantile temporal-difference learning. *J. Mach. Learn. Res.* **25** Paper No. [163], 47. MR4758134
[39] ROWLAND, M., TANG, Y., LYLE, C., MUNOS, R., BELLEMARE, M. G. and DABNEY, W. (2023). The statistical benefits of quantile temporal-difference learning for value estimation. In *International Conference on Machine Learning* 29210–29231. PMLR.
[40] 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
[41] RUDOLF, D., SMITH, A. and QUIROZ, M. (2024). Perturbations of Markov chains. arXiv preprint. Available at arXiv:2404.10251.
[42] SCHWEITZER, P. J. (1968). Perturbation theory and finite Markov chains. *J. Appl. Probab.* **5** 401–413. MR0234527 https://doi.org/10.2307/3212261
[43] SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 765–793. MR4460575
[44] SILVER, D., HUBERT, T., SCHRITTWIESER, J., ANTONOGLOU, I., LAI, M., GUEZ, A., LANCTOT, M., SIFRE, L., KUMARAN, D. et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362** 1140–1144. MR3888768 https://doi.org/10.1126/science.aar6404
[45] SIMON, H. A. (1956). Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica* **24** 74–81. MR0077847 https://doi.org/10.2307/1905261
[46] SINGH, R., ZHANG, Q. and CHEN, Y. (2020). Improving robustness via risk averse distributional reinforcement learning. In *Proceedings of the* 2*nd Conference on Learning for Dynamics and Control* (A. M. Bayen, A. Jadbabaie, G. Pappas, P. A. Parrilo, B. Recht, C. Tomlin and M. Zeilinger, eds.). *Proceedings of Machine Learning Research* **120** 958–968. PMLR.
[47] SUN, K., ZHAO, Y., LIU, Y., SHI, E., WANG, Y., YAN, X., JIANG, B. and KONG, L. (2022). *Interpreting Distributional Reinforcement Learning: A Regularization Perspective*.
[48] SUTTON, R. S. (2004). The reward hypothesis.
[49] SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951
[50] THEIL, H. (1957). A note on certainty equivalence in dynamic planning. *Econometrica* **25** 346–349. MR0091227 https://doi.org/10.2307/1910260
[51] THOMAS, P., THEOCHAROUS, G. and GHAVAMZADEH, M. (2015). High-confidence off-policy evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[52] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256
[53] VINYALS, O., BABUSCHKIN, I., CZARNECKI, W. M., MATHIEU, M., DUDZIK, A., CHUNG, J., CHOI, D. H., POWELL, R., EWALDS, T. et al. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* **575** 350–354.
[54] WANG, K., ZHOU, K., WU, R., KALLUS, N. and SUN, W. (2023). The benefits of being distributional: Small-loss bounds for reinforcement learning. *Adv. Neural Inf. Process. Syst.* **36** 2275–2312.
[55] WILTZER, H., FAREBROTHER, J., GRETTON, A., TANG, Y., BARRETO, A., DABNEY, W., BELLEMARE, M. G. and ROWLAND, M. (2024). A distributional analogue to the successor representation. In *Proceedings of the* 41*st International Conference on Machine Learning* 52994–53016.
[56] WU, R., UEHARA, M. and SUN, W. (2023). Distributional offline policy evaluation with predictive error guarantees. In *International Conference on Machine Learning* 37685–37712. PMLR.
[57] YANG, W., ZHANG, L. and ZHANG, Z. (2022). Toward theoretical understandings of robust Markov decision processes: Sample complexity and asymptotics. *Ann. Statist.* **50** 3223–3248. MR4524495 https://doi.org/10.1214/22-aos2225

[58] ZHANG, L., PENG, Y., LIANG, J., YANG, W. and ZHANG, Z. (2025). Supplement to "Estimation and Inference in Distributional Reinforcement Learning." https://doi.org/10.1214/25-AOS2527SUPPA, https://doi.org/10.1214/25-AOS2527SUPPB
[59] ZHU, Y., DONG, J. and LAM, H. (2024). Uncertainty quantification and exploration for reinforcement learning. *Oper. Res*. **72** 1689–1709. MR4812023

The Annals of Statistics
2025, Vol. 53, No. 5, 2012–2039
https://doi.org/10.1214/25-AOS2528

ADVANCES IN BAYESIAN MODEL SELECTION CONSISTENCY FOR HIGH-DIMENSIONAL GENERALIZED LINEAR MODELS

BY JEYONG LEE[1,a], MINWOO CHAE[1,b] AND RYAN MARTIN[2,c]

[1]*Department of Industrial and Management Engineering, Pohang University of Science and Technology,* [a]*jylee1024@postech.ac.kr,* [b]*mchae@postech.ac.kr*

[2]*Department of Statistics, North Carolina State University,* [c]*rgmarti3@ncsu.edu*

Uncovering genuine relationships between a response variable and a large collection of covariates is a fundamental and practically important problem. In the context of Gaussian linear models, both the Bayesian and non-Bayesian literature is well-developed and there are no substantial differences in the model selection consistency results available from the two schools. For generalized linear models (GLMs), however, Bayesian model selection consistency results are lacking in several ways. In this paper, we construct a Bayesian posterior distribution using an appropriate data-dependent prior and develop its asymptotic concentration properties using new theoretical techniques. In particular, we leverage Spokoiny's powerful nonasymptotic theory to obtain sharp quadratic approximations of the GLM's log-likelihood function, which leads to tight bounds on the errors associated with the model-specific maximum likelihood estimators and the Laplace approximation of our Bayesian marginal likelihood. In turn, these improved bounds lead to significantly stronger, near-optimal Bayesian model selection consistency results, for example, far weaker beta-min conditions, compared to those available in the existing literature. In particular, our results are applicable to the Poisson regression model, in which the score function is not sub-Gaussian.

REFERENCES

ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.* **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

BARBER, R. F. and DRTON, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electron. J. Stat.* **9** 567–607. MR3326135 https://doi.org/10.1214/15-EJS1012

BARBER, R. F., DRTON, M. and TAN, K. M. (2016). Laplace approximation in high-dimensional Bayesian regression. In *Statistical Analysis for High-Dimensional Data. Abel Symp.* **11** 15–36. Springer, Cham. MR3616262

BELITSER, E. and GHOSAL, S. (2020). Empirical Bayes oracle uncertainty quantification for regression. *Ann. Statist.* **48** 3113–3137. MR4185802 https://doi.org/10.1214/19-AOS1845

BHATTACHARYA, A., PATI, D. and YANG, Y. (2019). Bayesian fractional posteriors. *Ann. Statist.* **47** 39–66. MR3909926 https://doi.org/10.1214/18-AOS1712

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388

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

CAO, X. and LEE, K. (2024). Bayesian inference on hierarchical nonlocal priors in generalized linear models. *Bayesian Anal.* **19** 99–122. MR4692544 https://doi.org/10.1214/22-ba1350

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

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

MSC2020 subject classifications. Primary 62F15, 62E20; secondary 62E17.

Key words and phrases. Bayesian model selection consistency, beta-min condition, Laplace approximation, logistic regression, Poisson regression.

CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

CHAE, M., LIN, L. and DUNSON, D. B. (2019). Bayesian sparse linear regression with unknown symmetric error. *Inf. Inference* **8** 621–653. MR3994400 https://doi.org/10.1093/imaiai/iay022

CHEN, J. and CHEN, Z. (2012). Extended BIC for small-n-large-P sparse GLM. *Statist. Sinica* **22** 555–574. MR2954352 https://doi.org/10.5705/ss.2010.216

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

GEORGE, E. I. (2000). The variable selection problem. *J. Amer. Statist. Assoc.* **95** 1304–1308. MR1825282 https://doi.org/10.2307/2669776

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274 https://doi.org/10.1214/009053606000001172

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

GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. *Springer Series in Statistics*. Springer, New York. MR1992245

GÖTZE, F., SAMBALE, H. and SINULIS, A. (2021). Concentration inequalities for polynomials in α-sub-exponential random variables. *Electron. J. Probab.* **26** Paper No. 48. MR4247973 https://doi.org/10.1214/21-ejp606

GRÜNWALD, P. and VAN OMMEN, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* **12** 1069–1103. MR3724979 https://doi.org/10.1214/17-BA1085

HANS, C., DOBRA, A. and WEST, M. (2007). Shotgun stochastic search for "large p" regression. *J. Amer. Statist. Assoc.* **102** 507–516. MR2370849 https://doi.org/10.1198/016214507000000121

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

HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141

HOU, T., WANG, L. and ATCHADÉ, Y. (2024). Laplace approximation for Bayesian variable selection via Le Cam's one-step procedure. Available at arXiv:2407.20580.

HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** no. 52. MR2994877 https://doi.org/10.1214/ECP.v17-2079

ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158 https://doi.org/10.1214/009053604000001147

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

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

LEE, J., CHAE, M. and MARTIN, R. (2025). Supplement to "Advances in Bayesian model selection consistency for high-dimensional generalized linear models." https://doi.org/10.1214/25-AOS2528SUPP

LEE, K. and CAO, X. (2021). Bayesian group selection in logistic regression with application to MRI data analysis. *Biometrics* **77** 391–400. MR4307642 https://doi.org/10.1111/biom.13290

LIU, C. and MARTIN, R. (2019). An empirical G-Wishart prior for sparse high-dimensional Gaussian graphical models. Available at arXiv:1912.03807.

LOH, P.-L. and WAINWRIGHT, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *Ann. Statist.* **45** 2455–2482. MR3737898 https://doi.org/10.1214/16-AOS1530

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

MARTIN, R. and NING, B. (2020). Empirical priors and coverage of posterior credible sets in a sparse normal mean model. *Sankhya A* **82** 477–498. MR4136243 https://doi.org/10.1007/s13171-019-00189-w

MARTIN, R. and SYRING, N. (2022). Direct Gibbs posterior inference on risk minimizers: Construction, concentration, and calibration. In *Advancements in Bayesian Methods and Implementation* (A. S. R. Srinivasa Rao, G. A. Young and C. R. Rao, eds.). *Handbook of Statist.* **47** 1–41. Elsevier/Academic Press, Boston, MA. MR4599140

MARTIN, R. and TANG, Y. (2020b). Empirical priors for prediction in sparse high-dimensional linear regression. *J. Mach. Learn. Res.* **21** Paper No. 144. MR4138128

MARTIN, R. and WALKER, S. G. (2014). Asymptotically minimax empirical Bayes estimation of a sparse normal mean vector. *Electron. J. Stat.* **8** 2188–2206. MR3273623 https://doi.org/10.1214/14-EJS949

MARTIN, R. and WALKER, S. G. (2019). Data-driven priors and their posterior concentration rates. *Electron. J. Stat.* **13** 3049–3081. MR4010592 https://doi.org/10.1214/19-ejs1600

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). *SparseNet*: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769 https://doi.org/10.1198/jasa.2011.tm09738

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

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

NARISETTY, N. N., SHEN, J. and HE, X. (2019). Skinny Gibbs: A consistent and scalable Gibbs sampler for model selection. *J. Amer. Statist. Assoc.* **114** 1205–1217. MR4011773 https://doi.org/10.1080/01621459.2018.1482754

NIE, L. and ROČKOVÁ, V. (2023). Bayesian bootstrap spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **118** 2013–2028. MR4646623 https://doi.org/10.1080/01621459.2022.2025815

PIIRONEN, J. and VEHTARI, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11** 5018–5051. MR3738204 https://doi.org/10.1214/17-EJS1337SI

RAY, K. and SZABÓ, B. (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc.* **117** 1270–1281. MR4480711 https://doi.org/10.1080/01621459.2020.1847121

RAY, K., SZABÓ, B. and CLARA, G. (2020). Spike and slab variational Bayes for high dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **33** 14423–14434.

ROČKOVÁ, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Statist.* **46** 401–437. MR3766957 https://doi.org/10.1214/17-AOS1554

ROČKOVÁ, V. and GEORGE, E. I. (2018). The spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. MR3803476 https://doi.org/10.1080/01621459.2016.1260469

ROSSELL, D., ABRIL, O. and BHATTACHARYA, A. (2021). Approximate Laplace approximations for scalable model selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 853–879. MR4320004

ROSSELL, D. and TELESCA, D. (2017). Nonlocal priors for high-dimensional estimation. *J. Amer. Statist. Assoc.* **112** 254–265. MR3646569 https://doi.org/10.1080/01621459.2015.1130634

SHIN, M., BHATTACHARYA, A. and JOHNSON, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statist. Sinica* **28** 1053–1078. MR3791100

SPOKOINY, V. (2012). Parametric estimation. Finite sample theory. *Ann. Statist.* **40** 2877–2909. MR3097963 https://doi.org/10.1214/12-AOS1054

SPOKOINY, V. (2017). Penalized maximum likelihood estimation and effective dimension. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 389–429. MR3606746 https://doi.org/10.1214/15-AIHP720

SPOKOINY, V. (2025). Deviation bounds for the norm of a random vector under exponential moment conditions with applications. *Probab. Uncertain. Quant. Risk* **10** 135–158. MR4893242 https://doi.org/10.3934/puqr.2025007

SYRING, N. and MARTIN, R. (2023). Gibbs posterior concentration rates under sub-exponential type losses. *Bernoulli* **29** 1080–1108. MR4550216 https://doi.org/10.3150/22-bej1491

TANG, Y. and MARTIN, R. (2024). Empirical Bayes inference in sparse high-dimensional generalized linear models. *Electron. J. Stat.* **18** 3212–3246. MR4779882 https://doi.org/10.1214/24-ejs2274

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

VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809 https://doi.org/10.1214/009053607000000929

VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Adaptive posterior contraction rates for the horseshoe. *Electron. J. Stat.* **11** 3196–3225. MR3705450 https://doi.org/10.1214/17-EJS1316

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

WAINWRIGHT, M. J. (2009b). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory* **55** 5728–5741. MR2597190 https://doi.org/10.1109/TIT.2009.2032816

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

WALKER, S. and HJORT, N. L. (2001). On Bayesian consistency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 811–821. MR1872068 https://doi.org/10.1111/1467-9868.00314

WAN, K. Y. Y. and GRIFFIN, J. E. (2021). An adaptive MCMC method for Bayesian variable selection in logistic and accelerated failure time regression models. *Stat. Comput.* **31** Paper No. 6. MR4199466 https://doi.org/10.1007/s11222-020-09974-2

YANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. MR3576552 https://doi.org/10.1214/15-AOS1417

ZHANG, T. (2006). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Trans. Inf. Theory* **52** 1307–1321. MR2241190 https://doi.org/10.1109/TIT.2005.864439

ZHANG, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.* **11** 1081–1107. MR2629825

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

The Annals of Statistics
2025, Vol. 53, No. 5, 2040–2076
https://doi.org/10.1214/25-AOS2529

SYMMETRY: A GENERAL STRUCTURE IN NONPARAMETRIC REGRESSION

BY LOUIS GOLDWATER CHRISTIE[1,a] AND JOHN A. D. ASTON[2,b]

[1]*Department of Data Science and Operations, University of Southern California,* [a]*lgchrist@usc.edu*
[2]*Statistical Laboratory, University of Cambridge,* [b]*J.Aston@statslab.cam.ac.uk*

In this paper, we present the framework of symmetry in nonparametric regression. This generalises the framework of covariate sparsity, where the regression function $f : [0, 1]^d \to \mathbb{R}$ depends only on at most $s < d$ of the covariates, which is a special case of translation symmetry with linear orbits. In general, this extends to other types of functions that capture lower dimensional behavior even when these structures are nonlinear. We show both that known symmetries of regression functions can be exploited to give similarly faster rates, and that unknown symmetries with Lipschitz actions can be estimated sufficiently quickly to adaptively obtain the same rates. This is done by explicit constructions of partial symmetrisation operators that are then applied to usual estimators, and with a two-step M-estimator of the maximal symmetry of the regression function. We also demonstrate the finite sample performance of these estimators on synthetic data.

REFERENCES

BENTON, G., FINZI, M., IZMAILOV, P. and WILSON, A. G. (2020). Learning invariances in neural networks from training data. *Adv. Neural Inf. Process. Syst.* **33** 17605–17616.

BRONSTEIN, M. M., BRUNA, J., COHEN, T. and VELIČKOVIĆ, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv preprint. Available at arXiv:2104.13478.

CHIU, K. and BLOEM-REDDY, B. (2023). Hypothesis tests for distributional group symmetry with applications to particle physics. In *NeurIPS* 2023 *AI for Science Workshop*.

CHOU, C.-N., LOVE, P. J., SANDHU, J. S. and SHI, J. (2022). Limitations of local quantum algorithms on random MAX-k-XOR and beyond. In 49*th EATCS International Conference on Automata, Languages, and Programming. LIPIcs. Leibniz Int. Proc. Inform.* **229** 41:1–41:20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4473346 https://doi.org/10.4230/lipics.icalp.2022.41

CHRISTIE, L. G. and ASTON, J. A. (2025a). Estimating maximal symmetries of regression functions via subgroup lattices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, In Press. https://doi.org/10.1093/jrsssb/qkaf031

CHRISTIE, L. G and ASTON, J. A (2025b). Supplement to "Symmetry: A general structure in nonparametric regression." https://doi.org/10.1214/25-AOS2529SUPP

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

CUBUK, E. D., ZOPH, B., MANE, D., VASUDEVAN, V. and LE, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 113–123.

DE VITO, E., MÜCKE, N. and ROSASCO, L. (2021). Reproducing kernel Hilbert spaces on manifolds: Sobolev and diffusion spaces. *Anal. Appl.* (*Singap.*) **19** 363–396. MR4233705 https://doi.org/10.1142/S0219530520400114

DIECK, T. T. (1987). *Transform. Groups. De Gruyter Studies in Mathematics* **8**. de Gruyter, Berlin. MR0889050 https://doi.org/10.1515/9783110858372.312

ELESEDY, B. and ZAIDI, S. (2021). Provably strict generalisation benefit for equivariant models. In *International Conference on Machine Learning* 2959–2969. PMLR.

FLETCHER, T. (2010). Terse notes on Riemannian geometry. http://www.sci.utah.edu/~fletcher/RiemannianGeometryNotes.pdf. Accessed: 2024-02-05.

MSC2020 subject classifications. Primary 62G08; secondary 62R20.
Key words and phrases. Invariant models, dimension reduction, nonparametric regression.

GARCÍA-PORTUGUÉS, E., PAINDAVEINE, D. and VERDEBOUT, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.* **115** 1873–1887. MR4189764 https://doi.org/10.1080/01621459.2019.1665527

HAAR, A. (1933). Der Massbegriff in der Theorie der kontinuierlichen Gruppen. *Ann. of Math.* (2) **34** 147–169. MR1503103 https://doi.org/10.2307/1968346

HAWKINS, T. (2012). *Emergence of the Theory of Lie groups: An Essay in the History of Mathematics* 1869–1926. Springer, Berlin.

HENRIKSON, J. (1999). Completeness and total boundedness of the Hausdorff metric. *MIT Undergrad. J. Math.* **1** 10.

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

HUANG, K. H., ORBANZ, P. and AUSTERN, M. (2022). Quantifying the effects of data augmentation. arXiv preprint. Available at arXiv:2202.09134.

HUANG, Z. and SEN, B. (2023). Multivariate symmetry: Distribution-free testing via optimal transport. arXiv preprint. Available at arXiv:2305.01839.

JIANG, W. and TANG, L. (2017). Atomic cryo-EM structures of viruses. *Curr. Opin. Struck. Biol.* **46** 122–129.

KONDOR, R. and TRIVEDI, S. (2018). On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning* 2747–2755. PMLR.

LAFFERTY, J. and WASSERMAN, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *Ann. Statist.* **36** 28–63. MR2387963 https://doi.org/10.1214/009053607000000811

LIM, S., KIM, I., KIM, T., KIM, C. and KIM, S. (2019). Fast autoaugment. *Adv. Neural Inf. Process. Syst.* **32**.

LYLE, C., VAN DER WILK, M., KWIATKOWSKA, M., GAL, Y. and BLOEM-REDDY, B. (2020). On the benefits of invariance in neural networks. arXiv preprint. Available at arXiv:2005.00178.

MARDIA, K. V. and JUPP, P. E. (2009). *Directional Statistics* **494**. Wiley, New York.

MARRON, J. S. and DRYDEN, I. L. (2021). *Object Oriented Data Analysis*. CRC Press/CRC, Boca Raton.

MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2021). Learning with invariances in random features and kernel models. In *Conference on Learning Theory* 3351–3418. PMLR.

MILNOR, J. (1976). Curvatures of left invariant metrics on Lie groups. *Adv. Math.* **21** 293–329. MR0425012 https://doi.org/10.1016/S0001-8708(76)80002-3

MITZENMACHER, M. and UPFAL, E. (2017). *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR3674428

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NASH, J. (1956). The imbedding problem for Riemannian manifolds. *Ann. of Math.* (2) **63** 20–63. MR0075639 https://doi.org/10.2307/1969989

PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065 https://doi.org/10.1080/10485250500504828

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230. MR0760684 https://doi.org/10.1214/aos/1176346788

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

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

TAHMASEBI, B. and JEGELKA, S. (2023a). The exact sample complexity gain from invariances for kernel regression. *Adv. Neural Inf. Process. Syst.* **36**.

TAHMASEBI, B. and JEGELKA, S. (2023b). Sample complexity bounds for estimating probability divergences under invariances. arXiv preprint. Available at arXiv:2311.02868.

TRIEBEL, H. (2010). *Theory of Function Spaces II*, 1st ed. 1992. 2nd printing 2010. ed. *Modern Birkhäuser Classics*. Springer, Basel.

TSYBAKOV, A. B. (2008). *Introduction to Nonparametric Estimation*. Springer, Berlin.

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

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. MR0185765

The Annals of Statistics
2025, Vol. 53, No. 5, 2077–2102
https://doi.org/10.1214/25-AOS2530

A UNIFIED ANALYSIS OF LIKELIHOOD-BASED ESTIMATORS IN THE PLACKETT–LUCE MODEL

BY RUIJIAN HAN[1,a] AND YIMING XU[2,b]

[1]*Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University,* [a]*ruijian.han@polyu.edu.hk*
[2]*Department of Mathematics, University of Kentucky,* [b]*yiming.xu@uky.edu*

The Plackett–Luce model has been extensively used for rank aggregation in social choice theory. A central statistical question in this model concerns estimating the utility vector that governs the model's likelihood. In this paper, we investigate the asymptotic theory of utility vector estimation by maximizing different types of likelihood, such as full, marginal, and quasi-likelihood. Starting from interpreting the estimating equations of these estimators to gain some initial insights, we analyze their asymptotic behavior as the number of compared objects increases. In particular, we establish both uniform consistency and asymptotic normality of these estimators and discuss the trade-off between statistical efficiency and computational complexity. For generality, our results are proven for deterministic graph sequences under appropriate graph topology conditions. These conditions are shown to be informative when applied to common sampling scenarios, such as nonuniform random hypergraph models and hypergraph stochastic block models. Numerical results are provided to support our findings.

REFERENCES

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

AGARWAL, A., PATIL, P. and AGARWAL, S. (2018). Accelerated spectral ranking. In *ICML* 70–79. PMLR.

AZARI SOUFIANI, H., CHEN, W., PARKES, D. C. and XIA, L. (2013). Generalized method-of-moments for rank aggregation. In *NeurIPS* **26**.

BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *RecSys* 119–126.

BAUER, F. and JOST, J. (2013). Bipartite and neighborhood graphs and the spectrum of the normalized graph Laplace operator. *Comm. Anal. Geom.* **21** 787–845. MR3078942 https://doi.org/10.4310/CAG.2013.v21.n4.a2

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925 https://doi.org/10.2307/2334029

CARON, F. and DOUCET, A. (2012). Efficient Bayesian inference for generalized Bradley-Terry models. *J. Comput. Graph. Statist.* **21** 174–196. MR2913362 https://doi.org/10.1080/10618600.2012.638220

CARON, F., TEH, Y. W. and MURPHY, T. B. (2014). Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *Ann. Appl. Stat.* **8** 1145–1181. MR3262549 https://doi.org/10.1214/14-AOAS717

CHEN, P., GAO, C. and ZHANG, A. Y. (2022a). Optimal full ranking from pairwise comparisons. *Ann. Statist.* **50** 1775–1805. MR4441140 https://doi.org/10.1214/22-aos2175

CHEN, P., GAO, C. and ZHANG, A. Y. (2022b). Partial recovery for top-k ranking: Optimality of MLE and suboptimality of the spectral method. *Ann. Statist.* **50** 1618–1652. MR4441134 https://doi.org/10.1214/21-aos2166

CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023). Statistical inference for noisy incomplete binary matrix. *J. Mach. Learn. Res.* **24** Paper No. [95]. MR4582517

CHUNG, F. R. K. (1997). *Spectral Graph Theory*. *CBMS Regional Conference Series in Mathematics* **92**. Amer. Math. Soc., Providence, RI. MR1421568

MSC2020 subject classifications. Primary62F07; secondary62F12.
Key words and phrases. Asymptotic normality, entrywise error, hypergraph, likelihood-based estimation, Plackett–Luce model.

ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

FAN, J., LOU, Z., WANG, W. and YU, M. (2023). Spectral ranking inferences based on general multiway comparisons. arXiv preprint. Available at arXiv:2308.02918.

FAN, J., LOU, Z., WANG, W. and YU, M. (2025). Ranking inferences based on the top choice of multiway comparisons. *J. Amer. Statist. Assoc.* **120** 237–250. MR4893554 https://doi.org/10.1080/01621459.2024.2316364

FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *COLT* 943–959. PMLR.

GAO, C., SHEN, Y. and ZHANG, A. Y. (2023). Uncertainty quantification in the Bradley-Terry-Luce model. *Inf. Inference* **12** 1073–1140. MR4565758 https://doi.org/10.1093/imaiai/iaac032

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *NeurIPS* **27**.

GLICKMAN, M. E. and JONES, A. C. (1999). Rating the chess rating system. *Chance* **12** 21–28.

GUIVER, J. and SNELSON, E. (2009). Bayesian inference for Plackett–Luce ranking models. In *ICML* 377–384.

HAN, R. and XU, Y. (2025). Supplement to "A unified analysis of likelihood-based estimators in the Plackett–Luce model." https://doi.org/10.1214/25-AOS2530SUPP

HAN, R., XU, Y. and CHEN, K. (2023). A general pairwise comparison model for extremely sparse networks. *J. Amer. Statist. Assoc.* **118** 2422–2432. MR4681593 https://doi.org/10.1080/01621459.2022.2053137

HAN, R., YE, R., TAN, C. and CHEN, K. (2020). Asymptotic theory of sparse Bradley-Terry model. *Ann. Appl. Probab.* **30** 2491–2515. MR4149535 https://doi.org/10.1214/20-AAP1564

HASTIE, T. and TIBSHIRANI, R. (1998). Classification by pairwise coupling. *Ann. Statist.* **26** 451–471. MR1626055 https://doi.org/10.1214/aos/1028144844

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

HUNTER, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *Ann. Statist.* **32** 384–406. MR2051012 https://doi.org/10.1214/aos/1079120141

JANG, M., KIM, S. and SUH, C. (2018). Top-k rank aggregation from m-wise comparisons. *IEEE J. Sel. Top. Signal Process.* **12** 989–1004.

LIU, Y., FANG, E. X. and LU, J. (2023). Lagrangian inference for ranking problems. *Oper. Res.* **71** 202–223. MR4560195

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

MASSEY, K. (1997). Statistical models applied to the rating of sports teams. *Bluefield College* **1077**.

MAYSTRE, L. and GROSSGLAUSER, M. (2015). Fast and accurate inference of Plackett–Luce models. In *NeurIPS* **28**.

MCFADDEN, D. (1973). Conditional logit analysis of qualitative choice behavior. *Front. Econ.* 105–142.

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

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

OUYANG, L., WU, J., JIANG, X., ALMEIDA, D., WAINWRIGHT, C., MISHKIN, P., ZHANG, C., AGARWAL, S., SLAMA, K. et al. (2022). Training language models to follow instructions with human feedback. In *NeurIPS* **35** 27730–27744.

PAL, S. and ZHU, Y. (2021). Community detection in the sparse hypergraph stochastic block model. *Random Structures Algorithms* **59** 407–463. MR4295569 https://doi.org/10.1002/rsa.21006

PLACKETT, R. L. (1975). The analysis of permutations. *J. R. Stat. Soc., Ser. C* **24** 193–202. MR0391338 https://doi.org/10.2307/2346567

RAFAILOV, R., SHARMA, A., MITCHELL, E., MANNING, C. D., ERMON, S. and FINN, C. (2024). Direct preference optimization: Your language model is secretly a reward model. In *NeurIPS* **36**.

SHAH, N. B., BALAKRISHNAN, S., BRADLEY, J., PAREKH, A., RAMCHANDRAN, K. and WAINWRIGHT, M. J. (2016). Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *J. Mach. Learn. Res.* **17** Paper No. 58. MR3504618

SIMONS, G. and YAO, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.* **27** 1041–1060. MR1724040 https://doi.org/10.1214/aos/1018031267

THURSTONE, L. L. (1927). The method of paired comparisons for social values. *J. Abnorm. Psychol.* **21** 384.

YAN, T., YANG, Y. and XU, J. (2012). Sparse paired comparisons in the Bradley-Terry model. *Statist. Sinica* **22** 1305–1318. MR2987494 https://doi.org/10.5705/ss.2010.299

ZHAO, Z. and XIA, L. (2018). Composite marginal likelihood methods for random utility models. In *ICML* 5922–5931. PMLR.

ZHEN, Y. and WANG, J. (2023). Community detection in general hypergraph via graph embedding. *J. Amer. Statist. Assoc.* **118** 1620–1629. MR4646593 https://doi.org/10.1080/01621459.2021.2002157

The Annals of Statistics
2025, Vol. 53, No. 5, 2103–2127
https://doi.org/10.1214/25-AOS2531

THEORY OF FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS FOR DISCRETELY OBSERVED DATA

BY HANG ZHOU[1,a], DONGYI WEI[2,b] AND FANG YAO[3,c]

[1]*Department of Statistics and Operations Research & School of Data Science and Society, University of North Carolina at Chapel Hill,* [a]*hzh@unc.edu*

[2]*School of Mathematical Sciences, Peking University,* [b]*jnwdyi@pku.edu.cn*

[3]*School of Mathematical Sciences and Center for Statistical Science, Peking University,* [c]*fyao@math.pku.edu.cn*

Functional data analysis is an important research field in statistics which treats data as random functions drawn from some infinite-dimensional functional space, and functional principal component analysis (FPCA) based on eigen-decomposition plays a central role for data reduction and representation. After nearly three decades of research, there remains a key problem unsolved, namely, the perturbation analysis of covariance operator for diverging number of eigencomponents obtained from noisy and discretely observed data. This is fundamental for studying models and methods based on FPCA, while there has not been substantial progress since Hall, Müller and Wang (*Ann. Statist.* **34** (2006) 1493–1517)'s result for a fixed number of eigenfunction estimates. In this work, we aim to establish a unified theory for this problem, obtaining upper bounds for eigenfunctions with diverging indices in both the $\mathcal{L}^2$ and supremum norms, and deriving the asymptotic distributions of eigenvalues for a wide range of sampling schemes. Our results provide insight into the phenomenon when the $\mathcal{L}^2$ bound of eigenfunction estimates with diverging indices is minimax optimal as if the curves are fully observed, and reveal the transition of convergence rates from nonparametric to parametric regimes in connection to sparse or dense sampling. We also develop a double truncation technique to handle the uniform convergence of estimated covariance and eigenfunctions. The technical arguments in this work are useful for handling the perturbation series with noisy and discretely observed functional data and can be applied in models or those involving inverse problems based on FPCA as regularization, such as functional linear regression.

REFERENCES

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

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann*. *Statist*. **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694

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

BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann*. *Statist*. **50** 1144–1169. MR4404931 https://doi.org/10.1214/21-aos2143

BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans*. *Inf*. *Theory* **57** 2400–2419. MR2809098 https://doi.org/10.1109/TIT.2011.2111710

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann*. *Statist*. **1** 1071–1095. MR0348906

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

CAI, T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. Technical Report.

MSC2020 subject classifications. Primary 62R10; secondary 62G20.
Key words and phrases. Eigen-decomposition, functional data, perturbation series, phase transition.

CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830

CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870 https://doi.org/10.1214/11-AOS898

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

DING, X. and YANG, F. (2021). Spiked separable covariance matrices and principal components. *Ann. Statist.* **49** 1113–1138. MR4255121 https://doi.org/10.1214/20-aos1995

DOU, W. W., POLLARD, D. and ZHOU, H. H. (2012). Estimation in functional regression for general exponential families. *Ann. Statist.* **40** 2421–2451. MR3097608 https://doi.org/10.1214/12-AOS1027

EL KAROUI, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592 https://doi.org/10.1214/009117906000000917

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

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

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577 https://doi.org/10.1111/j.1467-9868.2005.00535.x

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365 https://doi.org/10.1214/009053606000000272

HÄRDLE, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179. MR1004333 https://doi.org/10.1016/0047-259X(89)90022-5

HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449. MR0964932 https://doi.org/10.1214/aos/1176351047

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. 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*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

INDRITZ, J. (1963). *Methods in Analysis*. The Macmillan Company, New York. MR0150991

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

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854 https://doi.org/10.1214/10-AOS813

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

NADLER, B., PENNA, F. and GARELLO, R. (2011). Performance of eigenvalue-based signal detectors with known and unknown noise level. In 2011 *IEEE Int. Conf. Commun.* 1–5. IEEE Press, New York.

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

PASTUR, L. and SHCHERBINA, M. (2011). *Eigenvalue Distribution of Large Random Matrices*. *Mathematical Surveys and Monographs* **171**. Amer. Math. Soc., Providence, RI. MR2808038 https://doi.org/10.1090/surv/171

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

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

PAUL, D. and PENG, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *Ann. Statist.* **37** 1229–1271. MR2509073 https://doi.org/10.1214/08-AOS608

QU, S., WANG, J.-L. and WANG, X. (2016). Optimal estimation for the functional Cox model. *Ann. Statist.* **44** 1708–1738. MR3519938 https://doi.org/10.1214/16-AOS1441

RAMSAY, J., HOOKER, G. and GRAVES, S. (2024). Package 'fda'.

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

RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259. MR1833314 https://doi.org/10.1111/j.0006-341X.2001.00253.x

SHAO, L., LIN, Z. and YAO, F. (2022). Intrinsic Riemannian functional data analysis for sparse longitudinal observations. *Ann. Statist.* **50** 1696–1721. MR4441137 https://doi.org/10.1214/22-aos2172

WAHL, M. (2022). Lower bounds for invariant statistical models with applications to principal component analysis. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 1565–1589. MR4452643 https://doi.org/10.1214/21-aihp1193

YAO, F. and LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 3–25. MR2212572 https://doi.org/10.1111/j.1467-9868.2005.00530.x

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). 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

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698 https://doi.org/10.1214/009053606000001505

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 https://doi.org/10.1214/16-AOS1446

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

ZHOU, H., WEI, D. and YAO, F. (2025). Supplement to "Theory of functional principal component analysis for discretely observed data." https://doi.org/10.1214/25-AOS2531SUPP

ZHOU, H., YAO, F. and ZHANG, H. (2023). Functional linear regression for discretely observed data: From ideal to reality. *Biometrika* **110** 381–393. MR4588359 https://doi.org/10.1093/biomet/asac053

The Annals of Statistics
2025, Vol. 53, No. 5, 2128–2152
https://doi.org/10.1214/25-AOS2533

ONLINE ESTIMATION AND INFERENCE FOR ROBUST POLICY EVALUATION IN REINFORCEMENT LEARNING

BY WEIDONG LIU[1,a], JIYUAN TU[2,b], XI CHEN[3,c] AND YICHEN ZHANG[4,d]

[1]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [a]*weidongl@sjtu.edu.cn*
[2]*School of Mathematics, Shanghai University of Finance and Economics,* [b]*tujy.19@gmail.com*
[3]*Stern School of Business, New York University,* [c]*xchen3@stern.nyu.edu*
[4]*Daniels School of Business, Purdue University,* [d]*zhang@purdue.edu*

Reinforcement learning has emerged as one of the prominent topics attracting attention in modern statistical learning, with policy evaluation being a key component. Unlike the traditional machine learning literature on this topic, our work emphasizes statistical inference for the model parameters and value functions of reinforcement learning algorithms. While most existing analyses assume random rewards to follow standard distributions, we embrace the concept of robust statistics in reinforcement learning by simultaneously addressing issues of outlier contamination and heavy-tailed rewards within a unified framework. In this paper, we develop a fully online robust policy evaluation procedure, and establish the Bahadur-type representation of our estimator. Furthermore, we develop an online procedure to efficiently conduct statistical inference based on the asymptotic distribution. This paper connects robust statistics and statistical inference in reinforcement learning, offering a more versatile and reliable approach to online policy evaluation. Finally, we validate the efficacy of our algorithm through numerical experiments conducted in simulations and real-world reinforcement learning experiments.

REFERENCES

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

BHANDARI, J., RUSSO, D. and SINGAL, R. (2018). A finite time analysis of temporal difference learning with linear function approximation. In *Conference on Learning Theory*.

BYRD, R. H., HANSEN, S. L., NOCEDAL, J. and SINGER, Y. (2016). A stochastic quasi-Newton method for large-scale optimization. *SIAM J. Optim.* **26** 1008–1031. MR3485979 https://doi.org/10.1137/140954362

CHARBONNIER, P., BLANC-FERAUD, L., AUBERT, G. and BARLAUD, M. (1994). Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of* 1*st International Conference on Image Processing*.

CHARBONNIER, P., BLANC-FERAUD, L., AUBERT, G. and BARLAUD, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.* **6** 298–311.

CHEN, H., LU, W. and SONG, R. (2021a). Statistical inference for online decision making: In a contextual bandit setting. *J. Amer. Statist. Assoc.* **116** 240–255. MR4227691 https://doi.org/10.1080/01621459.2020.1770098

CHEN, H., LU, W. and SONG, R. (2021b). Statistical inference for online decision making via stochastic gradient descent. *J. Amer. Statist. Assoc.* **116** 708–719. MR4270018 https://doi.org/10.1080/01621459.2020.1826325

CHEN, X., LAI, Z., LI, H. and ZHANG, Y. (2022). Online statistical inference for contextual bandits via stochastic gradient descent. Preprint. Available at arXiv:2212.14883.

CHEN, X., LAI, Z., LI, H. and ZHANG, Y. (2024). Online statistical inference for stochastic optimization via Kiefer-Wolfowitz methods. *J. Amer. Statist. Assoc.* **119** 2972–2982. MR4833930 https://doi.org/10.1080/01621459.2023.2296703

CHEN, X., LEE, J. D., TONG, X. T. and ZHANG, Y. (2020). Statistical inference for model parameters in stochastic gradient descent. *Ann. Statist.* **48** 251–273. MR4065161 https://doi.org/10.1214/18-AOS1801

MSC2020 subject classifications. Primary 62F12, 62M05; secondary 62F35.
Key words and phrases. Statistical inference, dependent samples, online policy evaluation, robust inference, Bahadur representation.

DAI, B., NACHUM, O., CHOW, Y., LI, L., SZEPESVÁRI, C. and SCHUURMANS, D. (2020). Coindice: Off-policy confidence interval estimation. *Adv. Neural Inf. Process. Syst.*

DESHPANDE, Y., MACKEY, L., SYRGKANIS, V. and TADDY, M. (2018). Accurate inference for adaptive linear models. In *International Conference on Machine Learning* 1194–1203. PMLR.

DIMAKOPOULOU, M., REN, Z. and ZHOU, Z. (2021). Online multi-armed bandits with adaptive inference. *Adv. Neural Inf. Process. Syst.* **34** 1939–1951.

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

DURMUS, A., MOULINES, E., NAUMOV, A., SAMSONOV, S. and WAI, H.-T. (2021). On the stability of random matrix product with Markovian noise: Application to linear stochastic approximation and TD learning. In *Proceedings of Thirty Fourth Conference on Learning Theory*.

EVERITT, T., KRAKOVNA, V., ORSEAU, L. and LEGG, S. (2017). Reinforcement learning with a corrupted reward channel. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*.

FAN, J., GUO, Y. and JIANG, B. (2022). Adaptive Huber regression on Markov-dependent data. *Stochastic Process. Appl.* **150** 802–818. MR4440165 https://doi.org/10.1016/j.spa.2019.09.004

FANG, Y., XU, J. and YANG, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *J. Mach. Learn. Res.* **19** Paper No. 78, 21. MR3899780

FENG, Y., TANG, Z., ZHANG, N. and LIU, Q. (2021). Non-asymptotic confidence intervals of off-policy evaluation: Primal and dual bounds. In *International Conference on Learning Representations*.

GIVCHI, A. and PALHANG, M. (2015). Quasi Newton temporal difference learning. In *Proceedings of the Sixth Asian Conference on Machine Learning*.

HADAD, V., HIRSHBERG, D. A., ZHAN, R., WAGER, S. and ATHEY, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proc. Natl. Acad. Sci. USA* **118** Paper No. e2014602118, 10. MR4294058 https://doi.org/10.1073/pnas.2014602118

HAN, Q., SUN, W. W. and ZHANG, Y. (2022). Online Statistical Inference in Decision Making with Matrix Context. Preprint. Available at arXiv:2212.11385.

HANNA, J., STONE, P. and NIEKUM, S. (2017). Bootstrapping with models: Confidence intervals for off-policy evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

HAO, B., JI, X., DUAN, Y., LU, H., SZEPESVARI, C. and WANG, M. (2021). Bootstrapping fitted q-evaluation for off-policy inference. In *International Conference on Machine Learning*.

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

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

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415 https://doi.org/10.1214/aoms/1177703732

HUBER, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

HUBER, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in Statistics*. Springer, Berlin.

JIANG, N. and HUANG, J. (2020). Minimax value interval for off-policy evaluation and policy optimization. *Adv. Neural Inf. Process. Syst.*

JOHNSON, A. E. W., POLLARD, T. J., SHEN, L. et al. (2016). MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**.

KOLTER, J. Z. and NG, A. Y. (2009). Regularization and feature selection in least-squares temporal difference learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*.

KOMOROWSKI, M., CELI, L. A., BADAWI, O. et al. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **24** 1716–1720.

KORMUSHEV, P., CALINON, S. and CALDWELL, D. G. (2013). Reinforcement learning in robotics: Applications and real-world challenges. *Robotics* **2** 122–148.

LI, X. and SUN, Q. (2023). Variance-aware robust reinforcement learning with linear function approximation under heavy-tailed rewards. e-prints. Available at arXiv:2303.05606.

LIU, W., TU, J., CHEN, X. and ZHANG, Y. (2025). Supplement to “Online estimation and inference for robust policy evaluation in reinforcement learning.” https://doi.org/10.1214/25-AOS2533SUPP

MNIH, V., KAVUKCUOGLU, K., SILVER, D., RUSU, A. A., VENESS, J., BELLEMARE, M. G., GRAVES, A., RIEDMILLER, M., FIDJELAND, A. K. et al. (2015). Human-level control through deep reinforcement learning. *Nature* **518** 529–533.

MOU, W., PANANJADY, A., WAINWRIGHT, M. J. and BARTLETT, P. L. (2024). Optimal and instance-dependent guarantees for Markovian linear stochastic approximation. *Math. Stat. Learn.* **7** 41–153. MR4733910 https://doi.org/10.4171/msl/44

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

NESTEROV, Y. (2004). *Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization* **87**. Kluwer Academic, Boston, MA. MR2142598 https://doi.org/10.1007/978-1-4419-8853-9

POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

PRASAD, N., CHENG, L.-F., CHIVERS, C., DRAUGELIS, M. and ENGELHARDT, B. E. (2017). A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. In *33rd Conference on Uncertainty in Artificial Intelligence*.

RAMPRASAD, P., LI, Y., YANG, Z., WANG, Z., SUN, W. W. and CHENG, G. (2023). Online bootstrap inference for policy evaluation in reinforcement learning. *J. Amer. Statist. Assoc.* **118** 2901–2914. MR4681629 https://doi.org/10.1080/01621459.2022.2096620

RIO, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes. Probability Theory and Stochastic Modelling* **80**. Springer, Berlin. MR3642873 https://doi.org/10.1007/978-3-662-54323-8

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

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

RUPPERT, D. (1985). A Newton-Raphson version of the multivariate Robbins-Monro procedure. *Ann. Statist.* **13** 236–245. MR0773164 https://doi.org/10.1214/aos/1176346589

RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process Technical Report, Cornell Univ. Operations Research and Industrial Engineering.

SCHRAUDOLPH, N. N., YU, J. and GÜNTER, S. (2007). A stochastic quasi-Newton method for online convex optimization. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*.

SHAO, Q.-M. and ZHANG, Z.-S. (2022). Berry-Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms. *Bernoulli* **28** 1548–1576. MR4411502 https://doi.org/10.3150/21-bej1336

SHEN, Y., CAI, H. and SONG, R. (2024). Doubly robust interval estimation for optimal policy evaluation in online learning. *J. Amer. Statist. Assoc.* **119** 2811–2821. MR4833917 https://doi.org/10.1080/01621459.2023.2279289

SHI, C., FAN, A., SONG, R. and LU, W. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Statist.* **46** 925–957. MR3797992 https://doi.org/10.1214/17-AOS1570

SHI, C., SONG, R., LU, W. and LI, R. (2021). Statistical inference for high-dimensional models via recursive online-score estimation. *J. Amer. Statist. Assoc.* **116** 1307–1318. MR4309274 https://doi.org/10.1080/01621459.2019.1710154

SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 765–793. MR4460575

SUN, Q., ZHOU, W.-X. and FAN, J. (2020). Adaptive Huber regression. *J. Amer. Statist. Assoc.* **115** 254–265. MR4078461 https://doi.org/10.1080/01621459.2018.1543124

SUTTON, R. (1988). Learning to predict by the method of temporal differences. *Mach. Learn.* **3** 9–44.

SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951

SUTTON, R. S., MAEI, H. R., PRECUP, D., BHATNAGAR, S., SILVER, D., SZEPESVÁRI, C. and WIEWIORA, E. (2009). Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *Proceedings of the 26th Annual International Conference on Machine Learning*.

SYRGKANIS, V. and ZHAN, R. (2023). Post-Episodic Reinforcement Learning Inference. e-prints. Available at arXiv:2302.08854.

THOMAS, P., THEOCHAROUS, G. and GHAVAMZADEH, M. (2015). High confidence policy improvement. In *International Conference on Machine Learning*. PMLR.

TSITSIKLIS, J. N. and VAN ROY, B. (1997). An analysis of temporal-difference learning with function approximation. *IEEE Trans. Automat. Control* **42** 674–690. MR1454208 https://doi.org/10.1109/9.580874

WANG, J., LIU, Y. and LI, B. (2020). Reinforcement learning with perturbed rewards. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

XIA, E., KHAMARU, K., WAINWRIGHT, M. J. and JORDAN, M. I. (2023). Instance-dependent confidence and early stopping for reinforcement learning. *J. Mach. Learn. Res.* **24** Paper No. [392], 43. MR4720848

ZHAN, R., HADAD, V., HIRSHBERG, D. A. and ATHEY, S. (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* 2125–2135.

ZHANG, K., JANSON, L. and MURPHY, S. (2021). Statistical inference with M-estimators on adaptively collected data. *Adv. Neural Inf. Process. Syst.* **34** 7460–7471.

ZHANG, K., JANSON, L. and MURPHY, S. (2022). Statistical inference after adaptive sampling in non-Markovian environments. Preprint. Available at arXiv:2202.07098.

ZHU, J., WAN, R., QI, Z., LUO, S. and SHI, C. (2024). Robust offline reinforcement learning with heavy-tailed rewards. In *International Conference on Artificial Intelligence and Statistics*.

The Annals of Statistics
2025, Vol. 53, No. 5, 2153–2178
https://doi.org/10.1214/25-AOS2536

TRIMMED SAMPLE MEANS FOR ROBUST UNIFORM MEAN ESTIMATION AND REGRESSION

BY ROBERTO I. OLIVEIRA[1,a] AND LUCAS RESENDE[2,b]

[1]*IMPA*, [a]*rimfo@impa.br*
[2]*ENSAE*, [b]*lucas.resende@ensae.fr*

It is well known that trimmed sample means are robust against heavy tails and data contamination. This paper analyzes the performance of trimmed means and related methods in two novel contexts. The first one consists of estimating expectations of functions in a given family, with uniform error bounds; this is closely related to the problem of estimating the mean of a random vector under a general norm. The second problem considered is that of regression with quadratic loss. In both cases, trimmed-mean-based estimators are the first to obtain optimal dependence on the (adversarial) contamination level. Moreover, they also match or improve upon the state of the art in terms of heavy tails. Experiments with synthetic data show that a natural "trimmed mean linear regression" method often performs better than both ordinary least squares and alternative methods based on median-of-means.

REFERENCES

[1] ABDALLA, P. and ZHIVOTOVSKIY, N. (2024). Covariance estimation: Optimal dimension-free guarantees for adversarial corruption and heavy tails. *J. Eur. Math. Soc.*

[2] ALON, N. and SPENCER, J. H. (2016). *The Probabilistic Method*, 4th ed. *Wiley Series in Discrete Mathematics and Optimization*. Wiley, Hoboken, NJ. MR3524748

[3] AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. MR2906886 https://doi.org/10.1214/11-AOS918

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

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

[6] BOUSQUET, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500. MR1890640 https://doi.org/10.1016/S1631-073X(02)02292-6

[7] BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. MR3405602 https://doi.org/10.1214/15-AOS1350

[8] BUBECK, S., CESA-BIANCHI, N. and LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Inf. Theory* **59** 7711–7717. MR3124669 https://doi.org/10.1109/TIT.2013.2277869

[9] CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

[10] CHERAPANAMJERI, Y., TRIPURANENI, N., BARTLETT, P. and JORDAN, M. (2022). Optimal mean estimation without a variance. In *Conference on Learning Theory* 356–357. PMLR.

[11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

[12] DEPERSIN, J. and LECUÉ, G. (2022). Robust sub-Gaussian estimation of a mean vector in nearly linear time. *Ann. Statist.* **50** 511–536. MR4382026 https://doi.org/10.1214/21-aos2118

[13] DEPERSIN, J. and LECUÉ, G. (2022). Optimal robust mean and location estimation via convex programs with respect to any pseudo-norms. *Probab. Theory Related Fields* **183** 997–1025. MR4453320 https://doi.org/10.1007/s00440-022-01127-y

[14] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440

MSC2020 subject classifications. Primary 62G35, 62G08; secondary 62F35, 62J05, 62C20.
Key words and phrases. Sub-Gaussian estimators, robust estimators, regression.

[15] DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput.* **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680

[16] DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., STEINHARDT, J. and STEWART, A. (2019). Sever: A robust meta-algorithm for stochastic optimization. In *Proceedings of the* 36*th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 1596–1606. PMLR.

[17] DIAKONIKOLAS, I., KANE, D. M., PENSIA, A. and PITTAS, T. (2022). Streaming algorithms for high-dimensional robust statistics. In *International Conference on Machine Learning*, *ICML* 2022, 17–23 *July* 2022, *Baltimore*, *Maryland*, *USA* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu and S. Sabato, eds.). *Proceedings of Machine Learning Research* **162**. 5061–5117. PMLR.

[18] DONG, Y., HOPKINS, S. and LI, J. (2019). Quantum entropy scoring for fast robust mean estimation and improved outlier detection. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook.

[19] GEER, S. A. (2000). *Empirical Processes in M-Estimation*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, Cambridge.

[20] HALL, P. (1981). Large sample properties of Jaeckel's adaptive trimmed mean. *Ann. Inst. Statist. Math.* **33** 449–462. MR0637757 https://doi.org/10.1007/BF02480955

[21] HOPKINS, S. B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843

[22] HOPKINS, S. B., LI, J. and ZHANG, F. (2020). Robust and heavy-tailed mean estimation made simple, via regret minimization. In *Advances in Neural Information Processing Systems* 33: *Annual Conference on Neural Information Processing Systems* 2020, *NeurIPS* 2020, *December* 6–12, 2020, *Virtual* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin, eds.).

[23] HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758. MR0185747 https://doi.org/10.1214/aoms/1177699803

[24] HUBER, P. J. (1972). The 1972 Wald lecture. Robust statistics: A review. *Ann. Math. Statist.* **43** 1041–1067. MR0314180 https://doi.org/10.1214/aoms/1177692459

[25] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

[26] JAECKEL, L. A. (1971). Some flexible estimates of location. *Ann. Math. Statist.* **42** 1540–1552. MR0350951 https://doi.org/10.1214/aoms/1177693152

[27] JANA JUREČKOVÁ, A. H. W., KOENKER, R. and WELSH, A. H. (1994). Adaptive choice of trimming proportions. *Ann. Inst. Statist. Math.* **46** 737–755. MR1325993 https://doi.org/10.1007/BF00773479

[28] LECUÉ, G. and LERASLE, M. (2019). Learning from MOM's principles: Le Cam's approach. *Stochastic Process. Appl.* **129** 4385–4410. MR4013866 https://doi.org/10.1016/j.spa.2018.11.024

[29] LECUÉ, G. and LERASLE, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. MR4102681 https://doi.org/10.1214/19-AOS1828

[30] LECUÉ, G. and MENDELSON, S. (2013). Learning subgaussian classes: Upper and minimax bounds. arXiv preprint. Available at arXiv:1305.4825.

[31] LEE, J. C. H. and VALIANT, P. (2022). Optimal sub-Gaussian mean estimation in $\mathbb{R}$. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 672–683. IEEE Comput. Soc., Los Alamitos, CA. MR4399724 https://doi.org/10.1109/FOCS52979.2021.00071

[32] LIAN, C., RONG, Y. and CHENG, W. (2025). On a novel skewed generalized t distribution: Properties, estimations, and its applications. *Comm. Statist. Theory Methods* **54** 396–417. MR4839831 https://doi.org/10.1080/03610926.2024.2313034

[33] LUGOSI, G. and MENDELSON, S. (2019). Near-optimal mean estimators with respect to general norms. *Probab. Theory Related Fields* **175** 957–973. MR4026610 https://doi.org/10.1007/s00440-019-00906-4

[34] LUGOSI, G. and MENDELSON, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x

[35] LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950 https://doi.org/10.1214/17-AOS1639

[36] LUGOSI, G. and MENDELSON, S. (2019). Risk minimization by median-of-means tournaments. *J. Eur. Math. Soc.* (*JEMS*) **22** 925–965. MR4055993 https://doi.org/10.4171/jems/937

[37] LUGOSI, G. and MENDELSON, S. (2021). Robust multivariate mean estimation: The optimality of trimmed mean. *Ann. Statist.* **49** 393–410. MR4206683 https://doi.org/10.1214/20-AOS1961

[38] MASSART, P. (2000). Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math.* **9** 245–303. MR1813803
[39] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25. MR3367000 https://doi.org/10.1145/2699439
[40] MENDELSON, S. and ZHIVOTOVSKIY, N. (2020). Robust covariance estimation under L_4-L_2 norm equivalence. *Ann. Statist.* **48** 1648–1664. MR4124338 https://doi.org/10.1214/19-AOS1862
[41] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645
[42] MINSKER, S. (2018). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Ann. Statist.* **46** 2871–2903. MR3851758 https://doi.org/10.1214/17-AOS1642
[43] MINSKER, S. (2025). Uniform bounds for robust mean estimators. *Stochastic Process. Appl.* **190** Paper No. 104724. MR4933834 https://doi.org/10.1016/j.spa.2025.104724
[44] MOURTADA, J., VAŠKEVIČIUS, T. and ZHIVOTOVSKIY, N. (2021). Distribution-free robust linear regression. *Math. Stat. Learn.* **4** 253–292. MR4383735 https://doi.org/10.4171/msl/27
[45] OLIVEIRA, R. I. and RESENDE, L. (2025). Supplement to "Trimmed sample means for robust uniform mean estimation and regression." https://doi.org/10.1214/25-AOS2536SUPPA, https://doi.org/10.1214/25-AOS2536SUPPB
[46] OLIVEIRA, R. I. and RICO, Z. F. (2024). Improved covariance estimation: Optimal robustness and sub-Gaussian guarantees under heavy tails. *Ann. Statist.* **52** 1953–1977. MR4829476 https://doi.org/10.1214/24-aos2407
[47] RICO, Z. F. (2022). Optimal statistical estimation: Sub-Gaussian properties, heavy-tailed data, and robustness Ph.D. thesis, Instituto de Matemática Pura e Aplicada (IMPA).
[48] SRIPERUMBUDUR, B. (2016). On the optimal estimation of probability measures in weak and strong topologies. *Bernoulli* **22** 1839–1893. MR3474835 https://doi.org/10.3150/15-BEJ713
[49] SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2012). On the empirical estimation of integral probability metrics. *Electron. J. Stat.* **6** 1550–1599. MR2988458 https://doi.org/10.1214/12-EJS722
[50] STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477. MR0359134
[51] STIGLER, S. M. (2010). The changing history of robustness. *Amer. Statist.* **64** 277–281. MR2758558 https://doi.org/10.1198/tast.2010.10159
[52] STIGLER, S. M. (2016). *The Seven Pillars of Statistical Wisdom*. Harvard Univ. Press, Cambridge, MA. MR3585675 https://doi.org/10.4159/9780674970199
[53] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006 https://doi.org/10.1007/s002220050108
[54] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

The Annals of Statistics
2025, Vol. 53, No. 5, 2179–2203
https://doi.org/10.1214/25-AOS2538

YURINSKII'S COUPLING FOR MARTINGALES

BY MATIAS D. CATTANEO[1,a], RICARDO P. MASINI[2,b] AND WILLIAM G. UNDERWOOD[3,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*cattaneo@princeton.edu*
[2]*Department of Statistics, University of California, Davis,* [b]*rmasini@ucdavis.edu*
[3]*Statistical Laboratory, University of Cambridge,* [c]*wgu21@cam.ac.uk*

Yurinskii's coupling is a popular theoretical tool for nonasymptotic distributional analysis in mathematical statistics and applied probability, offering a Gaussian strong approximation with an explicit error bound under easily verifiable conditions. Originally stated in ℓ_2-norm for sums of independent random vectors, it has recently been extended both to the ℓ_p-norm, for $1 \le p \le \infty$, and to vector-valued martingales in ℓ_2-norm, under some strong conditions. We present as our main result a Yurinskii coupling for approximate martingales in ℓ_p-norm, under substantially weaker conditions than those previously imposed. Our formulation further allows for the coupling variable to follow a more general Gaussian mixture distribution, and we provide a novel third-order coupling method, which gives tighter approximations in certain settings. We specialize our main result to mixingales, martingales, and independent data, and derive uniform Gaussian mixture strong approximations for martingale empirical processes. Applications to nonparametric partitioning-based and local polynomial regression procedures are provided, alongside central limit theorems for high-dimensional martingale vectors.

REFERENCES

[1] ANASTASIOU, A., BALASUBRAMANIAN, K. and ERDOGDU, M. A. (2019). Normal approximation for stochastic gradient descent via non-asymptotic rates of martingale CLT. In *Conference on Learning Theory* 115–137. PMLR.

[2] ATCHADÉ, Y. F. and CATTANEO, M. D. (2014). A martingale decomposition for quadratic forms of Markov chains (with applications). *Stochastic Process*. *Appl*. **124** 646–677. MR3131309 https://doi.org/10.1016/j.spa.2013.09.001

[3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and FERNÁNDEZ-VAL, I. (2019). Conditional quantile processes based on series or many regressors. *J. Econometrics* **213** 4–29. MR4013213 https://doi.org/10.1016/j.jeconom.2019.04.003

[4] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

[5] BELLONI, A. and OLIVEIRA, R. I. (2018). A high dimensional central limit theorem for martingales, with applications to context tree models. Available at arXiv:1809.02741.

[6] BERTHET, P. and MASON, D. M. (2006). Revisiting two strong approximation results of Dudley and Philipp. In *High Dimensional Probability*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 155–172. IMS, Beachwood, OH. MR2387767 https://doi.org/10.1214/074921706000000824

[7] BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[8] BIAU, G. and MASON, D. M. (2015). High-dimensional p-norms. In *Mathematical Statistics and Limit Theorems* 21–40. Springer, Cham. MR3380729

[9] 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

MSC2020 subject classifications. Primary 62E20, 62G20; secondary 60G42.
Key words and phrases. Coupling, strong approximation, mixingales, martingales, dependent data, Gaussian mixture approximation, time series, empirical processes, uniform inference, series estimation, local polynomial estimation, central limit theorems.

[10] BUZUN, N., SHVETSOV, N. and DYLOV, D. V. (2022). Strong Gaussian approximation for the sum of random vectors. In *Conference on Learning Theory* **178** 1693–1715. PMLR.

[11] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2020). Large sample properties of partitioning-based series estimators. *Ann. Statist.* **48** 1718–1741. MR4124341 https://doi.org/10.1214/19-AOS1865

[12] CATTANEO, M. D., FENG, Y. and UNDERWOOD, W. G. (2024). Uniform inference for kernel density estimators with dyadic data. *J. Amer. Statist. Assoc.* **119** 2695–2708. MR4833908 https://doi.org/10.1080/01621459.2023.2272785

[13] CATTANEO, M. D., MASINI, R. P. and UNDERWOOD, W. G. (2025). Supplement to "Yurinskii's Coupling for Martingales." https://doi.org/10.1214/25-AOS2538SUPP

[14] CATTANEO, M. D., MASINI, R. P. and UNDERWOOD, W. G. (2025). Sharp anti-concentration inequalities for extremum statistics via copulas. Available at arXiv:2502.07699.

[15] CATTANEO, M. D. and YU, R. R. (2025). Strong approximations for empirical processes indexed by Lipschitz functions. *Ann. Statist.* **53** 1203–1229. MR4925121 https://doi.org/10.1214/25-aos2500

[16] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

[17] 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

[18] 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

[19] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

[20] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

[21] 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

[22] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

[23] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. *Probability and Mathematical Statistics*. Academic Press, New York. MR0666546

[24] CUNY, C. and MERLEVÈDE, F. (2014). On martingale approximations and the quenched weak invariance principle. *Ann. Probab.* **42** 760–793. MR3178473 https://doi.org/10.1214/13-AOP856

[25] DEDECKER, J., MERLEVÈDE, F. and VOLNÝ, D. (2007). On the weak invariance principle for non-adapted sequences under projective criteria. *J. Theoret. Probab.* **20** 971–1004. MR2359065 https://doi.org/10.1007/s10959-007-0090-1

[26] DEHLING, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **63** 393–432. MR0705631 https://doi.org/10.1007/BF00542537

[27] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712 https://doi.org/10.1017/CBO9780511665622

[28] DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552. MR0690575 https://doi.org/10.1007/BF00534202

[29] EGGERMONT, P. P. B. and LARICCIA, V. N. (2009). *Maximum Penalized Likelihood Estimation. Volume II: Regression*. *Springer Series in Statistics*. Springer, Dordrecht. MR2817245 https://doi.org/10.1007/b12285

[30] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[31] FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press, Boca Raton.

[32] GIESSING, A. (2023). Anti-concentration of suprema of Gaussian processes and Gaussian order statistics. Available at arXiv:2310.12119.

[33] GINÉ, E., KOLTCHINSKII, V. and SAKHANENKO, L. (2004). Kernel density estimators: Convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167–198. MR2093761 https://doi.org/10.1007/s00440-004-0339-x

[34] KOCK, A. B. and PREINERSTORFER, D. (2024). A remark on moment-dependent phase transitions in high-dimensional Gaussian approximations. *Statist. Probab. Lett.* **211** 110149. MR4747018 https://doi.org/10.1016/j.spl.2024.110149

[35] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[36] LE CAM, L. (1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge. Technical Report.

[37] LI, J. and LIAO, Z. (2020). Uniform nonparametric inference for time series. *J. Econometrics* **219** 38–51. MR4152784 https://doi.org/10.1016/j.jeconom.2019.09.011

[38] LINDVALL, T. (1992). *Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR1180522

[39] 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

[40] MAGDA, P. and ZHANG, N. (2018). Martingale approximations for random fields. *Electron. Commun. Probab.* **23** 1–9.

[41] MCLEISH, D. L. (1975). Invariance principles for dependent variables. *Z. Wahrsch. Verw. Gebiete* **32** 165–178. MR0388483 https://doi.org/10.1007/BF00532611

[42] MONRAD, D. and PHILIPP, W. (1991). Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. *Probab. Theory Related Fields* **88** 381–404. MR1100898 https://doi.org/10.1007/BF01418867

[43] NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **1807** 169–187. Springer, Berlin. MR2083397 https://doi.org/10.1007/978-3-540-36428-3_15

[44] PELIGRAD, M. (2010). Conditional central limit theorem via martingale approximation. In *Dependence in Probability, Analysis and Number Theory* 295–309. Kendrick Press, Heber City, UT. MR2731064

[45] POLLARD, D. (2002). *A User's Guide to Measure Theoretic Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics* **8**. Cambridge Univ. Press, Cambridge. MR1873379

[46] RAKHLIN, A., SRIDHARAN, K. and TEWARI, A. (2015). Sequential complexities and uniform martingale laws of large numbers. *Probab. Theory Related Fields* **161** 111–153. MR3304748 https://doi.org/10.1007/s00440-013-0545-5

[47] RAY, K. and VAN DER VAART, A. (2021). On the Bernstein–von Mises theorem for the Dirichlet process. *Electron. J. Stat.* **15** 2224–2246. MR4255307 https://doi.org/10.1214/21-ejs1821

[48] RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45. MR1254823 https://doi.org/10.1007/BF01311347

[49] SHEEHY, A. and WELLNER, J. A. (1992). Uniform Donsker classes of functions. *Ann. Probab.* **20** 1983–2030. MR1188051

[50] VAN DE GEER, S. A. (2000). *Empirical Processes in M-Estimation* **6**. Cambridge Univ. Press, Cambridge.

[51] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[52] 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

[53] WU, W. B. and WOODROOFE, M. (2004). Martingale approximations for sums of stationary processes. *Ann. Probab.* **32** 1674–1690. MR2060314 https://doi.org/10.1214/009117904000000351

[54] YURINSKII, V. V. (1978). On the error of the Gaussian approximation for convolutions. *Theory Probab. Appl.* **22** 236–247.

[55] ZAITSEV, A. Y. (1987). Estimates of the Lévy–Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments. *Theory Probab. Appl.* **31** 203–220.

[56] ZAĬTSEV, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Theory Related Fields* **74** 535–566. MR0876255 https://doi.org/10.1007/BF00363515

[57] ZHAO, O. and WOODROOFE, M. (2008). On martingale approximations. *Ann. Appl. Probab.* **18** 1831–1847. MR2462550 https://doi.org/10.1214/07-AAP505

The Annals of Statistics
2025, Vol. 53, No. 5, 2204–2229
https://doi.org/10.1214/25-AOS2540

TESTS OF MISSING COMPLETELY AT RANDOM BASED ON SAMPLE COVARIANCE MATRICES

BY ALBERTO BORDINO[a] AND THOMAS B. BERRETT[b]

Department of Statistics, University of Warwick, [a]*alberto.bordino@warwick.ac.uk,* [b]*tom.berrett@warwick.ac.uk*

We study the problem of testing whether the missing values of a potentially high-dimensional dataset are Missing Completely at Random (MCAR). We relax the problem of testing MCAR to the problem of testing the compatibility of a collection of covariance matrices, motivated by the fact that this procedure is feasible when the dimension grows with the sample size. Our first contributions are to define a natural measure of the incompatibility of a collection of correlation matrices, which can be characterised as the optimal value of a Semidefinite Programming (SDP) problem, and to establish a key duality result allowing its practical computation and interpretation. By analysing the concentration properties of the natural plug-in estimator for this measure, we propose a novel hypothesis test, which is calibrated via a bootstrap procedure and demonstrates power against any distribution with incompatible covariance matrices. By considering key examples of missingness structures, we demonstrate that our procedures are minimax rate optimal in certain cases. We further validate our methodology with numerical simulations that provide evidence of validity and power, even when data are heavy tailed. Furthermore, tests of compatibility can be used to test the feasibility of positive semidefinite matrix completion problems with noisy observations, and thus our results may be of independent interest.

REFERENCES

ALEKSIĆ, D. (2024). A novel test of missing completely at random: U-statistics-based approach. *Statistics* **58** 1004–1023. MR4803554 https://doi.org/10.1080/02331888.2024.2386361

ALEKSIĆ, D., CUPARIĆ, M. and MILOŠEVIĆ, B. (2023). Non-degenerate U-statistics for data missing completely at random with application to testing independence. *Stat* **12** Paper No. e634, 14 pp. MR4681047 https://doi.org/10.1002/sta4.634

ANDREWS, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009 https://doi.org/10.1111/1468-0262.00114

BENTLER, P. M. and BERKANE, M. (1986). Greatest lower bound to the elliptical theory kurtosis parameter. *Biometrika* **73** 240–241. MR0836456 https://doi.org/10.1093/biomet/73.1.240

BERRETT, T. B., BORDINO, A., DUISENBEKOV, D., JAFFE, S. and SAMWORTH, R. J. (2022). MCARtest: Optimal nonparametric testing of missing completely at random. R package version 1.3.

BERRETT, T. B. and SAMWORTH, R. J. (2023). Optimal nonparametric testing of missing completely at random and its connections to compatibility. *Ann. Statist.* **51** 2170–2193. MR4678800 https://doi.org/10.1214/23-aos2326

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

BLEKHERMAN, G., PARRILO, P. A. and THOMAS, R. R. (2012). *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, Philadelphia, PA. https://doi.org/10.1137/1.9781611972290

BORDINO, A. and BERRETT, T. B. (2025). Supplement to "Tests of missing completely at random based on sample covariance matrices." https://doi.org/10.1214/25-AOS2540SUPPA, https://doi.org/10.1214/25-AOS2540SUPPB

MSC2020 subject classifications. 62G10.
Key words and phrases. Missing data, missing completely at random, minimax testing, semidefinite programming.

BROWN, M. B. and FORSYTHE, A. B. (1974). Robust tests for the equality of variances. *J. Amer. Statist. Assoc.* **69** 364–367. https://doi.org/10.1080/01621459.1974.10482955

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155

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

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

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

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

CAVALIERE, G., NIELSEN, H. B. and RAHBEK, A. (2017). On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *J. Time Series Anal.* **38** 513–534. MR3664645 https://doi.org/10.1111/jtsa.12214

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, Methodol.* **39** 1–38. MR0501537 https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

ELSENER, A. and VAN DE GEER, S. (2019). Sparse spectral estimation with missing and corrupted measurements. *Stat* **8** e229, 11 pp. MR3978409 https://doi.org/10.1002/sta4.229

FOLLAIN, B., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional changepoint estimation with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1023–1055. MR4460584 https://doi.org/10.1111/rssb.12540

FUCHS, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *J. Amer. Statist. Assoc.* **77** 270–278. https://doi.org/10.1080/01621459.1982.10477795

GASTWIRTH, J. L., GEL, Y. R. and MIAO, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statist. Sci.* **24** 343–360. MR2757435 https://doi.org/10.1214/09-STS301

GRONE, R., JOHNSON, C. R., DE SÁ, E. M. and WOLKOWICZ, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.* **58** 109–124. MR0739282 https://doi.org/10.1016/0024-3795(84)90207-6

HAWKINS, D. M. (1981). A new test for multivariate normality and homoscedasticity. *Technometrics* **23** 105–110. MR0604914 https://doi.org/10.2307/1267983

HOFERT, M., KOJADINOVIC, I., MAECHLER, M. and YAN, J. (2020). copula: Multivariate dependence with copulas. R package version 1.0-0.

JAMSHIDIAN, M. and JALAL, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika* **75** 649–674. MR2741492 https://doi.org/10.1007/s11336-010-9175-3

JIAO, J., HAN, Y. and WEISSMAN, T. (2018). Minimax estimation of the L_1 distance. *IEEE Trans. Inf. Theory* **64** 6672–6706. MR3860754 https://doi.org/10.1109/TIT.2018.2846245

KELLERER, H. G. (1984). Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete* **67** 399–432. MR0761565 https://doi.org/10.1007/BF00532047

KIM, K. H. and BENTLER, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika* **67** 609–623. MR2227891 https://doi.org/10.1007/BF02295134

LAURENT, M. (2009). Matrix completion problems. *Encycl. Optim.* **3** 221–229. https://doi.org/10.1007/978-0-387-74759-0_355

LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. *Springer Texts in Statistics*. Springer, New York. MR1663158 https://doi.org/10.1007/b98855

LI, J. and YU, Y. (2015). A nonparametric test of missing completely at random for incomplete multivariate data. *Psychometrika* **80** 707–726. MR3392026 https://doi.org/10.1007/s11336-014-9410-4

LITTLE, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.* **83** 1198–1202. MR0997603 https://doi.org/10.1080/01621459.1988.10478722

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*. *Wiley Series in Probability and Mathematical Statistics*. *Probability and Mathematical Statistics*. Wiley, New York. https://doi.org/10.1002/9781119013563

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

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

MOSTELLER, F. and FISHER, R. A. (1948). Questions and answers. *Amer. Statist.* **2** 30–31. https://doi.org/10.1080/00031305.1948.10483405

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932 https://doi.org/10.1002/9780470316559

NESTEROV, Y. and NEMIROVSKII, A. (1994). *Interior-Point Polynomial Algorithms in Convex Programming*. *SIAM Studies in Applied Mathematics* **13**. SIAM, Philadelphia, PA. MR1258086 https://doi.org/10.1137/1.9781611970791

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

ROCKEL, T. (2020). missMethods: Methods for missing data. R package version 0.2.0.

SAMWORTH, R. (2003). A note on methods of restoring consistency to the bootstrap. *Biometrika* **90** 985–990. MR2024773 https://doi.org/10.1093/biomet/90.4.985

SANTOS, M. S., PEREIRA, R. C., COSTA, A. F., SOARES, J. P., SANTOS, J. and ABREU, P. H. (2019). Generating synthetic missing data: A review by missing mechanism. *IEEE Access* **7** 11651–11667. https://doi.org/10.1109/ACCESS.2019.2891360

SEAMAN, S., GALATI, J., JACKSON, D. and CARLIN, J. (2013). What is meant by "missing at random"? *Statist. Sci.* **28** 257–268. MR3112409 https://doi.org/10.1214/13-sts415

SELL, T., BERRETT, T. B. and CANNINGS, T. I. (2024). Nonparametric classification with missing data. *Ann. Statist.* **52** 1178–1200. MR4784074 https://doi.org/10.1214/24-aos2389

SPOHN, M.-L., NÄF, J., MICHEL, L. and MEINSHAUSEN, N. (2025). PKLM: A flexible MCAR test using classification. *Psychometrika* **90** 280–303. MR4890638 https://doi.org/10.1017/psy.2024.14

STEKHOVEN, D. J. and BÜHLMANN, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28** 112–118. https://doi.org/10.1093/bioinformatics/btr597

THÉPAUT, S. and VERZELEN, N. (2024). Optimal estimation of Schatten norms of a rectangular matrix. *Ann. Statist.* **52** 1334–1359. MR4804811 https://doi.org/10.1214/24-aos2374

TIERNEY, N. and COOK, D. (2023). Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *J. Stat. Softw.* **105** 1–31. https://doi.org/10.18637/jss.v105.i07

VAN BUUREN, S. and GROOTHUIS-OUDSHOORN, K. (2011). mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45** 1–67. https://doi.org/10.18637/jss.v045.i03

VANDENBERGHE, L. and BOYD, S. (1996). Semidefinite programming. *SIAM Rev.* **38** 49–95. MR1379041 https://doi.org/10.1137/1038003

WAGHMARE, K. G. and PANARETOS, V. M. (2022). The completion of covariance kernels. *Ann. Statist.* **50** 3281–3306. MR4524497 https://doi.org/10.1214/22-aos2228

WILKS, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist.* **17** 257–281. MR0017498 https://doi.org/10.1214/aoms/1177730940

YANAGIDA, T. (2024). misty: Miscellaneous functions. R package version 0.6.7.

YATES, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agric.* **1** 129–142.

ZHU, Z., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional principal component analysis with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2000–2031. MR4515564 https://doi.org/10.1111/rssb.12550

The Annals of Statistics
2025, Vol. 53, No. 5, 2230–2257
https://doi.org/10.1214/25-AOS2541

CAUSALITY PURSUIT FROM HETEROGENEOUS ENVIRONMENTS VIA NEURAL ADVERSARIAL INVARIANCE LEARNING

BY YIHONG GU[1,a], CONG FANG[2,c], PETER BÜHLMANN[3,d] AND JIANQING FAN[1,b]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*yihongg@princeton.edu,* [b]*jqfan@princeton.edu*

[2]*School of Intelligence Science and Technology, Peking University,* [c]*congfang@pku.edu.cn*

[3]*Seminar for Statistics, ETH Zürich,* [d]*buhlmann@stat.math.ethz.ch*

Pursuing causality from data is a fundamental problem in scientific discovery, treatment intervention, and transfer learning. This paper introduces a novel algorithmic method for addressing nonparametric invariance and causality learning in regression models across multiple environments, where the joint distribution of response variables and covariates varies, but the conditional expectations of outcome given an unknown set of quasi-causal variables are invariant. The challenge of finding such an unknown set of quasi-causal or invariant variables is compounded by the presence of endogenous variables that have heterogeneous effects across different environments. The proposed Focused Adversarial Invariant Regularization (FAIR) framework utilizes an innovative minimax optimization approach that drives regression models toward prediction-invariant solutions through adversarial testing. Leveraging the representation power of neural networks, FAIR neural networks (FAIR-NN) are introduced for causality pursuit. It is shown that FAIR-NN can find the invariant variables and quasi-causal variables under a minimal identification condition and that the resulting procedure is adaptive to low-dimensional composition structures in a nonasymptotic analysis. Under a structural causal model, variables identified by FAIR-NN represent pragmatic causality and provably align with exact causal mechanisms under conditions of sufficient heterogeneity. Computationally, FAIR-NN employs a novel Gumbel approximation with decreased temperature and a stochastic gradient descent ascent algorithm. The procedures are demonstrated using simulated and real-data examples.

REFERENCES

AGARWAL, A. and ZHANG, T. (2022). Minimax regret optimization for robust machine learning under distribution shift. In *Conference on Learning Theory* 2704–2729. PMLR.

ARJOVSKY, M., BOTTOU, L., GULRAJANI, I. and LOPEZ-PAZ, D. (2019). Invariant risk minimization. arXiv preprint. Available at arXiv:1907.02893.

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

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

BREIMAN, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statist. Sci.* **16** 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

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

CHERNOZHUKOV, V., NEWEY, W., SINGH, R. and SYRGKANIS, V. (2020). Adversarial estimation of Riesz representers. arXiv preprint. Available at arXiv:2101.00009.

CHICKERING, D. M. (2002). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554.

MSC2020 subject classifications. Primary 62G05; secondary 62D20.

Key words and phrases. Adversarial estimation, causal discovery, conditional moment restriction, Gumbel approximation, invariance, neural networks.

DIKKALA, N., LEWIS, G., MACKEY, L. and SYRGKANIS, V. (2020). Minimax estimation of conditional moment models. *Adv. Neural Inf. Process. Syst.* **33** 12248–12262.

DUCHI, J. C. and NAMKOONG, H. (2021). Learning models with uniform performance via distributionally robust optimization. *Ann. Statist.* **49** 1378–1406. MR4298868 https://doi.org/10.1214/20-aos2004

FAN, J., FANG, C., GU, Y. and ZHANG, T. (2024). Environment invariant linear least squares. *Ann. Statist.* **52** 2268–2292. MR4829488 https://doi.org/10.1214/24-aos2435

FAN, J. and GU, Y. (2024). Factor augmented sparse throughput deep ReLU neural networks for high dimensional regression. *J. Amer. Statist. Assoc.* **119** 2680–2694. MR4833907 https://doi.org/10.1080/01621459.2023.2271605

FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press/CRC, Boca Raton.

FAN, J. and LIAO, Y. (2014). Endogeneity in high dimensions. *Ann. Statist.* **42** 872–917. MR3210990 https://doi.org/10.1214/13-AOS1202

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

GAMELLA, J. L., PETERS, J. and BÜHLMANN, P. (2025). The causal chambers: Real physical systems as a testbed for AI methodology. *Nat. Mach. Intell.* **7** 107–118.

GAUSS, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Cambridge University Press.

GEIGER, D. and PEARL, J. (1990). On the logic of causal models. In *Uncertainty in Artificial Intelligence*, 4. *Mach. Intelligence Pattern Recogn.* **9** 3–14. North-Holland, Amsterdam. MR1166825 https://doi.org/10.1016/B978-0-444-88650-7.50006-8

GHASSAMI, A., SALEHKALEYBAR, S., KIYAVASH, N. and ZHANG, K. (2017). Learning causal structures using regression invariance. *Adv. Neural Inf. Process. Syst.* **30**.

GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**.

GU, Y., FANG, C., BÜHLMANN, P. and FAN, J. (2025). Supplement to "Causality Pursuit from Heterogeneous Environments via Neural Adversarial Invariance Learning." https://doi.org/10.1214/25-AOS2541SUPPA, https://doi.org/10.1214/25-AOS2541SUPPB

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

HEINZE-DEML, C., PETERS, J. and MEINSHAUSEN, N. (2018). Invariant causal prediction for nonlinear models. *J. Causal Inference* **6** Art. No. 20170016. MR4335430 https://doi.org/10.1515/jci-2017-0016

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

HOYER, P., JANZING, D., MOOIJ, J. M., PETERS, J. and SCHÖLKOPF, B. (2008). Nonlinear causal discovery with additive noise models. *Adv. Neural Inf. Process. Syst.* **21**.

HYTTINEN, A., EBERHARDT, F. and JÄRVISALO, M. (2014). Constraint-based causal discovery: Conflict resolution with answer set programming. In *Conference on Uncertainty in Artificial Intelligence* 340–349. AUAI Press.

HYTTINEN, A., HOYER, P. O., EBERHARDT, F. and JÄRVISALO, M. (2013). Discovering cyclic causal models with latent variables: A general SAT-based procedure. In *Uncertainty in Artificial Intelligence* 301. Citeseer.

JANG, E., GU, S. and POOLE, B. (2017). Categorical reparameterization with Gumbel-softmax. In *International Conference on Learning Representations*.

JANZING, D., CHAVES, R. and SCHÖLKOPF, B. (2016). Algorithmic independence of initial condition and dynamical law in thermodynamics and causal inference. *New J. Phys.* **18** 093052.

JANZING, D., MOOIJ, J., ZHANG, K., LEMEIRE, J., ZSCHEISCHLER, J., DANIUŠIS, P., STEUDEL, B. and SCHÖLKOPF, B. (2012). Information-geometric approach to inferring causal directions. *Artificial Intelligence* **182/183** 1–31. MR2909089 https://doi.org/10.1016/j.artint.2012.01.002

KAMATH, P., TANGELLA, A., SUTHERLAND, D. and SREBRO, N. (2021). Does invariant risk minimization capture invariance? In *International Conference on Artificial Intelligence and Statistics* 4069–4077. PMLR.

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

LEGENDRE, A.-M. (1805). *Nouvelles Méthodes Pour la Détermination des Orbites des Comètes* [*New Methods for the Determination of the Orbits of Comets*] (*in French*). F. Didot, Paris.

LIANG, T. (2021). How well generative adversarial networks learn distributions. *J. Mach. Learn. Res.* **22** Paper No. 228. MR4329807

MADDISON, C. J., MNIH, A. and TEH, Y. W. (2017). The concrete distribution: A continuous relaxation of discrete random variables. In *International Conference on Learning Representations*.

MEINSHAUSEN, N. and BÜHLMANN, P. (2015). Maximin effects in inhomogeneous large-scale data. *Ann. Statist.* **43** 1801–1830. MR3357879 https://doi.org/10.1214/15-AOS1325

NELDER, J. A. and WEDDERBURN, R. W. (1972). Generalized linear models. *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* **135** 370–384.

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

PEARL, J., GLYMOUR, M. and JEWELL, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley, Chichester. MR3497861

PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012. MR3557186 https://doi.org/10.1111/rssb.12167

PETERS, J., MOOIJ, J. M., JANZING, D. and SCHÖLKOPF, B. (2014). Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.* **15** 2009–2053. MR3231600

PFISTER, N., BÜHLMANN, P. and PETERS, J. (2019). Invariant causal prediction for sequential data. *J. Amer. Statist. Assoc.* **114** 1264–1276. MR4011778 https://doi.org/10.1080/01621459.2018.1491403

PFISTER, N., WILLIAMS, E. G., PETERS, J., AEBERSOLD, R. and BÜHLMANN, P. (2021). Stabilizing variable selection and regression. *Ann. Appl. Stat.* **15** 1220–1246. MR4317406 https://doi.org/10.1214/21-aoas1487

RICHARDSON, T. (1996). Feedback models: Interpretation and discovery. Ph.D. thesis, Carnegie Mellon.

ROSENFELD, E., RAVIKUMAR, P. and RISTESKI, A. (2021). The risks of invariant risk minimization. *Int. Conf. Learn. Represent.*.

ROTHENHÄUSLER, D., BÜHLMANN, P. and MEINSHAUSEN, N. (2019). Causal Dantzig: Fast inference in linear structural equation models with hidden variables under additive interventions. *Ann. Statist.* **47** 1688–1722. MR3911127 https://doi.org/10.1214/18-AOS1732

ROTHENHÄUSLER, D., MEINSHAUSEN, N., BÜHLMANN, P. and PETERS, J. (2021). Anchor regression: Heterogeneous data meet causality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 215–246. MR4250274 https://doi.org/10.1111/rssb.12398

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

SAGAWA, S., KOH, P. W., HASHIMOTO, T. B. and LIANG, P. (2020). Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *Int. Conf. Learn. Represent.*.

SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function (with discussion). *Ann. Statist.* **48** 1875–1921. MR4134774 https://doi.org/10.1214/19-AOS1875

SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. MR2274431

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

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

YIN, M., WANG, Y. and BLEI, D. M. (2024). Optimization-based causal estimation from heterogeneous environments. *J. Mach. Learn. Res.* **25** Paper No. 168. MR4777410

ZHANG, K. and HYVÄRINEN, A. (2009). On the identifiability of the post-nonlinear causal model. In 25*th Conference on Uncertainty in Artificial Intelligence* (*UAI* 2009) 647–655. AUAI Press.

The Annals of Statistics
2025, Vol. 53, No. 5, 2258–2282
https://doi.org/10.1214/25-AOS2542

SPECTRAL DENSITY ESTIMATION OF FUNCTION-VALUED SPATIAL PROCESSES

BY RAFAIL KARTSIOUKAS[1,a], STILIAN STOEV[2,b] AND TAILEN HSING[2,c]

[1]*Medpace, Inc.*, [a]*rkarts@umich.edu*

[2]*Department of Statistics, University of Michigan, Ann Arbor*, [b]*sstoev@umich.edu*, [c]*thsing@umich.edu*

The spectral density function describes the second-order properties of a stationary stochastic process on $\mathbb{R}^d$. This paper considers the nonparametric estimation of the spectral density of a continuous-time stochastic process taking values in a separable Hilbert space. Our estimator is based on kernel smoothing and can be applied to a wide variety of spatial sampling schemes including those in which data are observed at irregular spatial locations. Thus, it finds immediate applications in spatial statistics, where irregularly sampled data naturally arise. The rates for the bias and variance of the estimator are obtained under general conditions in a mixed-domain asymptotic setting. When the data are observed on a regular grid, the optimal rate of the estimator matches the minimax rate for the class of operator-covariance functions that decay according to a power law. The asymptotic normality of the spectral density estimator is also established under general conditions for Gaussian Hilbert-space valued processes. Finally, with a view towards practical applications the asymptotic results are specialized to the case of discretely-sampled functional data in a reproducing kernel Hilbert space.

REFERENCES

BANDYOPADHYAY, S. and LAHIRI, S. N. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā* **71** 221–259. MR2639292

BARDET, J.-M. and BERTRAND, P. R. (2010). A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand. J. Stat.* **37** 458–476. MR2724508 https://doi.org/10.1111/j.1467-9469.2009.00684.x

BENTKUS, R. YU. (1985). The rate of uniform convergence of statistical estimates of a spectral density in spaces of differentiable functions. *Liet. Mat. Rink.* **25** 17–31. MR0823640

BILLINGSLEY, P. (2012). *Probability and Measure. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2893652

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR1015093

BOCHNER, S. (1948). *Vorlesungen über Fouriersche Integrale*. Chelsea, New York.

BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554 https://doi.org/10.1137/1.9780898719246

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series: Theory and Methods. Springer Series in Statistics*. Springer, New York. MR2839251

CEROVECKI, C. and HÖRMANN, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *J. Multivariate Anal.* **154** 282–295. MR3588570 https://doi.org/10.1016/j.jmva.2016.11.006

DÜKER, M.-C. and ZOUBOULOGLOU, P. (2024). Breuer–Major Theorems for Hilbert space-valued random variables. Preprint. Available at arXiv:2405.11452.

EFROMOVICH, S. (1998). Data-driven efficient estimation of the spectral density. *J. Amer. Statist. Assoc.* **93** 762–769. MR1631386 https://doi.org/10.2307/2670126

FAZEKAS, I. and KUKUSH, A. G. (2000). Infill asymptotics inside increasing domains for the least squares estimator in linear models. *Stat. Inference Stoch. Process.* **3** 199–223. MR1819396 https://doi.org/10.1023/A:1009914117739

FELLER, W. (2008). *An Introduction to Probability Theory and Its Applications, Vol.* 2. Wiley, New York.

MSC2020 subject classifications. Primary 62M10; secondary 62M15, 62R10, 60G10.
Key words and phrases. Functional data, spatial process, spectral density, minimax rate, central limit theorem.

FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545 https://doi.org/10.1198/016214506000000852

GINOVYAN, M. S. (2011). Efficient estimation of spectral functionals for continuous-time stationary models. *Acta Appl. Math.* **115** 233–254. MR2818916 https://doi.org/10.1007/s10440-011-9617-7

GUYON, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69** 95–105. MR0655674 https://doi.org/10.1093/biomet/69.1.95

HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields* **99** 399–424. MR1283119 https://doi.org/10.1007/BF01199899

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. MR0279952

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

KARTSIOUKAS, R., STOEV, S. and HSING, T. (2025). Supplement to "Spectral density estimation of function-valued spatial processes." https://doi.org/10.1214/25-AOS2542SUPP

KHINTCHINE, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **109** 604–615. MR1512911 https://doi.org/10.1007/BF01449156

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

KUUSELA, M. and STEIN, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. R. Soc. A, Math. Phys. Eng. Sci.* **474**.

MAITRA, T. and BHATTACHARYA, S. (2025). Increasing domain infill asymptotics for stochastic differential equations driven by fractional Brownian motion. *Statistics* **59** 989–1017. MR4929719 https://doi.org/10.1080/02331888.2025.2477199

MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. MR2655530 https://doi.org/10.1111/j.1467-9868.2008.00685.x

NEEB, K.-H. (1998). Operator-valued positive definite kernels on tubes. *Monatsh. Math.* **126** 125–160. MR1639387 https://doi.org/10.1007/BF01473583

NUALART, D. and PECCATI, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. MR2118863 https://doi.org/10.1214/009117904000000621

PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114 https://doi.org/10.1214/13-AOS1086

PERCIVAL, D. B. and WALDEN, A. T. (2006). *Wavelet Methods for Time Series Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **4**. Cambridge Univ. Press, Cambridge. MR2218866

PERCIVAL, D. B. and WALDEN, A. T. (2020). *Spectral Analysis for Univariate Time Series*. *Cambridge Series in Statistical and Probabilistic Mathematics* **51**. Cambridge Univ. Press, Cambridge. MR4406756 https://doi.org/10.1017/9781139235723

ROBINSON, P. M. (1983). Review of various approaches to power spectrum estimation. In *Time Series in the Frequency Domain*. *Handbook of Statist.* **3** 343–368. North-Holland, Amsterdam. MR0749793 https://doi.org/10.1016/S0169-7161(83)03018-7

ROEMMICH, D., GOULD, W. J. and GILSON, J. (2012). 135 years of global ocean warming between the challenger expedition and the argo programme. *Nat. Clim. Change* **2** 425–428. https://doi.org/10.1038/nclimate1461

SAMAROV, A. (1977). Lower bound on the risk for spectral density estimates. *Problemy Peredachi Informatsii* **13** 67–72.

SCHUSTER, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Magn.* **3** 13–41.

SHEN, J., STOEV, S. and HSING, T. (2022a). Tangent fields, intrinsic stationarity, and self similarity (with a supplement on Matheron Theory). Available at arXiv:2010.14715.

SHEN, J., STOEV, S. and HSING, T. (2022b). Tangent fields, intrinsic stationarity, and self similarity. *Electron. J. Probab.* **27** Paper No. 34, 56. MR4387842 https://doi.org/10.1214/22-ejp754

SIMON, B. (2015). *Advanced Complex Analysis*. *A Comprehensive Course in Analysis, Part 2B*. Amer. Math. Soc., Providence, RI. MR3364090 https://doi.org/10.1090/simon/002.2

SUBBA RAO, S. (2018). Statistical inference for spatial statistics defined in the Fourier domain. *Ann. Statist.* **46** 469–499. MR3782374 https://doi.org/10.1214/17-AOS1556

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

VAN DELFT, A. and DETTE, H. (2024). A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes. *Ann. Statist.* **52** 550–579. MR4744187 https://doi.org/10.1214/24-aos2358

VAN DELFT, A. and EICHLER, M. (2018). Locally stationary functional time series. *Electron. J. Stat.* **12** 107–170. MR3746979 https://doi.org/10.1214/17-EJS1384

VORONOI, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *J. Reine Angew. Math.* **133** 97–102. MR1580737 https://doi.org/10.1515/crll.1908.133.97

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

WERON, A. (1975). Prediction theory in Banach spaces. In *Probability—Winter School* (*Proc. Fourth Winter School*, *Karpacz*, 1975). *Lecture Notes in Math.* **472** 207–228. Springer, Berlin. MR0394849

YAN, D.-M., WANG, W., LÉVY, B. and LIU, Y. (2013). Efficient computation of clipped Voronoi diagram for mesh generation. *Comput.-Aided Design* **45** 843–852. MR3043264 https://doi.org/10.1016/j.cad.2011.09.004

YARGER, D., STOEV, S. and HSING, T. (2022). A functional-data approach to the Argo data. *Ann. Appl. Stat.* **16** 216–246. MR4400508 https://doi.org/10.1214/21-aoas1477

ZHANG, S. (2024). Statistical analysis of irregularly spaced spatial data in frequency domain. *J. Time Series Anal.* **45** 714–738. MR4794725 https://doi.org/10.1111/jtsa.12735

ZHU, T. and POLITIS, D. N. (2020). Higher-order accurate spectral density estimation of functional time series. *J. Time Series Anal.* **41** 3–20. MR4048679 https://doi.org/10.1111/jtsa.12473

The Annals of Statistics
2025, Vol. 53, No. 5, 2283–2302
https://doi.org/10.1214/25-AOS2543

IMPROVING KNOCKOFFS WITH CONDITIONAL CALIBRATION

BY YIXIANG LUO[1,a], WILLIAM FITHIAN[2,b] AND LIHUA LEI[3,c]

[1]*Department of Mathematics, University of California, Berkeley,* [a]*yixiangluo@berkeley.edu*
[2]*Department of Statistics, University of California, Berkeley,* [b]*wfithian@berkeley.edu*
[3]*Graduate School of Business, Stanford University,* [c]*lihualei@stanford.edu*

The knockoff filter of (*Ann. Statist.* **43** (2015) 2055–2085) is a flexible framework for multiple testing in supervised learning models, based on introducing synthetic predictor variables to control the false discovery rate (FDR). Using the conditional calibration framework of (*Ann. Statist.* **50** (2022) 3091–3118), we introduce the *calibrated knockoff procedure*, a method that uniformly improves the power of any fixed-X or model-X knockoff procedure. We show theoretically and empirically that the improvement is especially notable in two contexts where knockoff methods can be nearly powerless: when the rejection set is small, and when the structure of the design matrix in fixed-X knockoffs prevents us from constructing good knockoff variables. In these contexts, calibrated knockoffs even outperform competing FDR-controlling methods like the (dependence-adjusted) Benjamini–Hochberg procedure in many scenarios.

REFERENCES

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

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, Methodol.* **57** 289–300. MR1325392

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

DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

EMERY, K. and KEICH, U. (2019). Controlling the FDR in variable selection via multiple knockoffs. Preprint. Available at arXiv:1911.09442.

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

GIMENEZ, J. R. and ZOU, J. (2019). Improving the stability of the knockoff procedure: Multiple simultaneous knockoffs and entropy maximization. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2184–2192. PMLR.

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

KATSEVICH, E., SABATTI, C. and BOGOMOLOV, M. (2023). Filtering the rejection set while preserving false discovery rate control. *J. Amer. Statist. Assoc.* **118** 165–176. MR4571114 https://doi.org/10.1080/01621459.2021.1920958

LI, X. and FITHIAN, W. (2021). Whiteout: When do fixed-X knockoffs fail? Preprint. Available at arXiv:2107.06388.

LUO, Y., FITHIAN, W. and LEI, L. (2025a). Supplement to "Improving knockoffs with conditional calibration." https://doi.org/10.1214/25-AOS2543SUPPA

LUO, Y., FITHIAN, W. and LEI, L. (2025b). R package for "Improving knockoffs with conditional calibration." https://doi.org/10.1214/25-AOS2543SUPPB

LUO, Y., FITHIAN, W. and LEI, L. (2025c). Numerical experiment code for "Improving knockoffs with conditional calibration." https://doi.org/10.1214/25-AOS2543SUPPC

MSC2020 subject classifications. Primary 62H15; secondary 62J15.
Key words and phrases. Multiple hypotheses testing, linear model, knockoffs.

NGUYEN, T.-B., CHEVALIER, J.-A., THIRION, B. and ARLOT, S. (2020). Aggregation of multiple knockoffs. In *International Conference on Machine Learning* 7283–7293. PMLR.

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

RHEE, S.-Y., TAYLOR, J., WADHERA, G., BEN-HUR, A., BRUTLAG, D. L. and SHAFER, R. W. (2006). Genotypic predictors of human immunodeficiency virus type 1 drug resistance. *Proc. Natl. Acad. Sci. USA* **103** 17355–17360.

SARKAR, S. K. and TANG, C. Y. (2022). Adjusting the Benjamini–Hochberg method for controlling the false discovery rate in knockoff-assisted variable selection. *Biometrika* **109** 1149–1155. MR4519121 https://doi.org/10.1093/biomet/asab066

SHAH, R. D. and BÜHLMANN, P. (2023). Double-estimation-friendly inference for high-dimensional misspecified models. *Statist. Sci.* **38** 68–91. MR4535395 https://doi.org/10.1214/22-sts850

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SPECTOR, A. and JANSON, L. (2022). Powerful knockoffs via minimizing reconstructability. *Ann. Statist.* **50** 252–276. MR4382016 https://doi.org/10.1214/21-aos2104

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

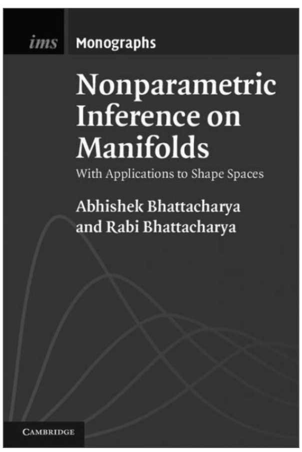
VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds
With Applications to Shape Spaces

Abhishek Bhattacharya, Rabi Bhattacharya

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

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

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

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