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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 50, No. 6—December 2022

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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

THE ANNALS OF STATISTICS Vol. 50, No. 6, pp. 3091–3676 December 2022

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 50, Number 6, December 2022. 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
2022, Vol. 50, No. 6, 3091–3118
https://doi.org/10.1214/21-AOS2137

CONDITIONAL CALIBRATION FOR FALSE DISCOVERY RATE CONTROL UNDER DEPENDENCE

BY WILLIAM FITHIAN[1,a] AND LIHUA LEI[2,b]

[1]*Department of Statistics, University of California, Berkeley,* [a]*wfithian@berkeley.edu*
[2]*Department of Statistics, Stanford University,* [b]*lihualei@stanford.edu*

We introduce a new class of methods for finite-sample false discovery rate (FDR) control in multiple testing problems with dependent test statistics where the dependence is known. Our approach separately calibrates a data-dependent p-value rejection threshold for each hypothesis, relaxing or tightening the threshold as appropriate to target exact FDR control. In addition to our general framework, we propose a concrete algorithm, the dependence-adjusted Benjamini–Hochberg (dBH) procedure, which thresholds the BH-adjusted p-value for each hypothesis. Under positive regression dependence, the dBH procedure uniformly dominates the standard BH procedure, and in general it uniformly dominates the Benjamini–Yekutieli (BY) procedure (also known as BH with log correction), which makes a conservative adjustment for worst-case dependence. Simulations and real data examples show substantial power gains over the BY procedure, and competitive performance with knockoffs in settings where both methods are applicable. When the BH procedure empirically controls FDR (as it typically does in practice), the dBH procedure performs comparably.

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

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

BENJAMINI, Y. and BOGOMOLOV, M. (2014). Selective inference on multiple families of hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 297–318. MR3153943 https://doi.org/10.1111/rssb.12028

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

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

BENJAMINI, Y. and HOCHBERG, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.* **24** 407–418. MR1481424 https://doi.org/10.1111/1467-9469.00072

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

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

BLANCHARD, G. and ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.* **2** 963–992. MR2448601 https://doi.org/10.1214/08-EJS180

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

BOCA, S. M. and LEEK, J. T. (2017). A regression framework for the proportion of true null hypotheses. Preprint. BioRxiv 35675.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

MSC2020 subject classifications. Primary 62F03; secondary 62G10.
Key words and phrases. False discovery rate, multiple testing, dependent test statistics, Benjamini–Hochberg procedure.

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

DELATTRE, S. and ROQUAIN, E. (2015). New procedures controlling the false discovery proportion via Romano-Wolf's heuristic. *Ann. Statist.* **43** 1141–1177. MR3346700 https://doi.org/10.1214/14-AOS1302

DOBRIBAN, E., FORTNEY, K., KIM, S. K. and OWEN, A. B. (2015). Optimal multiple testing under a Gaussian prior on the effect sizes. *Biometrika* **102** 753–766. MR3431551 https://doi.org/10.1093/biomet/asv050

FAN, J. and HAN, X. (2017). Estimation of the false discovery proportion with unknown dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1143–1164. MR3689312 https://doi.org/10.1111/rssb.12204

FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887 https://doi.org/10.1080/01621459.2012.720478

FARCOMENI, A. (2006). More powerful control of the false discovery rate under dependence. *Stat. Methods Appl.* **15** 43–73. MR2281214 https://doi.org/10.1007/s10260-006-0002-z

FARCOMENI, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Stat.* **34** 275–297. MR2346640 https://doi.org/10.1111/j.1467-9469.2006.00530.x

FERREIRA, J. A. and ZWINDERMAN, A. H. (2006). On the Benjamini–Hochberg method. *Ann. Statist.* **34** 1827–1849. MR2283719 https://doi.org/10.1214/009053606000000425

FINNER, H. (1999). Stepwise multiple test procedures and control of directional errors. *Ann. Statist.* **27** 274–289. MR1701111 https://doi.org/10.1214/aos/1018031111

FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. ArXiv preprint. Available at arXiv:1410.2597.

FITHIAN, W. and LEI, L. (2022). Supplement to "Conditional calibration for false discovery rate control under dependence." https://doi.org/10.1214/21-AOS2137SUPP

GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika* **93** 509–524. MR2261439 https://doi.org/10.1093/biomet/93.3.509

GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517. MR1924303 https://doi.org/10.1111/1467-9868.00347

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

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

GUO, W., HE, L. and SARKAR, S. K. (2014). Further results on controlling the false discovery proportion. *Ann. Statist.* **42** 1070–1101. MR3210996 https://doi.org/10.1214/14-AOS1214

GUO, W. and RAO, M. B. (2008). On control of the false discovery rate under no assumption of dependency. *J. Statist. Plann. Inference* **138** 3176–3188. MR2526229 https://doi.org/10.1016/j.jspi.2008.01.003

HELLER, R. and ROSSET, S. (2021). Optimal control of false discovery criteria in the two-group model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 133–155. MR4220987 https://doi.org/10.1111/rssb.12403

IGNATIADIS, N. and HUBER, W. (2017). Covariate-powered weighted multiple testing with false discovery rate control. ArXiv preprint. Available at arXiv:1701.05179.

IGNATIADIS, N., KLAUS, B., ZAUGG, J. B. and HUBER, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13** 577–580.

KIM, K. I. and VAN DE WIEL, M. A. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinform.* **9** 114. https://doi.org/10.1186/1471-2105-9-114

KORN, E. L., TROENDLE, J. F., MCSHANE, L. M. and SIMON, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371 https://doi.org/10.1016/S0378-3758(03)00211-8

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Stat.* **37** 1137–1153. MR0202228 https://doi.org/10.1214/aoms/1177699260

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

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

LEHMANN, E. L. and SCHEFFÉ, H. (1955). Completeness, similar regions, and unbiased estimation. II. *Sankhyā* **15** 219–236. MR0072410 https://doi.org/10.1007/978-1-4614-1412-4_24

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

LI, A. and BARBER, R. F. (2019). Multiple testing with the structure-adaptive Benjamini–Hochberg algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 45–74. MR3904779

LYNCH, G. and GUO, W. (2016). On procedures controlling the FDR for testing hierarchically ordered hypotheses. ArXiv preprint. Available at arXiv:1612.04467.

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

PATTERSON, E. and SESIA, M. (2018). knockoff: The Knockoff Filter for Controlled Variable Selection. R package version 0.3.2.

PERONE PACIFICO, M., GENOVESE, C., VERDINELLI, I. and WASSERMAN, L. (2004). False discovery control for random fields. *J. Amer. Statist. Assoc.* **99** 1002–1014. MR2109490 https://doi.org/10.1198/016214500000l655

RHEE, S.-Y., FESSEL, W. J., ZOLOPA, A. R., HURLEY, L., LIU, T., TAYLOR, J., NGUYEN, D. P., SLOME, S., KLEIN, D. et al. (2005). HIV-1 protease and reverse-transcriptase mutations: Correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance. *J. Infect. Dis.* **192** 456–465.

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.

ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST* **17** 417–442. MR2470085 https://doi.org/10.1007/s11749-008-0126-6

ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3** 678–711. MR2521216 https://doi.org/10.1214/09-EJS430

ROQUAIN, E. and VILLERS, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612. MR2797857 https://doi.org/10.1214/10-AOS847

SARKAR, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30** 239–257. MR1892663 https://doi.org/10.1214/aos/1015362192

SARKAR, S. K. and TANG, C. Y. (2021). Adjusting the Benjamini–Hochberg method for controlling the false discovery rate in knockoff assisted variable selection. ArXiv preprint. Available at arXiv:2102.09080.

SHAFFER, J. P. (1980). Control of directional errors with stagewise multiple test procedures. *Ann. Statist.* **8** 1342–1347. MR0594649

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302 https://doi.org/10.1111/1467-9868.00346

STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398 https://doi.org/10.1214/aos/1074290335

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

SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. MR2411657 https://doi.org/10.1198/016214507000000545

SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. MR2649603 https://doi.org/10.1111/j.1467-9868.2008.00694.x

TANSEY, W., WANG, Y., BLEI, D. and RABADAN, R. (2018). Black box FDR. In *International Conference on Machine Learning* 4867–4876. PMLR.

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

TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689 https://doi.org/10.1080/01621459.2015.1108848

TROENDLE, J. F. (2000). Stepwise normal theory multiple test procedures controlling the false discovery rate. *J. Statist. Plann. Inference* **84** 139–158. MR1747501 https://doi.org/10.1016/S0378-3758(99)00145-7

WEINSTEIN, A., FITHIAN, W. and BENJAMINI, Y. (2013). Selection adjusted confidence intervals with more power to determine the sign. *J. Amer. Statist. Assoc.* **108** 165–176. MR3174610 https://doi.org/10.1080/01621459.2012.737740

XIA, F., ZHANG, M. J., ZOU, J. Y. and TSE, D. (2017). Neuralfdr: Learning discovery thresholds from hypothesis features. In *Advances in Neural Information Processing Systems* 1541–1550.

XIE, J., CAI, T. T., MARIS, J. and LI, H. (2011). Optimal false discovery rate control for dependent data. *Stat. Interface* **4** 417–430. MR2868825 https://doi.org/10.4310/SII.2011.v4.n4.a1

YEKUTIELI, D. and BENJAMINI, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Plann. Inference* **82** 171–196.

The Annals of Statistics
2022, Vol. 50, No. 6, 3119–3143
https://doi.org/10.1214/22-AOS2189

A NO-FREE-LUNCH THEOREM FOR MULTITASK LEARNING

BY STEVE HANNEKE[1,a] AND SAMORY KPOTUFE[2,b]

[1]*Computer Science Department, Purdue University,* [a]*steve.hanneke@gmail.com*
[2]*Department of Statistics, Columbia University,* [b]*samory@columbia.edu*

Multitask learning and related areas such as multisource domain adaptation address modern settings where data sets from N related distributions $\{P_t\}$ are to be combined toward improving performance on any single such distribution $\mathcal{D}$. A perplexing fact remains in the evolving theory on the subject: while we would hope for performance bounds that account for the contribution from multiple tasks, the vast majority of analyses result in bounds that improve at best in the number n of samples per task, but most often do not improve in N. As such, it might seem at first that the distributional settings or aggregation procedures considered in such analyses might be somehow unfavorable; however, as we show, the picture happens to be more nuanced, with interestingly hard regimes that might appear otherwise favorable.

In particular, we consider a seemingly favorable classification scenario where all tasks P_t share a common optimal classifier h^*, and which can be shown to admit a broad range of regimes with improved *oracle rates* in terms of N and n. Some of our main results are:

• We show that, even though such regimes admit minimax rates accounting for both n and N, no adaptive algorithm exists, that is, without access to distributional information, *no* algorithm can guarantee rates that improve with large N for n fixed.

• With a bit of additional information, namely, a ranking of tasks $\{P_t\}$ according to their *distance* to a target $\mathcal{D}$, a simple rank-based procedure can achieve near optimal aggregations of tasks' data sets, despite a search space exponential in N. Interestingly, the optimal aggregation might exclude certain tasks, even though they all share the same h^*.

REFERENCES

ACHILLE, A., PAOLINI, G., MBENG, G. and SOATTO, S. (2021). The information complexity of learning tasks, their structure and their distance. *Inf. Inference* **10** 51–72. MR4228637 https://doi.org/10.1093/imaiai/iaaa033

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

ARJOVSKY, M., BOTTOU, L., GULRAJANI, I. and LOPEZ-PAZ, D. (2019). Invariant risk minimization. Available at arXiv:1907.02893.

ARORA, S., KHANDEPARKAR, H., KHODAK, M., PLEVRAKIS, O. and SAUNSHI, N. (2019). A theoretical analysis of contrastive unsupervised representation learning. Available at arXiv:1902.09229.

BARTLETT, P. L. (1992). Learning with a slowly changing distribution. In *Proceedings of the* 5*th Annual Workshop on Computational Learning Theory*.

BARTLETT, P. L., JORDAN, M. I. and MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156. MR2268032 https://doi.org/10.1198/016214505000000907

BARVE, R. D. and LONG, P. M. (1997). On the complexity of learning from drifting distributions. *Inform. and Comput.* **138** 170–193. MR1479321 https://doi.org/10.1006/inco.1997.2656

BAXTER, J. (1997). A Bayesian/information theoretic model of learning to learn via multiple task sampling. *Mach. Learn.* **28** 7–39. https://doi.org/10.1023/A:1007327622663

BEN-DAVID, S., BENEDEK, G. M. and MANSOUR, Y. (1989). A parametrization scheme for classifying models of learnability. In *Proceedings of the Second Annual Workshop on Computational Learning Theory* (*Santa Cruz*, *CA*, 1989) 285–302. Morgan Kaufmann, San Mateo, CA. MR1076235

MSC2020 subject classifications. Primary 68Q32, 62H30, 68T05; secondary 68T10.
Key words and phrases. Statistical learning theory, multitask learning, transfer learning, classification.

BEN-DAVID, S. and BORBELY, R. S. (2008). A notion of task relatedness yielding provable multiple-task learning guarantees. *Mach. Learn.* **73** 273–287.

BEN-DAVID, S., BLITZER, J., CRAMMER, K. and PEREIRA, F. (2007). Analysis of representations for domain adaptation. In *Advances in Neural Information Processing Systems*.

BEN-DAVID, S., BLITZER, J., CRAMMER, K., KULESZA, A., PEREIRA, F. and VAUGHAN, J. W. (2010a). A theory of learning from different domains. *Mach. Learn.* **79** 151–175. MR3108150 https://doi.org/10.1007/s10994-009-5152-4

BEN-DAVID, S., LU, T., LUU, T. and PÁL, D. (2010b). Impossibility theorems for domain adaptation. In *Proceedings of the* 13*th International Conference on Artificial Intelligence and Statistics*.

BLUM, A., HAGHTALAB, N., PROCACCIA, A. D. and QIAO, M. (2017). Collaborative PAC learning. In *Advances in Neural Information Processing Systems*.

CAI, T. T. and WEI, H. (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *Ann. Statist.* **49** 100–128. MR4206671 https://doi.org/10.1214/20-AOS1949

CARUANA, R. (1997). Multitask learning. *Mach. Learn.* **28** 41–75.

CORTES, C., MOHRI, M., RILEY, M. and ROSTAMIZADEH, A. (2008). Sample selection bias correction theory. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **5254** 38–53. Springer, Berlin. MR2540648 https://doi.org/10.1007/978-3-540-87987-9_8

CRAMMER, K., KEARNS, M. and WORTMAN, J. (2008). Learning from multiple sources. *J. Mach. Learn. Res.* **9** 1757–1774. MR2438823

DU, S. S., HU, W., KAKADE, S. M., LEE, J. D. and LEI, Q. (2020). Few-shot learning via learning the representation, provably. Available at arXiv:2002.09434.

GRETTON, A., SMOLA, A., HUANG, J., SCHMITTFULL, M., BORGWARDT, K. and SCHÖLKOPF, B. (2009). Covariate shift by kernel mean matching. In *Dataset Shift in Machine Learning* 131–160.

HANNEKE, S. and KPOTUFE, S. (2019). On the value of target data in transfer learning. In *Advances in Neural Information Processing Systems*.

HANNEKE, S. and KPOTUFE, S. (2022). Supplement to "A no-free-lunch theorem for multitask learning." https://doi.org/10.1214/22-AOS2189SUPP

HANNEKE, S. and YANG, L. (2019). Statistical learning under nonstationary mixing processes. In *Proceedings of the* 22*nd International Conference on Artificial Intelligence and Statistics*.

JALALI, A., SANGHAVI, S., RUAN, C. and RAVIKUMAR, P. (2010). A dirty model for multi-task learning. In *Advances in Neural Information Processing Systems*.

KLEIN, T. and RIO, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312 https://doi.org/10.1214/009117905000000044

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

KONSTANTINOV, N., FRANTAR, E., ALISTARH, D. and LAMPERT, C. H. (2020). On the Sample Complexity of Adversarial Multi-Source PAC Learning. Available at arXiv:2002.10384.

KPOTUFE, S. and MARTINET, G. (2021). Marginal singularity and the benefits of labels in covariate-shift. *Ann. Statist.* **49** 3299–3323. MR4352531 https://doi.org/10.1214/21-aos2084

LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865 https://doi.org/10.1214/11-AOS896

MAHLOUJIFAR, S., MAHMOODY, M. and MOHAMMED, A. (2019). Universal multi-party poisoning attacks. In *Proceedings of the* 36*th International Conference on Machine Learning*.

MANSOUR, Y., MOHRI, M. and ROSTAMIZADEH, A. (2009). Multiple source adaptation and the Rényi divergence. In *Proceedings of the* 25*th Conference on Uncertainty in Artificial Intelligence*.

MASSART, P. and NÉDÉLEC, É. (2006). Risk bounds for statistical learning. *Ann. Statist.* **34** 2326–2366. MR2291502 https://doi.org/10.1214/009053606000000786

MAURER, A., PONTIL, M. and ROMERA-PAREDES, B. (2013). Sparse coding for multitask and transfer learning. In *International Conference on Machine Learning*.

MAURER, A., PONTIL, M. and ROMERA-PAREDES, B. (2016). The benefit of multitask representation learning. *J. Mach. Learn. Res.* **17** Paper No. 81, 32. MR3517104

MCNAMARA, D. and BALCAN, M.-F. (2017). Risk bounds for transferring representations with and without fine-tuning. In *International Conference on Machine Learning*.

MOHRI, M. and MUÑOZ MEDINA, A. (2012). New analysis and algorithm for learning with drifting distributions. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **7568** 124–138. Springer, Heidelberg. MR3042886 https://doi.org/10.1007/978-3-642-34106-9_13

MUANDET, K., BALDUZZI, D. and SCHÖLKOPF, B. (2013). Domain generalization via invariant feature representation. In *International Conference on Machine Learning*.

NEGAHBAN, S. N. and WAINWRIGHT, M. J. (2011). Simultaneous support recovery in high dimensions: Benefits and perils of block ℓ_1/ℓ_∞-regularization. *IEEE Trans. Inf. Theory* **57** 3841–3863. MR2817058 https://doi.org/10.1109/TIT.2011.2144150

PENTINA, A. and LAMPERT, C. (2014). A PAC-Bayesian bound for lifelong learning. In *International Conference on Machine Learning*.

QIAO, M. (2018). Do Outliers Ruin Collaboration? Available at arXiv:1805.04720.

REDKO, I., HABRARD, A. and SEBBAN, M. (2017). Theoretical analysis of domain adaptation with optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2021). Adaptive transfer learning. *Ann. Statist.* **49** 3618–3649. MR4352543 https://doi.org/10.1214/21-aos2102

SCOTT, C. and ZHANG, J. (2019). Learning from Multiple Corrupted Sources, with Application to Learning from Label Proportions. Available at arXiv:1910.04665.

SHEN, J., QU, Y., ZHANG, W. and YU, Y. (2018). Wasserstein distance guided representation learning for domain adaptation. In 32*nd AAAI Conference on Artificial Intelligence*.

SLUD, E. V. (1977). Distribution inequalities for the binomial law. *Ann. Probab.* **5** 404–412. MR0438420 https://doi.org/10.1214/aop/1176995801

TRIPURANENI, N., JORDAN, M. I. and JIN, C. (2020). On the Theory of Transfer Learning: The Importance of Task Diversity. Available at arXiv:2006.11650.

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

VAN ERVEN, T., GRÜNWALD, P. D., MEHTA, N. A., REID, M. D. and WILLIAMSON, R. C. (2015). Fast rates in statistical and online learning. *J. Mach. Learn. Res.* **16** 1793–1861. MR3417799

YANG, L., HANNEKE, S. and CARBONELL, J. (2013). A theory of transfer learning with applications to active learning. *Mach. Learn.* **90** 161–189. MR3015741 https://doi.org/10.1007/s10994-012-5310-y

The Annals of Statistics
2022, Vol. 50, No. 6, 3144–3173
https://doi.org/10.1214/22-AOS2217

OPTIMIZATION HIERARCHY FOR FAIR STATISTICAL DECISION PROBLEMS

BY ANIL ASWANI[a] AND MATT OLFAT[b]

Department of Industrial Engineering and Operations Research, University of California, Berkeley, [a]*aaswani@berkeley.edu,* [b]*molfat@berkeley.edu*

Data-driven decision making has drawn scrutiny from policy makers due to fears of potential discrimination, and a growing literature has begun to develop fair statistical techniques. However, these techniques are often specialized to one model context and based on ad hoc arguments, which makes it difficult to perform theoretical analysis. This paper develops an optimization hierarchy, which is a sequence of optimization problems with an increasing number of constraints, for fair statistical decision problems. Because our hierarchy is based on the framework of statistical decision problems, this means it provides a systematic approach for developing and studying fair versions of hypothesis testing, decision making, estimation, regression, and classification. We use the insight that qualitative definitions of fairness are equivalent to statistical independence between the output of a statistical technique and a random variable that measures attributes for which fairness is desired. We use this insight to construct an optimization hierarchy that lends itself to numerical computation, and we use tools from variational analysis and random set theory to prove that higher levels of this hierarchy lead to consistency in the sense that it asymptotically imposes this independence as a constraint in corresponding statistical decision problems. We demonstrate numerical effectiveness of our hierarchy using several data sets, and we use our hierarchy to fairly perform automated dosing of morphine.

REFERENCES

[1] ADJIMAN, C. S., ANDROULAKIS, I. P. and FLOUDAS, C. A. (2000). Global optimization of mixed-integer nonlinear problems. *AIChE J*. **46** 1769–1797.

[2] AGARWAL, A., BEYGELZIMER, A., DUDÍK, M., LANGFORD, J. and WALLACH, H. (2018). A reductions approach to fair classification. In *International Conference on Machine Learning* 60–69.

[3] AGARWAL, A., DUDÍK, M. and WU, Z. S. (2019). Fair regression: Quantitative definitions and reduction-based algorithms. In *International Conference on Machine Learning*.

[4] ANGWIN, J., LARSON, J., MATTU, S. and KIRCHNER, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica* **23**.

[5] ASWANI, A. (2019). Statistics with set-valued functions: Applications to inverse approximate optimization. *Math*. *Program*. **174** 225–251. MR3935080 https://doi.org/10.1007/s10107-018-1257-5

[6] ASWANI, A. and OLFAT, M. (2022). Supplement to "Optimization hierarchy for fair statistical decision problems." https://doi.org/10.1214/22-AOS2217SUPPA, https://doi.org/10.1214/22-AOS2217SUPPB

[7] BAHARLOUEI, S., NOUIEHED, M., BEIRAMI, A. and RAZAVIYAYN, M. (2020). Rényi fair inference. In *International Conference on Learning Representations*.

[8] BANACH, S. (1938). Über homogene polynome in (L^2). *Studia Math*. **7** 36–44.

[9] BAROCAS, S. and SELBST, A. D. (2016). Big data's disparate impact. *Calif*. *Law Rev*. **104** 671–732.

[10] BERG, C. (1987). The multidimensional moment problem and semigroups. In *Moments in Mathematics* (*San Antonio*, *Tex*., 1987). *Proc*. *Sympos*. *Appl*. *Math*. **37** 110–124. Amer. Math. Soc., Providence, RI. MR0921086 https://doi.org/10.1090/psapm/037/921086

[11] BERG, C. (1988). The cube of a normal distribution is indeterminate. *Ann*. *Probab*. **16** 910–913. MR0929086

MSC2020 subject classifications. Primary 62C12, 62F12; secondary 49J53, 60D05.
Key words and phrases. Fairness, independence, optimization, statistical learning.

[12] BERK, R., HEIDARI, H., JABBARI, S., JOSEPH, M., KEARNS, M., MORGENSTERN, J., NEEL, S. and ROTH, A. (2017). A convex framework for fair regression. Preprint. Available at arXiv:1706.02409.

[13] BEUTEL, A., CHEN, J., ZHAO, Z. and CHI, E. H. (2017). Data decisions and theoretical implications when adversarially learning fair representations. Preprint. Available at arXiv:1707.00075.

[14] BILLOCK, J. (2018). Pain bias: The health inequality rarely discussed. *BBC*.

[15] BISGAARD, T. M. and SASVÁRI, Z. (2006). When does $E(X^k \cdot Y^l) = E(X^k) \cdot E(Y^l)$ imply independence? *Statist. Probab. Lett.* **76** 1111–1116. MR2269281 https://doi.org/10.1016/j.spl.2005.12.008

[16] BOCHNAK, J. and SICIAK, J. (1971). Polynomials and multilinear mappings in topological vector spaces. *Studia Math.* **39** 59–76. MR0313810 https://doi.org/10.4064/sm-39-1-59-76

[17] BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–598. MR0803258

[18] BURER, S. (2009). On the copositive representation of binary and continuous nonconvex quadratic programs. *Math. Program.* **120** 479–495. MR2505747 https://doi.org/10.1007/s10107-008-0223-z

[19] BURER, S. and LETCHFORD, A. N. (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surv. Oper. Res. Manag. Sci.* **17** 97–106. MR2997162 https://doi.org/10.1016/j.sorms.2012.08.001

[20] CALDERS, T., KAMIRAN, F. and PECHENIZKIY, M. (2009). Building classifiers with independency constraints. In *IEEE ICDMW* 13–18.

[21] CALDERS, T., KARIM, A., KAMIRAN, F., ALI, W. and ZHANG, X. (2013). Controlling attribute effect in linear regression. In *IEEE ICDM* 71–80.

[22] CALMON, F., WEI, D., VINZAMURI, B., RAMAMURTHY, K. N. and VARSHNEY, K. R. (2017). Optimized pre-processing for discrimination prevention. In *NeuRIPS* 3992–4001.

[23] CHEN, A. and BICKEL, P. J. (2005). Consistent independent component analysis and prewhitening. *IEEE Trans. Signal Process.* **53** 3625–3632. MR2239886 https://doi.org/10.1109/TSP.2005.855098

[24] CHEN, C., ATAMTÜRK, A. and OREN, S. S. (2017). A spatial branch-and-cut method for nonconvex QCQP with bounded complex variables. *Math. Program.* **165** 549–577. MR3707373 https://doi.org/10.1007/s10107-016-1095-2

[25] CHIERICHETTI, F., KUMAR, R., LATTANZI, S. and VASSILVITSKII, S. (2017). Fair clustering through fairlets. In *Advances in Neural Information Processing Systems* 5036–5044.

[26] CHOULDECHOVA, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Preprint. Available at arXiv:1703.00056.

[27] CHZHEN, E., DENIS, C., HEBIRI, M., ONETO, L. and PONTIL, M. (2020). Fair regression with Wasserstein barycenters. *Adv. Neural Inf. Process. Syst.* **33** 7321–7331.

[28] CORTEZ, P., CERDEIRA, A., ALMEIDA, F., MATOS, T. and REIS, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decis. Support Syst.* **47** 547–553.

[29] DEMPE, S. (2002). *Foundations of Bilevel Programming*. *Nonconvex Optimization and Its Applications* **61**. Kluwer Academic, Dordrecht. MR1921449

[30] DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS (1992). Census of population and housing 1990 United States: Summary tape file 1a and 3a (computer files).

[31] DEPARTMENT OF JUSTICE, BUREAU OF JUSTICE STATISTICS (1992). Law enforcement and administrative statistics (computer file).

[32] DEPARTMENT OF JUSTICE, FEDERAL BUREAU OF INVESTIGATION (1995). Crime in the United States (computer file). Available at http://www.fbi.gov/ucr/hc2004/openpage.htm.

[33] DEVROYE, L. and WISE, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** 480–488. MR0579432 https://doi.org/10.1137/0138038

[34] DONINI, M., ONETO, L., BEN-DAVID, S., SHAWE-TAYLOR, J. S. and PONTIL, M. (2018). Empirical risk minimization under fairness constraints. In *NeuRIPS* 2791–2801.

[35] DUSENBURY, M. (2018). "Everbody was telling me there was nothing wrong". *BBC*.

[36] DWORK, C., HARDT, M., PITASSI, T., REINGOLD, O. and ZEMEL, R. (2012). Fairness through awareness. In *Proceedings of the* 3*rd Innovations in Theoretical Computer Science Conference* 214–226. ACM, New York. MR3388391

[37] EDITORIAL (2016). More accountability for big-data algorithms. *Nature* **537** 449. https://doi.org/10.1038/537449a

[38] EDWARDS, H. and STORKEY, A. (2015). Censoring representations with an adversary. Preprint. Available at arXiv:1511.05897.

[39] ENSIGN, D., FRIEDLER, S. A., NEVILLE, S., SCHEIDEGGER, C. and VENKATASUBRAMANIAN, S. (2017). Runaway feedback loops in predictive policing. Preprint. Available at arXiv:1706.09847.

[40] FERNANDEZ-FRAGA, S. M., ACEVES-FERNANDEZ, M. A., PEDRAZA-ORTEGA, J. C. and TOVAR-ARRIAGA, S. (2018). Feature extraction of EEG signal upon BCI systems based on steady-state visual evoked potentials using the ant colony optimization algorithm. *Discrete Dyn. Nat. Soc.* **2018** Art. ID 2143873, 19 pp. MR3825223 https://doi.org/10.1155/2018/2143873

[41] FEUERVERGER, A. and MUREIKA, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5** 88–97. MR0428584

[42] FREY, P. W. and SLATE, D. J. (1991). Letter recognition using Holland-style adaptive classifiers. *Mach. Learn.* **6** 161–182.

[43] GOH, G., COTTER, A., GUPTA, M. and FRIEDLANDER, M. P. (2016). Satisfying real-world goals with dataset constraints. In *Advances in Neural Information Processing Systems* 2415–2423.

[44] GOUIC, T. L., LOUBES, J.-M. and RIGOLLET, P. (2020). Projection to fairness in statistical learning. Preprint. Available at arXiv:2005.11720.

[45] GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

[46] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.* **40** 385–411. MR3014311 https://doi.org/10.1214/11-AOS959

[47] GUROBI OPTIMIZATION, L. (2020). Gurobi optimizer reference manual.

[48] GUVENIR, H. A., ACAR, B., DEMIROZ, G. and CEKIN, A. (1997). A supervised machine learning algorithm for arrhythmia analysis. In *Computers in Cardiology* 1997 433–436.

[49] HARDT, M., PRICE, E. and SREBRO, N. (2016). Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems* 3315–3323.

[50] JOHNSON, K. D., FOSTER, D. P. and STINE, R. A. (2016). Impartial predictive modeling: Ensuring fairness in arbitrary models. Preprint. Available at arXiv:1608.00528.

[51] KAC, M. (1936). Sur les fonctions indépendantes (I) (propriétés générales). *Studia Math.* **6** 46–58.

[52] KAMISHIMA, T., AKAHO, S., ASOH, H. and SAKUMA, J. (2012). Fairness-aware classifier with prejudice remover regularizer. In *ECML-PKDD* 35–50.

[53] KILINÇ-KARZAN, F. and YILDIZ, S. (2015). Two-term disjunctions on the second-order cone. *Math. Program.* **154** 463–491. MR3421943 https://doi.org/10.1007/s10107-015-0903-4

[54] KITAGAWA, T., SAKAGUCHI, S. and TETENOV, A. (2021). Constrained classification and policy learning. Preprint. Available at arXiv:2106.12886.

[55] KLEBANOV, L. and MKRTCHYAN, S. (1984). An estimate of the nearness of the distributions in terms of the nearness of their characteristic functions on a finite interval. *J. Sov. Math.* **25** 1181–1186.

[56] KOROSTELËV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489. MR1332577 https://doi.org/10.1214/aos/1176324531

[57] KRAFT, D. (1988). A software package for sequential quadratic programming. Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt.

[58] LASSERRE, J. B. (2010). *Moments, Positive Polynomials and Their Applications. Imperial College Press Optimization Series* **1**. Imperial College Press, London. MR2589247

[59] LEHMANN, E. L. and ROMANO, J. P. (2006). *Testing Statistical Hypotheses. Springer Texts in Statistics*. Springer, New York. MR2135927

[60] LI, Z., PEREZ-SUAY, A., CAMPS-VALLS, G. and SEJDINOVIC, D. (2019). Kernel dependence regularizers and gaussian processes with applications to algorithmic fairness. Preprint. Available at arXiv:1911.04322.

[61] LICHMAN, M. (2013). UCI machine learning repository.

[62] LIITTSCHWAGER, J. M. and WANG, C. (1978). Integer programming solution of a classification problem. *Manage. Sci.* **24** 1515–1525.

[63] LIN, Y. and SCHRAGE, L. (2009). The global solver in the LINDO API. *Optim. Methods Softw.* **24** 657–668. MR2554904 https://doi.org/10.1080/10556780902753221

[64] LIU, L. T., DEAN, S., ROLF, E., SIMCHOWITZ, M. and HARDT, M. (2018). Delayed impact of fair machine learning. Preprint. Available at arXiv:1803.04383.

[65] MADRAS, D., CREAGER, E., PITASSI, T. and ZEMEL, R. (2018). Learning adversarially fair and transferable representations. Preprint. Available at arXiv:1802.06309.

[66] MAJUMDAR, A., VASUDEVAN, R., TOBENKIN, M. M. and TEDRAKE, R. (2014). Convex optimization of nonlinear feedback controllers via occupation measures. *Int. J. Robot. Res.* **33** 1209–1230.

[67] MANCHIKANTI, L., KAYE, A. M., KNEZEVIC, N. N., MCANALLY, H., SLAVIN, K., TRESCOT, A. M., BLANK, S., PAMPATI, V., ABDI, S. et al. (2017). Responsible, safe, and effective prescription of opioids for chronic non-cancer pain: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Phys.* **20** S3–S92.

[68] MANSOURI, K., RINGSTED, T., BALLABIO, D., TODESCHINI, R. and CONSONNI, V. (2013). Quantitative structure—Activity relationship models for ready biodegradability of chemicals. *J. Chem. Inf. Model.* **53** 867–878.

[69] MATHERON, G. (1975). *Random Sets and Integral Geometry. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0385969

[70] MOLCHANOV, I. (2006). *Theory of Random Sets*. *Probability and Its Applications* (*New York*). Springer London, Ltd., London. MR2132405

[71] MOSEK, APS (2002). The MOSEK Optimization Tools version 3.2 (Revision 8) user's manual and reference.

[72] MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **10** 1–141. https://doi.org/10.1561/2200000060

[73] EXECUTIVE OFFICE OF THE PRESIDENT (2016). *Big Data*: *A Report on Algorithmic Systems*, *Opportunity*, *and Civil Rights*.

[74] OLFAT, M. and ASWANI, A. (2018). Convex formulations for fair principal component analysis. In *AAAI* 663–670.

[75] OLFAT, M. and ASWANI, A. (2018). Spectral algorithms for computing fair support vector machines. In *AISTATS* 1933–1942.

[76] ONETO, L., DONINI, M. and PONTIL, M. (2019). General fair empirical risk minimization. Preprint. Available at arXiv:1901.10080.

[77] OUATTARA, A. and ASWANI, A. (2018). Duality approach to bilevel programs with a convex lower level. In *ACC* 1388–1395.

[78] PÁL, D., PÓCZOS, B. and SZEPESVÁRI, C. (2010). Estimation of Rényi entropy and mutual information based on generalized nearest-neighbor graphs. In *NeuRIPS* 1849–1857.

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

[80] PÉREZ-SUAY, A., LAPARRA, V., MATEO-GARCÍA, G., MUÑOZ-MARÍ, J., GÓMEZ-CHOVA, L. and CAMPS-VALLS, G. (2017). Fair kernel learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 339–355. Springer, Berlin.

[81] RACHEV, S. T., KLEBANOV, L. B., STOYANOV, S. V. and FABOZZI, F. J. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. Springer, New York. MR3024835 https://doi.org/10.1007/978-1-4614-4869-3

[82] REDMOND, M. and BAVEJA, A. (2002). A data-driven software tool for enabling cooperative information sharing among police departments. *European J. Oper. Res.* **141** 660–678.

[83] ROYSET, J. O. and WETS, R. J.-B. (2020). Variational analysis of constrained M-estimators. *Ann. Statist.* **48** 2759–2790. MR4152120 https://doi.org/10.1214/19-AOS1905

[84] ROZO, E. and RYKOFF, E. S. (2014). redMaPPer II: X-ray and SZ performance benchmarks for the SDSS catalog. *Astrophys. J.* **783** 80.

[85] SACHS, A.-L. (2015). The data-driven newsvendor with censored demand observations. In *Retail Analytics* 35–56. Springer, Berlin.

[86] SAEED, M., VILLARROEL, M., REISNER, A. T., CLIFFORD, G., LEHMAN, L.-W., MOODY, G., HELDT, T., KYAW, T. H., MOODY, B. et al. (2011). Multiparameter intelligent monitoring in intensive care II: A public-access intensive care unit database. *Crit. Care Med.* **39** 952–960. https://doi.org/10.1097/CCM.0b013e31820a92c6

[87] SAHINIDIS, N. V. (1996). BARON: A general purpose global optimization software package. *J. Global Optim.* **8** 201–205. MR1376505 https://doi.org/10.1007/BF00138693

[88] SALAM, M. (2017). The opioid epidemic: A crisis years in the making. *The New York Times* **26**.

[89] SCHÖLKOPF, B., PLATT, J. C., SHAWE-TAYLOR, J., SMOLA, A. J. and WILLIAMSON, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Comput.* **13** 1443–1471. https://doi.org/10.1162/089976601750264965

[90] SMITH, J. W., EVERHART, J., DICKSON, W., KNOWLER, W. and JOHANNES, R. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* 261.

[91] SMOLA, A. J., VISHWANATHAN, S. and HOFMANN, T. (2005). Kernel methods for missing variables. In *AISTATS*.

[92] SONG, L., SMOLA, A., GRETTON, A. and BORGWARDT, K. M. (2007). A dependence maximization view of clustering. In *International Conference on Machine Learning* 815–822.

[93] SONG, L., SMOLA, A., GRETTON, A., BORGWARDT, K. M. and BEDO, J. (2007). Supervised feature selection via dependence estimation. In *International Conference on Machine Learning* 823–830.

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

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

[96] THOMPSON, J. J., BLAIR, M. R., CHEN, L. and HENREY, A. J. (2013). Video game telemetry as a critical tool in the study of complex skill learning. *PLoS ONE* **8** e75129.
[97] TOMIOKA, R. and SUZUKI, T. (2014). Spectral norm of random tensors. Preprint. Available at arXiv:1407.1870.
[98] TSANAS, A. and XIFARA, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy Build*. **49** 560–567.
[99] TUY, H. (1995). D.C. optimization: Theory, methods and algorithms. In *Handbook of Global Optimization*. *Nonconvex Optim*. *Appl*. **2** 149–216. Kluwer Academic, Dordrecht. MR1377085 https://doi.org/10.1007/978-1-4615-2025-2_4
[100] TYRRELL ROCKAFELLAR, R. and WETS, R. J.-B. (2009). *Variational Analysis* **317**. Springer, Berlin.
[101] VIGERSKE, S. and GLEIXNER, A. (2018). SCIP: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optim*. *Methods Softw*. **33** 563–593. MR3783674 https://doi.org/10.1080/10556788.2017.1335312
[102] 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
[103] WOODWORTH, B., GUNASEKAR, S., OHANNESSIAN, M. I. and SREBRO, N. (2017). Learning non-discriminatory predictors. Preprint. Available at arXiv:1702.06081.
[104] WU, Z., SONG, S., KHOSLA, A., YU, F., ZHANG, L., TANG, X. and XIAO, J. (2015). 3D shapenets: A deep representation for volumetric shapes. In *IEEE CVPR* 1912–1920.
[105] YEH, I.-C. and LIEN, C.-H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Syst*. *Appl*. **36** 2473–2480.
[106] YUILLE, A. L. and RANGARAJAN, A. (2002). The concave-convex procedure (CCCP). In *Advances in Neural Information Processing Systems* 1033–1040.
[107] ZAFAR, M. B., VALERA, I., RODRIGUEZ, M. G. and GUMMADI, K. P. (2017). Fairness constraints: Mechanisms for fair classification. In *AISTATS*.
[108] ZAHEER, M., KOTTUR, S., RAVANBAKHSH, S., POCZOS, B., SALAKHUTDINOV, R. R. and SMOLA, A. J. (2017). Deep sets. In *Advances in Neural Information Processing Systems* 3391–3401.
[109] ZEMEL, R., WU, Y., SWERSKY, K., PITASSI, T. and DWORK, C. (2013). Learning fair representations. In *ICML* 325–333.
[110] ZHANG, B. H., LEMOINE, B. and MITCHELL, M. (2018). Mitigating unwanted biases with adversarial learning. Preprint. Available at arXiv:1801.07593.
[111] ZHAO, P., MOHAN, S. and VASUDEVAN, R. (2020). Optimal control of polynomial hybrid systems via convex relaxations. *IEEE Trans*. *Automat*. *Control* **65** 2062–2077. MR4091823 https://doi.org/10.1109/tac.2019.2929110
[112] ZHOU, F., CLAIRE, Q. and KING, R. D. (2014). Predicting the geographical origin of music. In 2014 *IEEE International Conference on Data Mining* 1115–1120. IEEE, New York.
[113] ZLIOBAITE, I. (2015). On the relation between accuracy and fairness in binary classification. Preprint. Available at arXiv:1505.05723.
[114] ZOLOTAREV, V. M. (1976). Metric distances in spaces of random variables and of their distributions. *Math*. *USSR*, *Sb*. **30** 373–401.

The Annals of Statistics
2022, Vol. 50, No. 6, 3174–3196
https://doi.org/10.1214/22-AOS2221

HALF-TREK CRITERION FOR IDENTIFIABILITY OF LATENT VARIABLE MODELS

BY RINA FOYGEL BARBER[1,a], MATHIAS DRTON[2,b], NILS STURMA[2,c] AND LUCA WEIHS[3,d]

[1]*Department of Statistics, University of Chicago,* [a]*rina@uchicago.edu*
[2]*School of Computation, Information and Technology, Technical University of Munich,* [b]*mathias.drton@tum.de,* [c]*nils.sturma@tum.de*
[3]*Allen Institute for AI,* [d]*lucaw@allenai.org*

We consider linear structural equation models with latent variables and develop a criterion to certify whether the direct causal effects between the observable variables are identifiable based on the observed covariance matrix. Linear structural equation models assume that both observed and latent variables solve a linear equation system featuring stochastic noise terms. Each model corresponds to a directed graph whose edges represent the direct effects that appear as coefficients in the equation system. Prior research has developed a variety of methods to decide identifiability of direct effects in a latent projection framework, in which the confounding effects of the latent variables are represented by correlation among noise terms. This approach is effective when the confounding is sparse and effects only small subsets of the observed variables. In contrast, the new latent-factor half-trek criterion (LF-HTC) we develop in this paper operates on the original unprojected latent variable model and is able to certify identifiability in settings, where some latent variables may also have dense effects on many or even all of the observables. Our LF-HTC is an effective sufficient criterion for rational identifiability, under which the direct effects can be uniquely recovered as rational functions of the joint covariance matrix of the observed random variables. When restricting the search steps in LF-HTC to consider subsets of latent variables of bounded size, the criterion can be verified in time that is polynomial in the size of the graph.

REFERENCES

BARBER, R. F, DRTON, M., STURMA, N. and WEIHS, L. (2022). Supplement to "Half-trek criterion for identifiability of latent variable models." https://doi.org/10.1214/22-AOS2221SUPP

BASU, S., POLLACK, R. and ROY, M.-F. (2006). *Algorithms in Real Algebraic Geometry*, 2nd ed. *Algorithms and Computation in Mathematics* **10**. Springer, Berlin. MR2248869

BENEDETTI, R. and RISLER, J.-J. (1990). *Real Algebraic and Semi-Algebraic Sets*. *Actualités Mathématiques*. [*Current Mathematical Topics*]. Hermann, Paris. MR1070358

BOCHNAK, J., COSTE, M. and ROY, M.-F. (1998). *Real Algebraic Geometry*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **36**. Springer, Berlin. Translated from the 1987 French original, Revised by the authors. MR1659509 https://doi.org/10.1007/978-3-662-03718-8

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0996025 https://doi.org/10.1002/9781118619179

BOWDEN, R. J. and TURKINGTON, D. A. (1984). *Instrumental Variables*. *Econometric Society Monographs in Quantitative Economics* **8**. Cambridge Univ. Press, Cambridge. MR0798790

BRITO, C. and PEARL, J. (2002). Generalized instrumental variables. In *Proceedings of the* 18*th Conference on Uncertainty in Artificial Intelligence* (*UAI*). *UAI'*02 85–93. Morgan Kaufmann Publishers Inc., San Francisco, CA.

MSC2020 subject classifications. 62H22, 62J05, 62R01.
Key words and phrases. Covariance matrix, factor analysis, graphical model, hidden variables, latent variables, parameter identification, structural equation model.

BRITO, C. and PEARL, J. (2006). Graphical condition for identification in recursive SEM. In *Proceedings of the* 22*nd Conference on Uncertainty in Artificial Intelligence* (*UAI*) 47–54. AUAI Press, Arlington, VA.

CHEN, B., KUMOR, D. and BAREINBOIM, E. (2017). Identification and model testing in linear structural equation models using auxiliary variables. In *Proceedings of the* 34*th International Conference on Machine Learning* (*ICML*). *ICML*'17 757–766. JMLR, USA.

CHEN, B., PEARL, J. and BAREINBOIM, E. (2016). Incorporating knowledge into structural equation models using auxiliary variables. In *Proceedings of the* 25*th International Joint Conference on Artificial Intelligence* (*IJCAI*). *IJCAI*'16 3577–3583. AAAI Press, Menlo Park, CA.

CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L. and STEIN, C. (2009). *Introduction to Algorithms*, 3rd ed. MIT Press, Cambridge, MA. MR2572804

COX, D., LITTLE, J. and O'SHEA, D. (2007). *Ideals*, *Varieties*, *and Algorithms*, 3rd ed. *Undergraduate Texts in Mathematics*. Springer, New York. An introduction to computational algebraic geometry and commutative algebra. MR2290010 https://doi.org/10.1007/978-0-387-35651-8

DRTON, M. (2018). Algebraic problems in structural equation modeling. In *The* 50*th Anniversary of Gröbner Bases*. *Adv*. *Stud*. *Pure Math*. **77** 35–86. Math. Soc. Japan, Tokyo. MR3839705 https://doi.org/10.2969/aspm/07710035

DRTON, M., STURMFELS, B. and SULLIVANT, S. (2007). Algebraic factor analysis: Tetrads, pentads and beyond. *Probab. Theory Related Fields* **138** 463–493. MR2299716 https://doi.org/10.1007/s00440-006-0033-2

DRTON, M. and WEIHS, L. (2016). Generic identifiability of linear structural equation models by ancestor decomposition. *Scand*. *J*. *Stat*. **43** 1035–1045. MR3573674 https://doi.org/10.1111/sjos.12227

DRTON, M. and YU, J. (2010). On a parametrization of positive semidefinite matrices with zeros. *SIAM J*. *Matrix Anal*. *Appl*. **31** 2665–2680. MR2740626 https://doi.org/10.1137/100783170

FOYGEL, R., DRAISMA, J. and DRTON, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *Ann*. *Statist*. **40** 1682–1713. MR3015040 https://doi.org/10.1214/12-AOS1012

FOYGEL BARBER, R., DRTON, M., STURMA, N. and WEIHS, L. (2022). SEMID: Identifiability of linear structural equation models. R package version 0.4.0.

GARCIA-PUENTE, L. D., SPIELVOGEL, S. and SULLIVANT, S. (2010). Identifying causal effects with computer algebra. In *Proceedings of the* 26*th Conference on Uncertainty in Artificial Intelligence* (*UAI*) AUAI Press, Arlington, VA.

GESSEL, I. and VIENNOT, G. (1985). Binomial determinants, paths, and hook length formulae. *Adv*. *Math*. **58** 300–321. MR0815360 https://doi.org/10.1016/0001-8708(85)90121-5

HARTSHORNE, R. (1977). *Algebraic Geometry*. *Graduate Texts in Mathematics*, *No*. 52. Springer, New York. MR0463157

KUMOR, D., CHEN, B. and BAREINBOIM, E. (2019). Efficient identification in linear structural causal models with instrumental cutsets. In *Advances in Neural Information Processing Systems* (*NeurIPS*) **32** 12477–12486. Neural Information Processing Systems Foundation. Curran Associates, Red Hook, NY.

KUMOR, D., CINELLI, C. and BAREINBOIM, E. (2020). Efficient identification in linear structural causal models with auxiliary cutsets. In *Proceedings of the* 37*th International Conference on Machine Learning* (*ICML*). *Proceedings of Machine Learning Research* **119** 5501–5510. PMLR, USA.

KUROKI, M. and PEARL, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101** 423–437. MR3215357 https://doi.org/10.1093/biomet/ast066

LEE, S. and BAREINBOIM, E. (2021). Causal identification with matrix equations. In *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang and J. W. Vaughan, eds.) **34** 9468–9479. Curran Associates, Red Hook, NY.

LEUNG, D., DRTON, M. and HARA, H. (2016). Identifiability of directed Gaussian graphical models with one latent source. *Electron*. *J*. *Stat*. **10** 394–422. MR3466188 https://doi.org/10.1214/16-EJS1111

LINDSTRÖM, B. (1973). On the vector representations of induced matroids. *Bull*. *Lond*. *Math*. *Soc*. **5** 85–90. MR0335313 https://doi.org/10.1112/blms/5.1.85

MAATHUIS, M., DRTON, M., LAURITZEN, S. and WAINWRIGHT, M., eds. (2019). *Handbook of Graphical Models*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3889064

MIAO, W., GENG, Z. and TCHETGEN TCHETGEN, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105** 987–993. MR3877879 https://doi.org/10.1093/biomet/asy038

OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann*. *Statist*. **1** 763–765. MR0331643

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

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. Foundations and learning algorithms. MR3822088

R CORE TEAM, (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SHAFAREVICH, I. R. (2013). *Basic Algebraic Geometry*. 1, 3rd ed. Springer, Heidelberg. MR3100243

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

STANGHELLINI, E. and WERMUTH, N. (2005). On the identification of path analysis models with one hidden variable. *Biometrika* **92** 337–350. MR2201363 https://doi.org/10.1093/biomet/92.2.337

SULLIVANT, S., TALASKA, K. and DRAISMA, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist*. **38** 1665–1685. MR2662356 https://doi.org/10.1214/09-AOS760

TIAN, J. (2005). Identifying direct causal effects in linear models. In *Proceedings of the* 20*th National Conference on Artificial Intelligence* (*AAAI*). *AAAI'*05 346–352. Association for the Advancement of Artificial Intelligence. AAAI Press, Menlo Park, CA.

TIAN, J. (2009). Parameter identification in a class of linear structural equation models. In *Proceedings of the* 21*st International Joint Conference on Artificial Intelligence* (*IJCAI*). *IJCAI'*09 1970–1975. AAAI Press, Menlo Park, CA.

VAN DER ZANDER, B., TEXTOR, J. and LISKIEWICZ, M. (2015). Efficiently finding conditional instruments for causal inference. In *Proceedings of the* 24*th International Joint Conference on Artificial Intelligence* (*IJCAI*). *IJCAI'*15 3243–3249. AAAI Press, Menlo Park, CA.

WEIHS, L., ROBINSON, B., DUFRESNE, E., KENKEL, J., KUBJAS, K., MCGEE, R. II, NGUYEN, N., ROBEVA, E. and DRTON, M. (2018). Determinantal generalizations of instrumental variables. *J. Causal Inference* **6** 20170009. MR4351483 https://doi.org/10.1515/jci-2017-0009

The Annals of Statistics
2022, Vol. 50, No. 6, 3197–3222
https://doi.org/10.1214/22-AOS2222

ON RESAMPLING SCHEMES FOR PARTICLE FILTERS WITH WEAKLY INFORMATIVE OBSERVATIONS

BY NICOLAS CHOPIN[1,a], SUMEETPAL S. SINGH[2,b], TOMÁS SOTO[3,c] AND MATTI VIHOLA[4,d]

[1]*ENSAE, Institut Polytechnique de Paris,* [a]*nicolas.chopin@ensae.fr*

[2]*Department of Engineering, University of Cambridge,* [b]*sss40@eng.cam.ac.uk*

[3]*School of Engineering Science, LUT University,* [c]*Tomas.Soto@lut.fi*

[4]*Department of Mathematics and Statistics, University of Jyväskylä,* [d]*matti.s.vihola@jyu.fi*

We consider particle filters with weakly informative observations (or 'potentials') relative to the latent state dynamics. The particular focus of this work is on particle filters to approximate time-discretisations of continuous-time Feynman–Kac path integral models—a scenario that naturally arises when addressing filtering and smoothing problems in continuous time—but our findings are indicative about weakly informative settings beyond this context too. We study the performance of different resampling schemes, such as systematic resampling, SSP (Srinivasan sampling process) and stratified resampling, as the time-discretisation becomes finer and also identify their continuous-time limit, which is expressed as a suitably defined 'infinitesimal generator.' By contrasting these generators, we find that (certain modifications of) systematic and SSP resampling 'dominate' stratified and independent 'killing' resampling in terms of their limiting overall resampling rate. The reduced intensity of resampling manifests itself in lower variance in our numerical experiment. This efficiency result, through an ordering of the resampling rate, is new to the literature. The second major contribution of this work concerns the analysis of the limiting behaviour of the entire population of particles of the particle filter as the time discretisation becomes finer. We provide the first proof, under general conditions, that the particle approximation of the discretised continuous-time Feynman–Kac path integral models converges to a (uniformly weighted) continuous-time particle system.

REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

[2] ARNAUDON, M. and DEL MORAL, P. (2020). A duality formula and a particle Gibbs sampler for continuous time Feynman–Kac measures on path spaces. *Electron. J. Probab.* **25** Paper No. 157. MR4193898 https://doi.org/10.1214/20-ejp546

[3] BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Rev.* **59** 65–98. MR3605826 https://doi.org/10.1137/141000671

[4] CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068 https://doi.org/10.1214/10-AIHP358

[5] CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo. Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

[6] CHOPIN, N., SINGH, S. S., SOTO, T. and VIHOLA, M. (2022). Supplement to "On resampling schemes for particle filters with weakly informative observations." https://doi.org/10.1214/22-AOS2222SUPPA, https://doi.org/10.1214/22-AOS2222SUPPB

[7] CRISAN, D., DEL MORAL, P. and LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields* **5** 293–318. MR1710982

MSC2020 subject classifications. Primary 65C35; secondary 65C05, 65C60, 60J25.
Key words and phrases. Feynman–Kac model, hidden Markov model, particle filter, path integral, resampling.

[8] Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications (New York)*. Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

[9] Del Moral, P. (2013). *Mean Field Simulation for Monte Carlo Integration. Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

[10] Del Moral, P., Jacob, P. E., Lee, A., Murray, L. and Peters, G. W. (2013). Feynman–Kac particle integration with geometric interacting jumps. *Stoch. Anal. Appl.* **31** 830–871. MR3175799 https://doi.org/10.1080/07362994.2013.817247

[11] Del Moral, P. and Miclo, L. (2000). A Moran particle system approximation of Feynman–Kac formulae. *Stochastic Process. Appl.* **86** 193–216. MR1741805 https://doi.org/10.1016/S0304-4149(99)00094-0

[12] Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV. Lecture Notes in Math.* **1729** 1–145. Springer, Berlin. MR1768060 https://doi.org/10.1007/BFb0103798

[13] Douc, R. and Cappé, O. (2005). Comparison of resampling schemes for particle filtering. In *Proc. ISPA* 2005 64–69. https://doi.org/10.1109/ISPA.2005.195385

[14] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[15] Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704 https://doi.org/10.1214/09-AOS735

[16] Gerber, M., Chopin, N. and Whiteley, N. (2019). Negative association, ordering and convergence of resampling methods. *Ann. Statist.* **47** 2236–2260. MR3953450 https://doi.org/10.1214/18-AOS1746

[17] Glynn, P. W. and Whitt, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.* **40** 505–520. MR1180030 https://doi.org/10.1287/opre.40.3.505

[18] Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F* **140** 107–113. https://doi.org/10.1049/ip-f-2.1993.0015

[19] Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737

[20] Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators. *J. Stat. Comput. Simul.* **59** 1–23. MR1474178 https://doi.org/10.1080/00949659708811843

[21] Hoare, C. A. R. (1962). Quicksort. *Comput. J.* **5** 10–15. MR0142216 https://doi.org/10.1093/comjnl/5.1.10

[22] Liu, J. S. and Chen, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576. MR3363399 https://doi.org/10.1080/01621459.1995.10476549

[23] Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198 https://doi.org/10.2307/2669847

[24] Murray, L. M., Lee, A. and Jacob, P. E. (2016). Parallel resampling in the particle filter. *J. Comput. Graph. Statist.* **25** 789–805. MR3533638 https://doi.org/10.1080/10618600.2015.1062015

[25] Rousset, M. (2006). On the control of an interacting particle estimation of Schrödinger ground states. *SIAM J. Math. Anal.* **38** 824–844. MR2262944 https://doi.org/10.1137/050640667

[26] Vihola, M. (2020). Ergonomic and reliable Bayesian inference with adaptive Markov chain Monte Carlo. In *Wiley StatsRef: Statistics Reference Online* (M. Davidian, R. S. Kenett, N. T. Longford, G. Molenberghs, W. W. Piegorsch and F. Ruggeri, eds.) Wiley, New York. https://doi.org/10.1002/9781118445112.stat08286

[27] Vihola, M., Helske, J. and Franks, J. (2020). Importance sampling type estimators based on approximate marginal Markov chain Monte Carlo. *Scand. J. Stat.* **47** 1339–1376. MR4178196 https://doi.org/10.1111/sjos.12492

[28] Whiteley, N., Lee, A. and Heine, K. (2016). On the role of interaction in sequential Monte Carlo algorithms. *Bernoulli* **22** 494–529. MR3449791 https://doi.org/10.3150/14-BEJ666

The Annals of Statistics
2022, Vol. 50, No. 6, 3223–3248
https://doi.org/10.1214/22-AOS2225

TOWARD THEORETICAL UNDERSTANDINGS OF ROBUST MARKOV DECISION PROCESSES: SAMPLE COMPLEXITY AND ASYMPTOTICS

BY WENHAO YANG[1,a], LIANGYU ZHANG[1,b] AND ZHIHUA ZHANG[2,c]

[1]*Academy for Advanced Interdisciplinary Studies, Peking University,* [a]*yangwenhaosms@pku.edu.cn,* [b]*zhangliangyu@pku.edu.cn*

[2]*School of Mathematical Sciences, Peking University,* [c]*zhzhang@math.pku.edu.cn*

In this paper, we study the nonasymptotic and asymptotic performances of the optimal robust policy and value function of robust Markov Decision Processes (MDPs), where the optimal robust policy and value function are estimated from a generative model. While prior work focusing on nonasymptotic performances of robust MDPs is restricted in the setting of the KL uncertainty set and (s,a)-rectangular assumption, we improve their results and also consider other uncertainty sets, including the L_1 and χ^2 balls. Our results show that when we assume (s,a)-rectangular on uncertainty sets, the sample complexity is about $\widetilde{\mathcal{O}}(\frac{|\mathcal{S}|^2|\mathcal{A}|}{\varepsilon^2\rho^2(1-\gamma)^4})$. In addition, we extend our results from the (s,a)-rectangular assumption to the s-rectangular assumption. In this scenario, the sample complexity varies with the choice of uncertainty sets and is generally larger than the case under the (s,a)-rectangular assumption. Moreover, we also show that the optimal robust value function is asymptotically normal with a typical rate $\sqrt{n}$ under the (s,a) and s-rectangular assumptions from both theoretical and empirical perspectives.

REFERENCES

[1] AGARWAL, A., KAKADE, S. and YANG, L. F. (2020). Model-based reinforcement learning with a generative model is minimax optimal. In *Proceedings of Thirty Third Conference on Learning Theory* 67–83.

[2] AGARWAL, R., SCHUURMANS, D. and NOROUZI, M. (2020). An optimistic perspective on offline reinforcement learning. In *Proceedings of the* 37*th International Conference on Machine Learning* 104–114.

[3] BEHZADIAN, B., RUSSEL, R. H., PETRIK, M. and HO, C. P. (2021). Optimizing percentile criterion using robust MDPs. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* 1009–1017.

[4] BEN-TAL, A., DEN HERTOG, D., DE WAEGENAERE, A., MELENBERG, B. and RENNEN, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Manage. Sci.* **59** 341–357.

[5] BERTSEKAS, D. P. and TSITSIKLIS, J. N. (1995). Neuro-dynamic programming: An overview. In *Proceedings of* 1995 34*th IEEE Conference on Decision and Control* **1** 560–564.

[6] BERTSIMAS, D., GUPTA, V. and KALLUS, N. (2018). Data-driven robust optimization. *Math. Program.* **167** 235–292. MR3755733 https://doi.org/10.1007/s10107-017-1125-8

[7] BLANCHET, J. and MURTHY, K. (2019). Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* **44** 565–600. MR3959085 https://doi.org/10.1287/moor.2018.0936

[8] CHEN, J. and JIANG, N. (2019). Information-theoretic considerations in batch reinforcement learning. In *Proceedings of the* 36*th International Conference on Machine Learning* 1042–1051.

[9] 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.* **33** 9398–9411.

[10] DANN, C., NEUMANN, G. and PETERS, J. (2014). Policy evaluation with temporal differences: A survey and comparison. *J. Mach. Learn. Res.* **15** 809–883. MR3195332

MSC2020 subject classifications. Primary 62C05, 62F12; secondary 68Q32.
Key words and phrases. Model-based reinforcement learning, robust MDPs, distributional robustness, f-divergence set.

[11] DELAGE, E. and YE, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58** 595–612. MR2680566 https://doi.org/10.1287/opre.1090.0741

[12] DERMAN, E. and MANNOR, S. (2020). Distributional robustness and regularization in reinforcement learning. ArXiv preprint. Available at arXiv:2003.02894.

[13] DUAN, Y., JIA, Z. and WANG, M. (2020). Minimax-optimal off-policy evaluation with linear function approximation. In *Proceedings of the* 37*th International Conference on Machine Learning* 2701–2709.

[14] DUAN, Y., JIN, C. and LI, Z. (2021). Risk bounds and Rademacher complexity in batch reinforcement learning. In *Proceedings of the* 38*th International Conference on Machine Learning* 2892–2902.

[15] DUCHI, J. and NAMKOONG, H. (2019). Variance-based regularization with convex objectives. *J. Mach. Learn. Res.* **20** Paper No. 68. MR3960922

[16] DUCHI, J. C., GLYNN, P. W. and NAMKOONG, H. (2021). Statistics of robust optimization: A generalized empirical likelihood approach. *Math. Oper. Res.* **46** 946–969. MR4312583 https://doi.org/10.1287/moor.2020.1085

[17] 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

[18] DUDÍK, M., ERHAN, D., LANGFORD, J. and LI, L. (2014). Doubly robust policy evaluation and optimization. *Statist. Sci.* **29** 485–511. MR3300356 https://doi.org/10.1214/14-STS500

[19] EPSTEIN, L. G. and SCHNEIDER, M. (2003). Recursive multiple-priors. *J. Econom. Theory* **113** 1–31. MR2017864 https://doi.org/10.1016/S0022-0531(03)00097-8

[20] FARAJTABAR, M., CHOW, Y. and GHAVAMZADEH, M. (2018). More robust doubly robust off-policy evaluation. In *Proceedings of the* 35*th International Conference on Machine Learning* 1447–1456.

[21] FUJIMOTO, S., MEGER, D. and PRECUP, D. (2019). Off-policy deep reinforcement learning without exploration. In *Proceedings of the* 36*th International Conference on Machine Learning* 2052–2062.

[22] GAO, R. and KLEYWEGT, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. ArXiv Preprint. Available at arXiv:1604.02199.

[23] GHAVAMZADEH, M., PETRIK, M. and CHOW, Y. (2016). Safe policy improvement by minimizing robust baseline regret. *Adv. Neural Inf. Process. Syst.* **29** 2298–2306.

[24] GHESHLAGHI AZAR, M., MUNOS, R. and KAPPEN, H. J. (2013). Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model. *Mach. Learn.* **91** 325–349. MR3064431 https://doi.org/10.1007/s10994-013-5368-1

[25] GOYAL, V. and GRAND-CLEMENT, J. (2022). Robust Markov decision processes: Beyond rectangularity. *Math. Oper. Res.*

[26] GRÜNEWÄLDER, S., LEVER, G., BALDASSARRE, L., PONTIL, M. and GRETTON, A. (2012). Modelling transition dynamics in MDPs with RKHS embeddings. In *Proceedings of the* 29*th International Coference on International Conference on Machine Learning* 1603–1610.

[27] HAARNOJA, T., ZHOU, A., ABBEEL, P. and LEVINE, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the* 35*th International Conference on Machine Learning* 1861–1870.

[28] 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

[29] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

[30] HO, C. P., PETRIK, M. and WIESEMANN, W. (2018). Fast Bellman updates for robust MDPs. In *Proceedings of the* 35*th International Conference on Machine Learning* 1979–1988.

[31] HO, C. P., PETRIK, M. and WIESEMANN, W. (2021). Partial policy iteration for L_1-robust Markov decision processes. *J. Mach. Learn. Res.* **22** Paper No. 275. MR4353054

[32] IYENGAR, G. N. (2005). Robust dynamic programming. *Math. Oper. Res.* **30** 257–280. MR2142033 https://doi.org/10.1287/moor.1040.0129

[33] JIANG, N. and LI, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *Proceedings of the* 33*rd International Conference on Machine Learning* 652–661.

[34] JIN, C., ALLEN-ZHU, Z., BUBECK, S. and JORDAN, M. I. (2018). Is Q-learning provably efficient? *Adv. Neural Inf. Process. Syst.* **31**.

[35] JIN, C., YANG, Z., WANG, Z. and JORDAN, M. I. (2020). Provably efficient reinforcement learning with linear function approximation. In *Proceedings of Thirty Third Conference on Learning Theory* 2137–2143.

[36] JIN, Y., YANG, Z. and WANG, Z. (2021). Is pessimism provably efficient for offline rl? In *Proceedings of the* 38*th International Conference on Machine Learning* 5084–5096.

[37] JONG, N. K. and STONE, P. (2007). Model-based function approximation in reinforcement learning. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems* 1–8.

[38] KALLUS, N. and UEHARA, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** Paper No. 167. MR4209453

[39] KAUFMAN, D. L. and SCHAEFER, A. J. (2013). Robust modified policy iteration. *INFORMS J. Comput.* **25** 396–410. MR3085321 https://doi.org/10.1287/ijoc.1120.0509

[40] LAM, H. (2016). Robust sensitivity analysis for stochastic systems. *Math. Oper. Res.* **41** 1248–1275. MR3544795 https://doi.org/10.1287/moor.2015.0776

[41] LAZARIC, A., GHAVAMZADEH, M. and MUNOS, R. (2012). Finite-sample analysis of least-squares policy iteration. *J. Mach. Learn. Res.* **13** 3041–3074. MR2997720

[42] LE, H., VOLOSHIN, C. and YUE, Y. (2019). Batch policy learning under constraints. In *Proceedings of the 36th International Conference on Machine Learning* 3703–3712.

[43] LEE, J. and RAGINSKY, M. (2018). Minimax statistical learning with Wasserstein distances. *Adv. Neural Inf. Process. Syst.* **31**.

[44] LI, L., MUNOS, R. and SZEPESVÁRI, C. (2015). Toward minimax off-policy value estimation. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics* 608–616.

[45] LI, X., YANG, W., ZHANG, Z. and JORDAN, M. I. (2021). Polyak–Ruppert averaged Q-leaning is statistically efficient. ArXiv Preprint. Available at arXiv:2112.14582.

[46] LILLICRAP, T. P., HUNT, J. J., PRITZEL, A., HEESS, N., EREZ, T., TASSA, Y., SILVER, D. and WIERSTRA, D. (2015). Continuous control with deep reinforcement learning. In *Proceedings of the 4th International Conference on Learning Representations*.

[47] LIM, S. H., XU, H. and MANNOR, S. (2013). Reinforcement learning in robust Markov decision processes. *Adv. Neural Inf. Process. Syst.* **26** 701–709.

[48] LIU, Q., LI, L., TANG, Z. and ZHOU, D. (2018). Breaking the curse of horizon: Infinite-horizon off-policy estimation. *Adv. Neural Inf. Process. Syst.* **31**.

[49] MANNOR, S., MEBEL, O. and XU, H. (2012). Lightning does not strike twice: Robust MDPs with coupled uncertainty. In *Proceedings of the 29th International Conference on Machine Learning* 451–458.

[50] MANNOR, S., SIMESTER, D., SUN, P. and TSITSIKLIS, J. N. (2004). Bias and variance in value function estimation. In *Proceedings of the 21st International Conference on Machine Learning* 72.

[51] MNIH, V., BADIA, A. P., MIRZA, M., GRAVES, A., LILLICRAP, T., HARLEY, T., SILVER, D. and KAVUKCUOGLU, K. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning* 1928–1937.

[52] 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.

[53] MOHRI, M., ROSTAMIZADEH, A. and TALWALKAR, A. (2018). *Foundations of Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3931734

[54] MUNOS, R. and SZEPESVÁRI, C. (2008). Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **9** 815–857. MR2417255

[55] NILIM, A. and EL GHAOUI, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Oper. Res.* **53** 780–798. MR2171651 https://doi.org/10.1287/opre.1050.0216

[56] PANAGANTI, K. and KALATHIL, D. (2022). Sample complexity of robust reinforcement learning with a generative model. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics* 9582–9602.

[57] PENG, X. B., ANDRYCHOWICZ, M., ZAREMBA, W. and ABBEEL, P. (2018). Sim-to-real transfer of robotic control with dynamics randomization. In 2018 *IEEE International Conference on Robotics and Automation* (*ICRA*) 3803–3810.

[58] PETRIK, M. (2012). Approximate dynamic programming by minimizing distributionally robust bounds. In *Proceedings of the 29th International Conference on Machine Learning* 1595–1602.

[59] PETRIK, M. and RUSSEL, R. H. (2019). Beyond confidence regions: Tight Bayesian ambiguity sets for robust mdps. *Adv. Neural Inf. Process. Syst.* **32**.

[60] PETRIK, M. and SUBRAMANIAN, D. (2014). RAAM: The benefits of robustness in approximating aggregated MDPs in reinforcement learning. *Adv. Neural Inf. Process. Syst.* **27**.

[61] PUTERMAN, M. L. (1994). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1270015

[62] QI, Z. and LIAO, P. (2020). Robust batch policy learning in markov decision processes. ArXiv preprint. Available at arXiv:2011.04185.

[63] SASON, I. and VERDÚ, S. (2016). f-divergence inequalities. *IEEE Trans. Inf. Theory* **62** 5973–6006. MR3565096 https://doi.org/10.1109/TIT.2016.2603151

[64] SHAPIRO, A. (2017). Distributionally robust stochastic programming. *SIAM J. Optim.* **27** 2258–2275. MR3715383 https://doi.org/10.1137/16M1058297

[65] SI, N., ZHANG, F., ZHOU, Z. and BLANCHET, J. (2020). Distributionally robust policy evaluation and learning in offline contextual bandits. In *Proceedings of the* 37*th International Conference on Machine Learning* 8884–8894.

[66] SIDFORD, A., WANG, M., WU, X., YANG, L. F. and YE, Y. (2018). Near-optimal time and sample complexities for solving Markov decision processes with a generative model. In *Advances in Neural Information Processing Systems* **31**.

[67] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* **529** 484–489.

[68] SMIRNOVA, E., DOHMATOB, E. and MARY, J. (2019). Distributionally robust reinforcement learning. ArXiv Preprint. Available at arXiv:1902.08708.

[69] SWAMINATHAN, A. and JOACHIMS, T. (2015). The self-normalized estimator for counterfactual learning. *Adv. Neural Inf. Process. Syst.* **28**.

[70] THOMAS, P. and BRUNSKILL, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *Proceedings of the* 33*rd International Conference on Machine Learning* 2139–2148.

[71] 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

[72] 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

[73] WANG, R., FOSTER, D. and KAKADE, S. M. (2020). What are the statistical limits of offline RL with linear function approximation? In *Proceedings of the* 9*th International Conference on Learning Representations*.

[74] WIESEMANN, W., KUHN, D. and RUSTEM, B. (2013). Robust Markov decision processes. *Math. Oper. Res.* **38** 153–183. MR3029483 https://doi.org/10.1287/moor.1120.0566

[75] WOZABAL, D. (2012). A framework for optimization under ambiguity. *Ann. Oper. Res.* **193** 21–47. MR2874755 https://doi.org/10.1007/s10479-010-0812-0

[76] XIAO, C., WU, Y., MEI, J., DAI, B., LATTIMORE, T., LI, L., SZEPESVARI, C. and SCHUURMANS, D. (2021). On the optimality of batch policy optimization algorithms. In *Proceedings of the* 38*th International Conference on Machine Learning* 11362–11371.

[77] XIE, T. and JIANG, N. (2021). Batch value-function approximation with only realizability. In *Proceedings of the* 38*th International Conference on Machine Learning* 11404–11413.

[78] XIE, T., MA, Y. and WANG, Y.-X. (2019). Toward optimal off-policy evaluation for reinforcement learning with marginalized importance sampling. *Adv. Neural Inf. Process. Syst.* **32**.

[79] XU, H. and MANNOR, S. (2006). The robustness-performance tradeoff in Markov decision processes. *Adv. Neural Inf. Process. Syst.* **19**.

[80] XU, H. and MANNOR, S. (2009). Parametric regret in uncertain Markov decision processes. In *Proceedings of the* 48*h IEEE Conference on Decision and Control* (*CDC*) *Held Jointly with* 2009 28*th Chinese Control Conference* 3606–3613.

[81] YANG, W., ZHANG, L. and ZHANG, Z. (2022). Supplement to "Toward theoretical understandings of robust Markov decision processes: Sample complexity and asymptotics." https://doi.org/10.1214/22-AOS2225SUPP

[82] YIN, M., BAI, Y. and WANG, Y.-X. (2021). Near-optimal provable uniform convergence in offline policy evaluation for reinforcement learning. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* 1567–1575.

[83] YIN, M. and WANG, Y.-X. (2020). Asymptotically efficient off-policy evaluation for tabular reinforcement learning. In *Proceedings of the* 23*rd International Conference on Artificial Intelligence and Statistics* 3948–3958.

[84] ZHAO, W., QUERALTA, J. P. and WESTERLUND, T. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: A survey. In 2020 *IEEE Symposium Series on Computational Intelligence* (*SSCI*) 737–744.

[85] ZHOU, Z., BAI, Q., ZHOU, Z., QIU, L., BLANCHET, J. and GLYNN, P. (2021). Finite-sample regret bound for distributionally robust offline tabular reinforcement learning. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* 3331–3339.

The Annals of Statistics
2022, Vol. 50, No. 6, 3249–3280
https://doi.org/10.1214/22-AOS2226

LIMITING DISTRIBUTIONS FOR EIGENVALUES OF SAMPLE CORRELATION MATRICES FROM HEAVY-TAILED POPULATIONS

BY JOHANNES HEINY[1,a] AND JIANFENG YAO[2,b]

[1]*Department of Mathematics, Ruhr University Bochum,* [a]*johannes.heiny@rub.de*
[2]*School of Data Science, The Chinese University of Hong Kong (Shenzhen),* [b]*jeffyao@cuhk.edu.cn*

Consider a p-dimensional population $\mathbf{x} \in \mathbb{R}^p$ with i.i.d. coordinates that are regularly varying with index $\alpha \in (0, 2)$. Since the variance of $\mathbf{x}$ is infinite, the diagonal elements of the sample covariance matrix $\mathbf{S}_n = n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'$ based on a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ from the population tend to infinity as n increases and it is of interest to use instead the sample correlation matrix $\mathbf{R}_n = \{\operatorname{diag}(\mathbf{S}_n)\}^{-1/2} \mathbf{S}_n \{\operatorname{diag}(\mathbf{S}_n)\}^{-1/2}$. This paper finds the limiting distributions of the eigenvalues of $\mathbf{R}_n$ when both the dimension p and the sample size n grow to infinity such that $p/n \to \gamma \in (0, \infty)$. The family of limiting distributions $\{H_{\alpha,\gamma}\}$ is new and depends on the two parameters α and γ. The moments of $H_{\alpha,\gamma}$ are fully identified as sum of two contributions: the first from the classical Marčenko–Pastur law and a second due to heavy tails. Moreover, the family $\{H_{\alpha,\gamma}\}$ has continuous extensions at the boundaries $\alpha = 2$ and $\alpha = 0$ leading to the Marčenko–Pastur law and a modified Poisson distribution, respectively.

Our proofs use the method of moments, the path-shortening algorithm developed in [18] (*Stochastic Process. Appl.* **128** (2018) 2779–2815) and some novel graph counting combinatorics. As a consequence, the moments of $H_{\alpha,\gamma}$ are expressed in terms of combinatorial objects such as Stirling numbers of the second kind. A simulation study on these limiting distributions $H_{\alpha,\gamma}$ is also provided for comparison with the Marčenko–Pastur law.

REFERENCES

[1] ALBRECHER, H. and TEUGELS, J. L. (2007). Asymptotic analysis of a measure of variation. *Theory Probab. Math. Statist.* **74** 1–10.
[2] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662
[3] AUFFINGER, A., BEN AROUS, G. and PÉCHÉ, S. (2009). Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 589–610. MR2548495 https://doi.org/10.1214/08-AIHP188
[4] 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
[5] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845
[6] BAI, Z. D. and YIN, Y. Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** 1729–1741. MR0958213
[7] BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416
[8] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** no. 88, 32 pp. MR2988403 https://doi.org/10.1214/EJP.v17-1962
[9] BASRAK, B., CHO, Y., HEINY, J. and JUNG, P. (2021). Extreme eigenvalue statistics of m-dependent heavy-tailed matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 2100–2127. MR4415391 https://doi.org/10.1214/21-aihp1152

MSC2020 subject classifications. Primary 60B20; secondary 60F05, 60G10, 60G57, 60G70.
Key words and phrases. Sample correlation matrix, limiting spectral distribution, method of moments, infinite variance, Marčenko–Pastur law, stable distribution.

[10] BELINSCHI, S., DEMBO, A. and GUIONNET, A. (2009). Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.* **289** 1023–1055. MR2511659 https://doi.org/10.1007/s00220-009-0822-4

[11] BEN AROUS, G. and GUIONNET, A. (2008). The spectrum of heavy tailed random matrices. *Comm. Math. Phys.* **278** 715–751. MR2373441 https://doi.org/10.1007/s00220-007-0389-x

[12] COMTET, L. (1974). *Advanced Combinatorics*: *The Art of Finite and Infinite Expansions*, enlarged ed. D. Reidel Publishing Co., Dordrecht. MR0460128

[13] DAVIS, R. A., HEINY, J., MIKOSCH, T. and XIE, X. (2016). Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes* **19** 517–547. MR3535965 https://doi.org/10.1007/s10687-016-0251-7

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

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

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

[17] HEINY, J. and MIKOSCH, T. (2017). Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: The iid case. *Stochastic Process. Appl.* **127** 2179–2207. MR3652410 https://doi.org/10.1016/j.spa.2016.10.006

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

[19] HEINY, J. and MIKOSCH, T. (2019). The eigenstructure of the sample covariance matrices of high-dimensional stochastic volatility models with heavy tails. *Bernoulli* **25** 3590–3622. MR4010966 https://doi.org/10.3150/18-bej1103

[20] HEINY, J. and MIKOSCH, T. (2021). Large sample autocovariance matrices of linear processes with heavy tails. *Stochastic Process. Appl.* **141** 344–375. MR4301551 https://doi.org/10.1016/j.spa.2021.07.010

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

[22] JIANG, T. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48. MR2082906

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

[24] 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

[25] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72 (114)** 507–536. MR0208649

[26] MASON, D. M. and ZINN, J. (2005). Acknowledgment of priority: "When does a randomly weighted self-normalized sum converge in distribution?" [Electron. Comm. Probab. **10** (2005), 70–81 (electronic); MR2133894]. *Electron. Commun. Probab.* **10** 297. MR2198604 https://doi.org/10.1214/ECP.v10-1170

[27] PÉCHÉ, S. (2012). Universality in the bulk of the spectrum for complex sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 80–106. MR2919199 https://doi.org/10.1214/11-AIHP442

[28] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042 https://doi.org/10.1214/12-AOS1022

[29] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978 https://doi.org/10.1214/13-AAP939

[30] RENNIE, B. C. and DOBSON, A. J. (1969). On Stirling numbers of the second kind. *J. Combin. Theory* **7** 116–121. MR0241310

[31] SOSHNIKOV, A. (2004). Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electron. Commun. Probab.* **9** 82–91. MR2081462 https://doi.org/10.1214/ECP.v9-1112

[32] SOSHNIKOV, A. (2006). Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical Physics of Quantum Mechanics*. *Lecture Notes in Physics* **690** 351–364. Springer, Berlin. MR2234922 https://doi.org/10.1007/3-540-34273-7_26

[33] TIKHOMIROV, K. (2015). The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math.* **284** 1–20. MR3391069 https://doi.org/10.1016/j.aim.2015.07.020

[34] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

[35] ZHENG, S., CHENG, G., GUO, J. and ZHU, H. (2019). Test for high-dimensional correlation matrices. *Ann*. *Statist*. **47** 2887–2921. MR3988776 https://doi.org/10.1214/18-AOS1768

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

The Annals of Statistics
2022, Vol. 50, No. 6, 3281–3306
https://doi.org/10.1214/22-AOS2228

THE COMPLETION OF COVARIANCE KERNELS

BY KARTIK G. WAGHMARE[a] AND VICTOR M. PANARETOS[b]

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, [a]*kartik.waghmare@epfl.ch,* [b]*victor.panaretos@epfl.ch*

We consider the problem of positive-semidefinite continuation: extending a partially specified covariance kernel from a subdomain Ω of a rectangular domain $I \times I$ to a covariance kernel on the entire domain $I \times I$. For a broad class of domains Ω called *serrated domains*, we are able to present a complete theory. Namely, we demonstrate that a canonical completion always exists and can be explicitly constructed. We characterise all possible completions as suitable perturbations of the canonical completion, and determine necessary and sufficient conditions for a unique completion to exist. We interpret the canonical completion via the graphical model structure it induces on the associated Gaussian process. Furthermore, we show how the estimation of the canonical completion reduces to the solution of a system of linear statistical inverse problems in the space of Hilbert–Schmidt operators, and derive rates of convergence. We conclude by providing extensions of our theory to more general forms of domains, and by demonstrating how our results can be used to construct covariance estimators from sample path fragments of the associated stochastic process. Our results are illustrated numerically by way of a simulation study and a real example.

REFERENCES

[1] ARTJOMENKO, A. P. (1941). Hermitian positive functions and positive functionals. (Russian), Dissertation, Odessa State University. Published in *Teor. Funkcii, Funkcional. Anal. i Prilozen.* **41** 1–16 (1983); **42** 1–21, (1984).

[2] BACHRACH, L. K., HASTIE, T., WANG, M. C., NARASIMHAN, B. and MARCUS, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: A longitudinal study. *J. Clin. Endocrinol. Metab.* **84** 4702–4712.

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

[4] CALDERÓN, A. and PEPINSKY, R. (1952). On the phases of Fourier coefficients for positive real periodic functions. In *Computing Methods and the Phase Problem in X-Ray Crystal Analysis* 339–348. The X-Ray Crystal Analysis Laboratory, Department of Physics, The Pennsylvania State College.

[5] CARATHÉODORY, C. (1907). Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.* **64** 95–115. MR1511425 https://doi.org/10.1007/BF01449883

[6] CARROLL, C., GAJARDO, A., CHEN, Y., DAI, X., FAN, J., HADJIPANTELIS, P. Z., HAN, K., JI, H., MUELLER, H.-G. et al. (2021). *Fdapace*: *Functional Data Analysis and Empirical Dynamics*. R package version 0.5.6. Available at https://CRAN.R-project.org/package=fdapace.

[7] DELAIGLE, A. and HALL, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika* **103** 779–799. MR3620439 https://doi.org/10.1093/biomet/asw040

[8] DELAIGLE, A., HALL, P., HUANG, W. and KNEIP, A. (2021). Estimating the covariance of fragmented and other related types of functional data. *J. Amer. Statist. Assoc.* **116** 1383–1401. MR4309280 https://doi.org/10.1080/01621459.2020.1723597

[9] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175. MR3931974

[10] DESCARY, M.-H. and PANARETOS, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145–160. MR3912388 https://doi.org/10.1093/biomet/asy055

[11] DESCARY, M.-H. and PANARETOS, V. M. (2019). Functional data analysis by matrix completion. *Ann. Statist.* **47** 1–38. MR3909925 https://doi.org/10.1214/17-AOS1590

MSC2020 subject classifications. Primary 62M20, 62H22, 62G05; secondary 47A57, 15A83, 45Q05.

Key words and phrases. Positive-definite continuation, functional data analysis, graphical model, identifiability, inverse problem, fragments.

[12] DYM, H. and GOHBERG, I. (1981). Extensions of band matrices with band inverses. *Linear Algebra Appl.* **36** 1–24. MR0604325 https://doi.org/10.1016/0024-3795(81)90215-9

[13] GNEITING, T. and SASVÁRI, Z. (1999). The characterization problem for isotropic covariance functions. *Math. Geol.* **31** 105–111. MR1671159

[14] GOHBERG, I., KAASHOEK, M. A. and WOERDEMAN, H. J. (1989). The band method for positive and contractive extension problems. *J. Operator Theory* **22** 109–155. MR1026078

[15] 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

[16] JOHNSON, C. R., ed. (1990). *Matrix Theory and Applications. Proceedings of Symposia in Applied Mathematics* **40**. Amer. Math. Soc., Providence, RI. MR1059481 https://doi.org/10.1090/psapm/040

[17] JOUZDANI, N. M. and PANARETOS, V. M. (2021). Functional Data Analysis with Rough Sampled Paths? arXiv preprint. Available at arXiv:2105.12035.

[18] 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

[19] KRAUS, D. (2015). Components and completion of partially observed functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 777–801. MR3382597 https://doi.org/10.1111/rssb.12087

[20] KREIN, M. (1940). Sur le problème du prolongement des fonctions hermitiennes positives et continues. *C. R. (Dokl.) Acad. Sci. URSS* **26** 17–22. MR0004333

[21] KREIN, M. G. and LANGER, H. (2014). Continuation of Hermitian positive definite functions and related questions. *Integral Equations Operator Theory* **78** 1–69. MR3147401 https://doi.org/10.1007/s00020-013-2091-z

[22] LIN, Z., WANG, J.-L. and ZHONG, Q. (2021). Basis expansions for functional snippets. *Biometrika* **108** 709–726. MR4298773 https://doi.org/10.1093/biomet/asaa088

[23] LOÈVE, M. (2017). *Probability Theory*, 3rd ed. Dover, New York.

[24] PAULSEN, V. I., POWER, S. C. and SMITH, R. R. (1989). Schur products and matrix completions. *J. Funct. Anal.* **85** 151–178. MR1005860 https://doi.org/10.1016/0022-1236(89)90050-5

[25] RUDIN, W. (1963). The extension problem for positive-definite functions. *Illinois J. Math.* **7** 532–539. MR0151796

[26] SAITOH, S. and SAWANO, Y. (2016). *Theory of Reproducing Kernels and Applications. Developments in Mathematics* **44**. Springer, Singapore. MR3560890 https://doi.org/10.1007/978-981-10-0530-5

[27] SASVÁRI, Z. (2006). The extension problem for positive definite functions. A short historical survey. In *Operator Theory and Indefinite Inner Product Spaces. Oper. Theory Adv. Appl.* **163** 365–379. Birkhäuser, Basel. MR2215871 https://doi.org/10.1007/3-7643-7516-7_16

[28] STEWART, J. (1976). Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.* **6** 409–434. MR0430674 https://doi.org/10.1216/RMJ-1976-6-3-409

[29] R CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[30] WAGHMARE, K. (2022). covcomp: Covariance Completion. R package. https://github.com/kgwstat/covcomp.

[31] WAGHMARE, K. G and PANARETOS, V. M (2022). Supplement to "The Completion of Covariance Kernels." https://doi.org/10.1214/22-AOS2228SUPP

The Annals of Statistics
2022, Vol. 50, No. 6, 3307–3333
https://doi.org/10.1214/22-AOS2229

LIKELIHOOD ESTIMATION OF SPARSE TOPIC DISTRIBUTIONS IN TOPIC MODELS AND ITS APPLICATIONS TO WASSERSTEIN DOCUMENT DISTANCE CALCULATIONS

BY XIN BING[1,a], FLORENTINA BUNEA[2,b], SETH STRIMAS-MACKEY[2,c] AND MARTEN WEGKAMP[3,d]

[1]*Department of Statistical Sciences, University of Toronto,* [a]*xin.bing@utoronto.ca*
[2]*Department of Statistics and Data Science, Cornell University,* [b]*fb238@cornell.edu,* [c]*scs324@cornell.edu*
[3]*Departments of Mathematics, and of Statistics and Data Science, Cornell University,* [d]*mhw73@cornell.edu*

This paper studies the estimation of high-dimensional, discrete, possibly sparse, mixture models in the context of topic models. The data consists of observed multinomial counts of p words across n independent documents. In topic models, the $p \times n$ expected word frequency matrix is assumed to be factorized as a $p \times K$ word-topic matrix A and a $K \times n$ topic-document matrix T. Since columns of both matrices represent conditional probabilities belonging to probability simplices, columns of A are viewed as p-dimensional mixture components that are common to all documents while columns of T are viewed as the K-dimensional mixture weights that are document specific and are allowed to be sparse.

The main interest is to provide sharp, finite sample, ℓ_1-norm convergence rates for estimators of the mixture weights T when A is either known or unknown. For known A, we suggest MLE estimation of T. Our nonstandard analysis of the MLE not only establishes its ℓ_1 convergence rate, but also reveals a remarkable property: the MLE, with no extra regularization, can be exactly sparse and contain the true zero pattern of T. We further show that the MLE is both minimax optimal and adaptive to the unknown sparsity in a large class of sparse topic distributions. When A is unknown, we estimate T by optimizing the likelihood function corresponding to a plug in, generic, estimator $\widehat{A}$ of A. For any estimator $\widehat{A}$ that satisfies carefully detailed conditions for proximity to A, we show that the resulting estimator of T retains the properties established for the MLE. Our theoretical results allow the ambient dimensions K and p to grow with the sample sizes.

Our main application is to the estimation of 1-Wasserstein distances between document generating distributions. We propose, estimate and analyze new 1-Wasserstein distances between alternative probabilistic document representations, at the word and topic level, respectively. We derive finite sample bounds on the estimated proposed 1-Wasserstein distances. For word level document-distances, we provide contrast with existing rates on the 1-Wasserstein distance between standard empirical frequency estimates. The effectiveness of the proposed 1-Wasserstein distances is illustrated by an analysis of an IMDB movie reviews data set. Finally, our theoretical results are supported by extensive simulation studies.

REFERENCES

[1] AGRESTI, A. (2013). *Categorical Data Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR3087436

[2] ANANDKUMAR, A., FOSTER, D. P., HSU, D. J., KAKADE, S. M. and LIU, Y.-K. (2012). A spectral algorithm for latent Dirichlet allocation. In *Advances in Neural Information Processing Systems* 25

MSC2020 subject classifications. Primary 62H12, 62H30; secondary 62F10.

Key words and phrases. Adaptive estimation, high-dimensional estimation, maximum likelihood estimation, minimax estimation, multinomial distribution, mixture model, sparse estimation, nonnegative matrix factorization, topic models, anchor words.

(F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 917–925. Curran Associates, Red Hook.

[3] ARORA, S., GE, R., HALPERN, Y., MIMNO, D. M., MOITRA, A., SONTAG, D., WU, Y. and ZHU, M. (2013). A practical algorithm for topic modeling with provable guarantees. In *ICML* (2) 280–288.

[4] ARORA, S., GE, R., KOEHLER, F., MA, T. and MOITRA, A. (2016). Provable algorithms for inference in topic models. In *Proceedings of the* 33*rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 2859–2867. PMLR, New York, New York, USA.

[5] ARORA, S., GE, R. and MOITRA, A. (2012). Learning topic models—going beyond SVD. In 2012 *IEEE* 53*rd Annual Symposium on Foundations of Computer Science—FOCS* 2012 1–10. IEEE Computer Soc., Los Alamitos, CA. MR3185945

[6] BANSAL, T., BHATTACHARYYA, C. and KANNAN, R. (2014). A provable SVD-based algorithm for learning topics in dominant admixture corpus. *Adv. Neural Inf. Process. Syst.* **27**.

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

[8] BING, X., BUNEA, F., STRIMAS-MACKEY, S. and WEGKAMP, M. (2022). Supplement to "Likelihood estimation of sparse topic distributions in topic models and its applications to Wasserstein document distance calculations." https://doi.org/10.1214/22-AOS2229SUPP

[9] BING, X., BUNEA, F. and WEGKAMP, M. (2020). A fast algorithm with minimax optimal guarantees for topic models with an unknown number of topics. *Bernoulli* **26** 1765–1796. MR4091091 https://doi.org/10.3150/19-BEJ1166

[10] BING, X., BUNEA, F. and WEGKAMP, M. (2020). Optimal estimation of sparse topic models. *J. Mach. Learn. Res.* **21** 177. MR4209463

[11] BIRCH, M. W. (1964). A new proof of the Pearson–Fisher theorem. *Ann. Math. Stat.* **35** 817–824. MR0169324 https://doi.org/10.1214/aoms/1177703581

[12] BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer, New York. With the collaboration of Richard J. Light and Frederick Mosteller, Reprint of the 1975 original. MR2344876

[13] BITTORF, V., RECHT, B., RE, C. and TROPP, J. A. (2012). Factoring nonnegative matrices with linear programs. Available at arXiv:1206.1270.

[14] BLEI, D. M. (2012). Introduction to probabilistic topic models. *Commun. ACM* **55** 77–84.

[15] BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* 993–1022.

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

[17] CAO, Y., ZHANG, A. and LI, H. (2020). Multisample estimation of bacterial composition matrices in metagenomics data. *Biometrika* **107** 75–92. MR4064141 https://doi.org/10.1093/biomet/asz062

[18] CHEN, S., RIVAUD, P., PARK, J. H., TSOU, T., CHARLES, E., HALIBURTON, J. R., PICHIORRI, F. and THOMSON, M. (2020). Dissecting heterogeneous cell populations across drug and disease conditions with PopAlign. *Proc. Natl. Acad. Sci. USA* **117** 28784–28794.

[19] DING, W., ISHWAR, P. and SALIGRAMA, V. (2015). Most large topic models are approximately separable. In 2015 *Information Theory and Applications Workshop* (*ITA*) 199–203.

[20] DING, W., ROHBAN, M. H., ISHWAR, P. and SALIGRAMA, V. (2013). Topic discovery through data dependent and random projections. In *Proceedings of the* 30*th International Conference on Machine Learning* (S. Dasgupta and D. McAllester, eds.). *Proceedings of Machine Learning Research* **28** 1202–1210. PMLR, Atlanta, GA, USA.

[21] GIBBS, A. L. and SU, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** 419–435.

[22] GONZÁLEZ-BLAS, C. B., MINNOYE, L., PAPASOKRATI, D., AIBAR, S., HULSELMANS, G., CHRISTIAENS, V., DAVIE, K., WOUTERS, J. and AERTS, S. (2019). cisTopic: Cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16** 397–400. https://doi.org/10.1038/s41592-019-0367-1

[23] GRIFFITHS, T. L. and STEYVERS, M. (2004). Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **101** 5228–5235.

[24] KE, T. Z. and WANG, M. (2017). A new SVD approach to optimal topic estimation. Available at arXiv:1704.07016.

[25] KLEINBERG, J. and SANDLER, M. (2008). Using mixture models for collaborative filtering. *J. Comput. System Sci.* **74** 49–69. MR2364181 https://doi.org/10.1016/j.jcss.2007.04.013

[26] KLOPP, O., PANOV, M., SIGALLA, S. and TSYBAKOV, A. (2021). Assigning topics to documents by successive projections. ArXiv preprint. Available at arXiv:2107.03684.

[27] KUSNER, M., KOLKIN, Y. S. N. I. and WEINBERGER, K. Q. (2015). From word embeddings to document distances. http://proceedings.mlr.press/v37/kusnerb15.pdf.

[28] LE, Q. V. and MIKOLOV, T. (2014). Distributed Representations of Sentences and Documents.

[29] MA, W.-K., BIOUCAS-DIAS, J. M., CHAN, T.-H., GILLIS, N., GADER, P., PLAZA, A. J., AMBIKAPATHI, A. and CHI, C.-Y. (2013). A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.* **31** 67–81.

[30] MIKOLOV, T., CHEN, K., CORRADO, G. and DEAN, J. (2013). Efficient Estimation of Word Representations in Vector Space.

[31] QIU, X., SUN, T., XU, Y., SHAO, Y., DAI, N. and HUANG, X. (2020). Pre-trained Models for Natural Language Processing: A Survey.

[32] RAO, C. R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankhyā* **18** 139–148. MR0105183

[33] RAO, C. R. (1958). Maximum likelihood estimation for the multinomial distribution with infinite number of cells. *Sankhyā* **20** 211–218. MR0107334

[34] REIMERS, N. and GUREVYCH, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.

[35] SOMMERFELD, M. and MUNK, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238. MR3744719 https://doi.org/10.1111/rssb.12236

[36] TAMELING, C., SOMMERFELD, M. and MUNK, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.* **29** 2744–2781. MR4019874 https://doi.org/10.1214/19-AAP1463

[37] WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

[38] ZHU, Z., LI, X., WANG, M. and ZHANG, A. (2021). Learning Markov models via low-rank optimization. *Oper. Res.*

The Annals of Statistics
2022, Vol. 50, No. 6, 3334–3363
https://doi.org/10.1214/22-AOS2230

BAYESIAN FIXED-DOMAIN ASYMPTOTICS FOR COVARIANCE PARAMETERS IN A GAUSSIAN PROCESS MODEL

BY CHENG LI[a]

Department of Statistics and Data Science, National University of Singapore, [a]*stalic@nus.edu.sg*

Gaussian process models typically contain finite-dimensional parameters in the covariance function that need to be estimated from the data. We study the Bayesian fixed-domain asymptotics for the covariance parameters in a universal kriging model with an isotropic Matérn covariance function, which has many applications in spatial statistics. We show that when the dimension of domain is less than or equal to three, the joint posterior distribution of the microergodic parameter and the range parameter can be factored independently into the product of their marginal posteriors under fixed-domain asymptotics. The posterior of the microergodic parameter is asymptotically close in total variation distance to a normal distribution with shrinking variance, while the posterior distribution of the range parameter does not converge to any point mass distribution in general. Our theory allows an unbounded prior support for the range parameter and flexible designs of sampling points. We further study the asymptotic efficiency and convergence rates in posterior prediction for the Bayesian kriging predictor with covariance parameters randomly drawn from their posterior distribution. In the special case of one-dimensional Ornstein–Uhlenbeck process, we derive explicitly the limiting posterior of the range parameter and the posterior convergence rate for asymptotic efficiency in posterior prediction. We verify these asymptotic results in numerical experiments.

REFERENCES

[1] ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700 https://doi.org/10.1214/09-AOS725

[2] ARNOLD, B. C. (2015). *Pareto Distributions*, 2nd ed. *Monographs on Statistics and Applied Probability* **140**. CRC Press, Boca Raton, FL. MR3618736

[3] BACHOC, F., BEVILACQUA, M. and VELANDIA, D. (2019). Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. *J. Multivariate Anal.* **174** 104534. MR3985049 https://doi.org/10.1016/j.jmva.2019.104534

[4] BACHOC, F. and LAGNOUX, A. (2020). Fixed-domain asymptotic properties of maximum composite likelihood estimators for Gaussian processes. *J. Statist. Plann. Inference* **209** 62–75. MR4096254 https://doi.org/10.1016/j.jspi.2020.02.008

[5] BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 https://doi.org/10.1111/j.1467-9868.2008.00663.x

[6] BERGER, J. O., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* **96** 1361–1374. MR1946582 https://doi.org/10.1198/016214501753382282

[7] BEVILACQUA, M., FAOUZI, T., FURRER, R. and PORCU, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Ann. Statist.* **47** 828–856. MR3909952 https://doi.org/10.1214/17-AOS1652

[8] BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *Ann. Statist.* **40** 206–237. MR3013185 https://doi.org/10.1214/11-AOS921

[9] BOCHKINA, N. A. and GREEN, P. J. (2014). The Bernstein–von Mises theorem and nonregular models. *Ann. Statist.* **42** 1850–1878. MR3262470 https://doi.org/10.1214/14-AOS1239

MSC2020 subject classifications. 62F15, 62H11, 62E20.

Key words and phrases. Fixed-domain asymptotics, limiting posterior distribution, Matérn covariance function, asymptotic efficiency in posterior prediction.

[10] BRAZAUSKAS, V. (2002). Fisher information matrix for the Feller–Pareto distribution. *Statist. Probab. Lett.* **59** 159–167. MR1927527 https://doi.org/10.1016/S0167-7152(02)00143-8

[11] CHAE, M. and WALKER, S. G. (2020). Wasserstein upper bounds of the total variation for smooth densities. *Statist. Probab. Lett.* **163** 108771. MR4083852 https://doi.org/10.1016/j.spl.2020.108771

[12] CHANG, C.-H., HUANG, H.-C. and ING, C.-K. (2014). Asymptotic theory of generalized information criterion for geostatistical regression model selection. *Ann. Statist.* **42** 2441–2468. MR3269985 https://doi.org/10.1214/14-AOS1258

[13] CHANG, C.-H., HUANG, H.-C. and ING, C.-K. (2017). Mixed domain asymptotics for a stochastic process model with time trend and measurement error. *Bernoulli* **23** 159–190. MR3556770 https://doi.org/10.3150/15-BEJ740

[14] CHEN, H.-S., SIMPSON, D. G. and YING, Z. (2000). Infill asymptotics for a stochastic process model with measurement error. *Statist. Sinica* **10** 141–156. MR1742105

[15] CHEN, X., CHRISTENSEN, T. M. and TAMER, E. (2018). Monte Carlo confidence sets for identified sets. *Econometrica* **86** 1965–2018. MR3901244 https://doi.org/10.3982/ECTA14525

[16] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489 https://doi.org/10.1111/j.1468-0262.2004.00540.x

[17] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

[18] CROWDER, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. Roy. Statist. Soc. Ser. B* **38** 45–53. MR0403035 https://doi.org/10.1111/j.2517-6161.1976.tb01565.x

[19] DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706 https://doi.org/10.1080/01621459.2015.1044091

[20] DE OLIVEIRA, V., KEDEM, B. and SHORT, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *J. Amer. Statist. Assoc.* **92** 1422–1433. MR1615252 https://doi.org/10.2307/2965412

[21] DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. MR2549562 https://doi.org/10.1214/08-AOS676

[22] FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *J. Amer. Statist. Assoc.* **114** 445–452. MR3941267 https://doi.org/10.1080/01621459.2017.1415907

[23] 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

[24] GNEITING, T. (2002). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508. MR1945966 https://doi.org/10.1006/jmva.2001.2056

[25] GU, M. and ANDERSON, K. (2018). Calibration of imperfect mathematical models by multiple sources of data with measurement bias. ArXiv Preprint. Available at arXiv:1810.11664.

[26] GU, M., WANG, X. and BERGER, J. O. (2018). Robust Gaussian stochastic process emulation. *Ann. Statist.* **46** 3038–3066. MR3851764 https://doi.org/10.1214/17-AOS1648

[27] GUHANIYOGI, R., LI, C., SAVITSKY, T. D. and SRIVASTAVA, S. (2022). Distributed Bayesian inference in massive spatial data. *Statist. Sci.* To appear.

[28] GUSTAFSON, P. (2014). Bayesian inference in partially identified models: Is the shape of the posterior distribution useful? *Electron. J. Stat.* **8** 476–496. MR3205730 https://doi.org/10.1214/14-EJS891

[29] GUSTAFSON, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data. Monographs on Statistics and Applied Probability* **141**. CRC Press, Boca Raton, FL. MR3642458

[30] HANDCOCK, M. S. and STEIN, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35** 403–410. https://doi.org/10.1080/00401706.1993.10485354

[31] HEATON, J. H., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–825. https://doi.org/10.1007/s13253-018-00348-w

[32] JIANG, W. (2017). On limiting distribution of quasi-posteriors under partial identification. *Econom. Stat.* **3** 60–72. MR3666242 https://doi.org/10.1016/j.ecosta.2017.03.006

[33] JIANG, W. and LI, C. (2019). On Bayesian oracle properties. *Bayesian Anal.* **14** 235–260. MR3910045 https://doi.org/10.1214/18-BA1097

[34] JUN, S. J., PINKSE, J. and WAN, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *J. Econometrics* **187** 201–216. MR3347303 https://doi.org/10.1016/j.jeconom.2015.01.005

[35] KANAGAWA, M., HENNIG, P., SEJDINOVIC, D. and SRIPERUMBUDUR, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. ArXiv Preprint. Available at arXiv:1807.02582.

[36] KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 https://doi.org/10.1198/016214508000000959

[37] KAUFMAN, C. G. and SHABY, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100** 473–484. MR3068447 https://doi.org/10.1093/biomet/ass079

[38] KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 https://doi.org/10.1111/1467-9868.00294

[39] KLEIJN, B. J. K. and KNAPIK, B. (2012). Semiparametric posterior limits under local asymptotic exponentiality. ArXiv Preprint. Available at arXiv:1210.6204v3.

[40] KREH, M. (2012). *Bessel Functions*. *Lecture Notes*, *Penn State—Göttingen Summer School on Number Theory*.

[41] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875 https://doi.org/10.1007/b98854

[42] LI, C. (2022). Supplement to "Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model." https://doi.org/10.1214/22-AOS2230SUPPA, https://doi.org/10.1214/22-AOS2230SUPPB

[43] LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. MR3694589 https://doi.org/10.1093/biomet/asx033

[44] LOH, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *Ann. Statist.* **33** 2344–2394. MR2211089 https://doi.org/10.1214/009053605000000516

[45] LOH, W.-L. (2015). Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.* **43** 2766–2794. MR3405611 https://doi.org/10.1214/15-AOS1365

[46] LOH, W.-L., SUN, S. and WEN, J. (2021). On fixed-domain asymptotics, parameter estimation and isotropic Gaussian random fields with Matérn covariance functions. *Ann. Statist.* **49** 3127–3152. MR4352525 https://doi.org/10.1214/21-aos2077

[47] MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. *Springer Series in Statistics*. Springer, New York. MR2151380 https://doi.org/10.1007/b97478

[48] MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. MR0738334 https://doi.org/10.1093/biomet/71.1.135

[49] MOON, H. R. and SCHORFHEIDE, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica* **80** 755–782. MR2951948 https://doi.org/10.3982/ECTA8360

[50] PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *J. Amer. Statist. Assoc.* **117** 969–982. MR4436326 https://doi.org/10.1080/01621459.2020.1833889

[51] PUTTER, H. and YOUNG, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli* **7** 421–438. MR1836738 https://doi.org/10.2307/3318494

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

[53] RITTER, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Springer, Berlin. MR1763973 https://doi.org/10.1007/BFb0103934

[54] SANG, H. and HUANG, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 111–132. MR2885842 https://doi.org/10.1111/j.1467-9868.2011.01007.x

[55] SANG, H., JUN, M. and HUANG, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *Ann. Appl. Stat.* **5** 2519–2548. MR2907125 https://doi.org/10.1214/11-AOAS478

[56] SHABY, B. and RUPPERT, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *J. Comput. Graph. Statist.* **21** 433–452. MR2945475 https://doi.org/10.1080/10618600.2012.680819

[57] SHEN, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97** 222–235. MR1947282 https://doi.org/10.1198/016214502753479365

[58] STEIN, M. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872. MR1056340 https://doi.org/10.1214/aos/1176347629

[59] STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856 https://doi.org/10.1214/aos/1176350690

[60] STEIN, M. L. (1990). Bounds on the efficiency of linear predictions using an incorrect covariance function. *Ann. Statist.* **18** 1116–1138. MR1062701 https://doi.org/10.1214/aos/1176347742

[61] STEIN, M. L. (1990). A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process. *Ann. Statist.* **18** 1139–1157. MR1062702 https://doi.org/10.1214/aos/1176347743

[62] STEIN, M. L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statist. Probab. Lett.* **17** 399–404. MR1237787 https://doi.org/10.1016/0167-7152(93)90261-G

[63] STEIN, M. L. (1997). Efficiency of linear predictors for periodic processes using an incorrect covariance function. *J. Statist. Plann. Inference* **58** 321–331. MR1450019 https://doi.org/10.1016/S0378-3758(96)00088-2

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

[65] STEIN, M. L. (1999). Predicting random fields with increasing dense observations. *Ann. Appl. Probab.* **9** 242–273. MR1682572 https://doi.org/10.1214/aoap/1029962604

[66] SUN, Q., MIAO, C., DUAN, Q., ASHOURI, H., SOROOSHIAN, S. and HSU, K. L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.* **56** 79–107. https://doi.org/10.1002/2017RG000574

[67] TAMER, E. (2010). Partial identification in econometrics. *Ann. Rev. Econ.* **3** 167–195. https://doi.org/10.1146/annurev.economics.050708.143401

[68] TANG, W., ZHANG, L. and BANERJEE, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 1044–1070. MR4349127 https://doi.org/10.1111/rssb.12472

[69] TUO, R. and WANG, W. (2020). Kriging prediction with isotropic Matérn correlations: Robustness and experimental designs. *J. Mach. Learn. Res.* **21** Paper No. 187. MR4209473

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

[71] 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

[72] 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

[73] 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

[74] VELANDIA, D., BACHOC, F., BEVILACQUA, M., GENDRE, X. and LOUBES, J.-M. (2017). Maximum likelihood estimation for a bivariate Gaussian process under fixed domain asymptotics. *Electron. J. Stat.* **11** 2978–3007. MR3694574 https://doi.org/10.1214/17-EJS1298

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

[76] WANG, D. and LOH, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Stat.* **5** 238–269. MR2792553 https://doi.org/10.1214/11-EJS607

[77] WANG, W., TUO, R. and WU, C. F. J. (2020). On prediction properties of kriging: Uniform error bounds and robustness. *J. Amer. Statist. Assoc.* **115** 920–930. MR4107689 https://doi.org/10.1080/01621459.2019.1598868

[78] WENDLAND, H. (2005). *Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724 https://doi.org/10.1017/CBO9780511617539

[79] WU, Z. M. and SCHABACK, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.* **13** 13–27. MR1199027 https://doi.org/10.1093/imanum/13.1.13

[80] WYNNE, G., BRIOL, F. X. and GIROLAMI, M. (2021). Convergence guarantees for Gaussian process approximations under several observation models. *J. Mach. Learn. Res.* **123** 1–40.

[81] YAKOWITZ, S. J. and SZIDAROVSZKY, F. (1985). A comparison of kriging with nonparametric regression methods. *J. Multivariate Anal.* **16** 21–53. MR0778488 https://doi.org/10.1016/0047-259X(85)90050-8

[82] 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

[83] YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36** 280–296. MR1096671 https://doi.org/10.1016/0047-259X(91)90062-7

[84] YING, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21** 1567–1590. MR1241279 https://doi.org/10.1214/aos/1176349272

[85] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

[86] ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195 https://doi.org/10.1093/biomet/92.4.921

The Annals of Statistics
2022, Vol. 50, No. 6, 3364–3387
https://doi.org/10.1214/22-AOS2231

BATCH POLICY LEARNING IN AVERAGE REWARD MARKOV DECISION PROCESSES

BY PENG LIAO[1,a], ZHENGLING QI[2,c], RUNZHE WAN[3,d], PREDRAG KLASNJA[4,e] AND SUSAN A. MURPHY[1,b]

[1]*Department of Statistics, Harvard University,* [a]*pliao9122@gmail.com,* [b]*samurphy11@gmail.com*
[2]*Department of Decision Sciences, George Washington University,* [c]*qizhengling@gwu.edu*
[3]*Amazon,* [d]*runzhe.wan@gmail.com*
[4]*School of Information, University of Michigan, Ann Arbor,* [e]*klasnja@umich.edu*

We consider the batch (off-line) policy learning problem in the infinite horizon Markov decision process. Motivated by mobile health applications, we focus on learning a policy that maximizes the long-term average reward. We propose a doubly robust estimator for the average reward and show that it achieves semiparametric efficiency. Further, we develop an optimization algorithm to compute the optimal policy in a parameterized stochastic policy class. The performance of the estimated policy is measured by the difference between the optimal average reward in the policy class and the average reward of the estimated policy and we establish a finite-sample regret guarantee. The performance of the method is illustrated by simulation studies and an analysis of a mobile health study promoting physical activity.

REFERENCES

ABOUNADI, J., BERTSEKAS, D. and BORKAR, V. S. (2001). Learning algorithms for Markov decision processes with average cost. *SIAM J. Control Optim.* **40** 681–698. MR1871450 https://doi.org/10.1137/S0363012999361974

AGARWAL, R., SCHUURMANS, D. and NOROUZI, M. (2020). An optimistic perspective on offline reinforcement learning. In *International Conference on Machine Learning* 104–114. PMLR.

ANTOS, A., SZEPESVÁRI, C. and MUNOS, R. (2008a). Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path. *Mach. Learn.* **71** 89–129. https://doi.org/10.1007/s10994-007-5038-2

ANTOS, A., SZEPESVÁRI, C. and MUNOS, R. (2008b). Fitted Q-iteration in continuous action-space MDPs. In *Advances in Neural Information Processing Systems* 9–16.

ATHEY, S. and WAGER, S. (2017). Efficient policy learning. arXiv preprint arXiv:1702.02896.

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

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

DUDÍK, M., ERHAN, D., LANGFORD, J. and LI, L. (2014). Doubly robust policy evaluation and optimization. *Statist. Sci.* **29** 485–511. MR3300356 https://doi.org/10.1214/14-STS500

ERNST, D., GEURTS, P. and WEHENKEL, L. (2005). Tree-based batch mode reinforcement learning. *J. Mach. Learn. Res.* **6** 503–556. MR2249830

ERTEFAIE, A. and STRAWDERMAN, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika* **105** 963–977. MR3877877 https://doi.org/10.1093/biomet/asy043

FARAHMAND, A. and SZEPESVÁRI, C. (2011). Model selection in reinforcement learning. *Mach. Learn.* **85** 299–332. MR3108236 https://doi.org/10.1007/s10994-011-5254-7

FARAHMAND, A., GHAVAMZADEH, M., SZEPESVÁRI, C. and MANNOR, S. (2016). Regularized policy iteration with nonparametric function spaces. *J. Mach. Learn. Res.* **17** Paper No. 139, 66. MR3555030

MSC2020 subject classifications. 62G05.
Key words and phrases. Markov decision process, average reward, policy optimization, doubly robust estimator.

FUJIMOTO, S., MEGER, D. and PRECUP, D. (2019). Off-policy deep reinforcement learning without exploration. In *International Conference on Machine Learning* 2052–2062. PMLR.

FUKUMIZU, K., GRETTON, A., LANCKRIET, G. R., SCHÖLKOPF, B. and SRIPERUMBUDUR, B. K. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Advances in Neural Information Processing Systems* 1750–1758.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2006). *A Distribution-Free Theory of Nonparametric Regression*. Springer Science & Business Media. MR1920390 https://doi.org/10.1007/b97848

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning* **1**. *Springer Series in Statistics New York*. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

HERNÁNDEZ-LERMA, O. and LASSERRE, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. *Applications of Mathematics* (*New York*) **42**. Springer, New York. MR1697198 https://doi.org/10.1007/978-1-4612-0561-6

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

KAKADE, S. and LANGFORD, J. (2002). Approximately optimal approximate reinforcement learning. In *Proc*. 19*th International Conference on Machine Learning*. Citeseer.

KALLUS, N. and UEHARA, M. (2019). Efficiently breaking the curse of horizon: Double reinforcement learning in infinite-horizon processes. arXiv preprint arXiv:1909.05850.

KALLUS, N. and UEHARA, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** Paper No. 167, 63. MR4209453

KLASNJA, P., HEKLER, E. B., SHIFFMAN, S., BORUVKA, A., ALMIRALL, D., TEWARI, A. and MURPHY, S. A. (2015). Micro-randomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* **34** 1220.

KLASNJA, P., SMITH, S., SEEWALD, N. J., LEE, A., HALL, K., LUERS, B., HEKLER, E. B. and MURPHY, S. A. (2018). Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Ann. Behav. Med.*

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

KUMAR, A., FU, J., SOH, M., TUCKER, G. and LEVINE, S. (2019). Stabilizing off-policy q-learning via bootstrapping error reduction. *Adv. Neural Inf. Process. Syst.* **32**.

LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **8** 1225–1272. MR3263118 https://doi.org/10.1214/14-EJS920

LAGOUDAKIS, M. G. and PARR, R. (2004). Least-squares policy iteration. *J. Mach. Learn. Res.* **4** 1107–1149. MR2125347 https://doi.org/10.1162/1532443041827907

LIAO, P., KLASNJA, P. and MURPHY, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *J. Amer. Statist. Assoc.* **116** 382–391. MR4227701 https://doi.org/10.1080/01621459.2020.1807993

LIAO, P., KLASNJA, P., TEWARI, A. and MURPHY, S. A. (2016). Sample size calculations for micro-randomized trials in mHealth. *Stat. Med.* **35** 1944–1971. MR3513494 https://doi.org/10.1002/sim.6847

LIAO, P., QI, Z., WAN, R., KLASNJA, P. and MURPHY, S. A. (2022). Supplement to "Batch policy learning in average reward Markov decision processes." https://doi.org/10.1214/22-AOS2231SUPP

LIU, D. C. and NOCEDAL, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* **45** 503–528. MR1038245 https://doi.org/10.1007/BF01589116

LIU, Q., LI, L., TANG, Z. and ZHOU, D. (2018). Breaking the curse of horizon: Infinite-horizon off-policy estimation. In *Advances in Neural Information Processing Systems* 5356–5366.

LIU, Y., SWAMINATHAN, A., AGARWAL, A. and BRUNSKILL, E. (2019). Off-policy policy gradient with state distribution correction. arXiv preprint arXiv:1904.08473.

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403 https://doi.org/10.1214/16-AOS1471

LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919

MAHADEVAN, S. (1996). Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Mach. Learn.* **22** 159–195.

MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

MUNOS, R. and SZEPESVÁRI, C. (2008). Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **9** 815–857. MR2417255

MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586 https://doi.org/10.1198/016214501753382327

MURPHY, S. A., DENG, Y., LABER, E. B., MAEI, H. R., SUTTON, R. S. and WITKIEWITZ, K. (2016). A batch, off-policy, actor-critic algorithm for optimizing the average reward. arXiv preprint arXiv:1607.05047.

NACHUM, O., CHOW, Y., DAI, B. and LI, L. (2019). Dualdice: Behavior-agnostic estimation of discounted stationary distribution corrections. In *Advances in Neural Information Processing Systems* 2315–2325.

NAHUM-SHANI, I., SMITH, S. N., SPRING, B. J., COLLINS, L. M., WITKIEWITZ, K., TEWARI, A. and MURPHY, S. A. (2016). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* 1–17.

NAIK, A., SHARIFF, R., YASUI, N. and SUTTON, R. S. (2019). Discounted reinforcement learning is not an optimization problem. arXiv preprint arXiv:1910.02140.

NEWEY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5** 99–135.

ORMONEIT, D. and SEN, S. (2003). Kernel-based reinforcement learning. In *Machine Learning* 161–178.

PRECUP, D. (2000). Eligibility traces for off-policy policy evaluation. Computer Science Department Faculty Publication Series 80.

PUTERMAN, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR1270015

RICHARDSON, G. B. (1995). The theory of the market economy. *Revue* 1487–1496.

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

SHARMA, H., JAFARNIA-JAHROMI, M. and JAIN, R. (2020). Approximate relative value learning for average-reward continuous state MDPs. In *Uncertainty in Artificial Intelligence* 956–964. PMLR.

SHI, C., WAN, R., CHERNOZHUKOV, V. and SONG, R. (2021). Deeply-debiased off-policy interval estimation. arXiv preprint arXiv:2105.04646.

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

SNOEK, J., LAROCHELLE, H. and ADAMS, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* **25**.

STEINWART, I. and CHRISTMANN, A. (2008). *Support Vector Machines*. Springer Science & Business Media.

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

TANG, Z., FENG, Y., LI, L., ZHOU, D. and LIU, Q. (2020). Doubly robust bias reduction in infinite horizon off-policy estimation. In *International Conference on Learning Representations*.

THOMAS, P. and BRUNSKILL, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning* 2139–2148.

UEHARA, M. and JIANG, N. (2019). Minimax weight and q-function learning for off-policy evaluation. arXiv preprint arXiv:1910.12809.

VAN ROY, B. (1998). *Learning and Value Function Approximation in Complex Decision Processes*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Massachusetts Institute of Technology. MR2716807

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

VOLOSHIN, C., LE, H. M., JIANG, N. and YUE, Y. (2019). Empirical study of off-policy policy evaluation for reinforcement learning. arXiv preprint arXiv:1911.06854.

WAN, Y., NAIK, A. and SUTTON, R. S. (2021). Learning and planning in average-reward Markov decision processes. In *International Conference on Machine Learning* 10653–10662. PMLR.

WU, Y. and WANG, L. (2021). Resampling-based confidence intervals for model-free robust inference on optimal treatment regimes. *Biometrics* **77** 465–476. MR4307648 https://doi.org/10.1111/biom.13337

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007 https://doi.org/10.1111/j.1541-0420.2012.01763.x

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445 https://doi.org/10.1093/biomet/ast014

ZHANG, R., DAI, B., LI, L. and SCHUURMANS, D. (2020). Gen{DICE}: Generalized offline estimation of stationary values. In *International Conference on Learning Representations*.

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488

ZHAO, Y.-Q., LABER, E. B., NING, Y., SAHA, S. and SANDS, B. E. (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *J. Mach. Learn. Res.* **20** Paper No. 48, 23. MR3948088

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

The Annals of Statistics
2022, Vol. 50, No. 6, 3388–3414
https://doi.org/10.1214/22-AOS2233

LOCAL PERMUTATION TESTS FOR CONDITIONAL INDEPENDENCE

BY ILMUN KIM[1,a], MATEY NEYKOV[2,b], SIVARAMAN BALAKRISHNAN[2,c] AND LARRY WASSERMAN[2,d]

[1]*Department of Statistics and Data Science, Department of Applied Statistics, Yonsei University,* [a]*ilmun@yonsei.ac.kr*
[2]*Department of Statistics and Data Science, Carnegie Mellon University,* [b]*mneykov@stat.cmu.edu,* [c]*siva@stat.cmu.edu,* [d]*larry@stat.cmu.edu*

In this paper, we investigate local permutation tests for testing conditional independence between two random vectors X and Y given Z. The local permutation test determines the significance of a test statistic by locally shuffling samples, which share similar values of the conditioning variables Z, and it forms a natural extension of the usual permutation approach for unconditional independence testing. Despite its simplicity and empirical support, the theoretical underpinnings of the local permutation test remain unclear. Motivated by this gap, this paper aims to establish theoretical foundations of local permutation tests with a particular focus on binning-based statistics. We start by revisiting the hardness of conditional independence testing and provide an upper bound for the power of any valid conditional independence test, which holds when the probability of observing "collisions" in Z is small. This negative result naturally motivates us to impose additional restrictions on the possible distributions under the null and alternate. To this end, we focus our attention on certain classes of smooth distributions and identify provably tight conditions under which the local permutation method is universally valid, that is, it is valid when applied to any (binning-based) test statistic. To complement this result on type I error control, we also show that in some cases, a binning-based statistic calibrated via the local permutation method can achieve minimax optimal power. We also introduce a double-binning permutation strategy, which yields a valid test over less smooth null distributions than the typical single-binning method without compromising much power. Finally, we present simulation results to support our theoretical findings.

REFERENCES

[1] AGRESTI, A. (1992). A survey of exact inference for contingency tables. *Statist. Sci.* **7** 131–177. With comments and a rejoinder by the author. MR1173420

[2] ALBERT, M., LAURENT, B., MARREL, A. and MEYNAOUI, A. (2022). Adaptive test of independence based on HSIC measures. *Ann. Statist.* **50** 858–879. MR4404921 https://doi.org/10.1214/21-aos2129

[3] AZADKIA, M. and CHATTERJEE, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. MR4352523 https://doi.org/10.1214/21-aos2073

[4] BALAKRISHNAN, S. and WASSERMAN, L. (2018). Hypothesis testing for high-dimensional multinomials: A selective review. *Ann. Appl. Stat.* **12** 727–749. MR3834283 https://doi.org/10.1214/18-AOAS1155SF

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

[6] BARBER, R. F. (2020). Is distribution-free inference possible for binary regression? *Electron. J. Stat.* **14** 3487–3524. MR4154845 https://doi.org/10.1214/20-EJS1749

[7] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). The limits of distribution-free conditional predictive inference. *Inf. Inference* **10** 455–482. MR4270755 https://doi.org/10.1093/imaiai/iaaa017

[8] BELLOT, A. and VAN DER SCHAAR, M. (2019). Conditional independence testing using generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **32**.

MSC2020 subject classifications. Primary 62G10, 62C20; secondary 62H20, 62G09.
Key words and phrases. Conditional independence, minimax optimality, permutation test.

[9] BERGSMA, W. (2010). Nonparametric testing of conditional independence by means of the partial copula. Available at SSRN 1702981.

[10] BERGSMA, W. P. (2004). Testing for continuous random variables. Citeseer.

[11] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981

[12] 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

[13] CANONNE, C. L. (2017). A short note on Poisson tail bounds. http://www.cs.columbia.edu/~ccanonne/files/misc/2017-poissonconcentration.pdf.

[14] CANONNE, C. L., DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2018). Testing conditional independence of discrete distributions. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 735–748. ACM, New York. MR3826290 https://doi.org/10.1145/3188745.3188756

[15] DAI, B., SHEN, X. and PAN, W. (2021). Significance tests of feature relevance for a blackbox learner. arXiv preprint arXiv:2103.04985.

[16] DAUDIN, J.-J. (1980). Partial association measures and an application to qualitative regression. *Biometrika* **67** 581–590. MR0601095 https://doi.org/10.1093/biomet/67.3.581

[17] DE CAMPOS, L. M. and HUETE, J. F. (2000). A new approach for learning belief networks using independence criteria. *Internat. J. Approx. Reason.* **24** 11–37. MR1759736 https://doi.org/10.1016/S0888-613X(99)00042-0

[18] DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* 685–694. IEEE Computer Soc., Los Alamitos, CA. MR3631031

[19] DORAN, G., MUANDET, K., ZHANG, K. and SCHÖLKOPF, B. (2014). A permutation-based kernel conditional independence test. In *UAI* 132–141.

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

[21] FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* 489–496.

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

[23] GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

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

[25] HUANG, T.-M. (2010). Testing conditional independence using maximal nonlinear conditional correlation. *Ann. Statist.* **38** 2047–2091. MR2676883 https://doi.org/10.1214/09-AOS770

[26] HUANG, Z., DEB, N. and SEN, B. (2020). Kernel partial correlation coefficient—a measure of conditional dependence. arXiv preprint arXiv:2012.14804.

[27] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

[28] KAMPS, U. (1989). Hellinger distances and α-entropy in a one-parameter class of density functions. *Stat. Hefte* **30** 263–269. MR1024023 https://doi.org/10.1007/BF02924332

[29] KIM, D. and AGRESTI, A. (1997). Nearly exact tests of conditional independence and marginal homogeneity for sparse contingency tables. *Comput. Statist. Data Anal.* **24** 89–104. MR1439566 https://doi.org/10.1016/S0167-9473(96)00038-2

[30] KIM, I. (2021). Comparing a large number of multivariate distributions. *Bernoulli* **27** 419–441. MR4177375 https://doi.org/10.3150/20-BEJ1244

[31] KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Minimax optimality of permutation tests. *Ann. Statist.* **50** 225–251. MR4382015 https://doi.org/10.1214/21-aos2103

[32] KIM, I., NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Supplement to "Local permutation tests for conditional independence." https://doi.org/10.1214/22-AOS2233SUPP

[33] KIM, I., RAMDAS, A., SINGH, A. and WASSERMAN, L. (2021). Classification accuracy as a proxy for two-sample testing. *Ann. Statist.* **49** 411–434. MR4206684 https://doi.org/10.1214/20-AOS1962

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

[35] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

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

[37] LI, C. and FAN, X. (2020). On nonparametric conditional independence tests for continuous variables. *Wiley Interdiscip*. *Rev*.: *Comput*. *Stat*. **12** e1489, 11. MR4098477 https://doi.org/10.1002/wics.1489

[38] LI, L., TCHETGEN TCHETGEN, E., VAN DER VAART, A. and ROBINS, J. M. (2011). Higher order inference on a treatment effect under low regularity conditions. *Statist*. *Probab. Lett*. **81** 821–828. MR2793749 https://doi.org/10.1016/j.spl.2011.02.030

[39] LUNDBORG, A. R., SHAH, R. D. and PETERS, J. (2021). Conditional independence testing in Hilbert spaces with applications to functional data analysis. arXiv preprint arXiv:2101.07108.

[40] MARGARITIS, D. (2005). Distribution-free learning of Bayesian network structure in continuous domains. In *AAAI* **5** 825–830.

[41] MARX, A. and VREEKEN, J. (2019). Testing conditional independence on discrete data using stochastic complexity. In *The* 22*nd International Conference on Artificial Intelligence and Statistics* 496–505. PMLR.

[42] MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found*. *Trends Mach*. *Learn*.

[43] NEWEY, W. K. and ROBINS, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. arXiv preprint arXiv:1801.09138.

[44] NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2021). Minimax optimal conditional independence testing. *Ann*. *Statist*. **49** 2151–2177. MR4319245 https://doi.org/10.1214/20-aos2030

[45] PARK, J. and MUANDET, K. (2020). A measure-theoretic approach to kernel conditional mean embeddings. *Adv. Neural Inf. Process*. *Syst*. **33**.

[46] PATRA, R. K., SEN, B. and SZÉKELY, G. J. (2016). On a nonparametric notion of residual and its applications. *Statist*. *Probab. Lett*. **109** 208–213. MR3434980 https://doi.org/10.1016/j.spl.2015.10.011

[47] PEARL, J. (2014). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*. Elsevier. MR0965765

[48] PETERSEN, L. and HANSEN, N. R. (2021). Testing conditional independence via quantile regression based partial copulas. *J. Mach. Learn*. *Res*. **22** Paper No. 70, 47. MR4253763

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

[50] RUNGE, J. (2018). Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In *International Conference on Artificial Intelligence and Statistics* 938–947. PMLR.

[51] SASON, I. and VERDÚ, S. (2016). f-divergence inequalities. *IEEE Trans*. *Inf. Theory* **62** 5973–6006. MR3565096 https://doi.org/10.1109/TIT.2016.2603151

[52] SCHRAB, A., KIM, I., ALBERT, M., LAURENT, B., GUEDJ, B. and GRETTON, A. (2021). MMD aggregated two-sample test. arXiv preprint arXiv:2110.15073.

[53] SEN, R., SHANMUGAM, K., ASNANI, H., RAHIMZAMANI, A. and KANNAN, S. (2018). Mimic and classify: A meta-algorithm for conditional independence testing. arXiv preprint arXiv:1806.09708.

[54] SEN, R., SURESH, A. T., SHANMUGAM, K., DIMAKIS, A. G. and SHAKKOTTAI, S. (2017). Model-powered conditional independence test. In *Advances in Neural Information Processing Systems* 2951–2961.

[55] 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

[56] SHENG, T. and SRIPERUMBUDUR, B. K. (2019). On distance and kernel measures of conditional independence. arXiv preprint arXiv:1912.01103.

[57] SHI, C., XU, T., BERGSMA, W. and LI, L. (2021). Double generative adversarial networks for conditional independence testing. *J*. *Mach*. *Learn*. *Res*. **22** Paper No. [285], 32. MR4353064 https://doi.org/10.1515/ijnsns-2016-0151

[58] SONG, K. (2009). Testing conditional independence via Rosenblatt transforms. *Ann*. *Statist*. **37** 4011–4045. MR2572451 https://doi.org/10.1214/09-AOS704

[59] SPOHN, W. (1994). On the properties of conditional independence. In *Patrick Suppes*: *Scientific Philosopher*, *Vol*. 1. *Synthese Lib*. **233** 173–196. Kluwer Academic, Dordrecht. MR1349350 https://doi.org/10.1007/978-94-011-0774-7_7

[60] STROBL, E. V., ZHANG, K. and VISWESWARAN, S. (2019). Approximate kernel-based conditional independence tests for fast non-parametric causal discovery. *J. Causal Inference* **7** Art. No. 20180017, 24. MR4350065 https://doi.org/10.1515/jci-2018-0017

[61] SU, L. and WHITE, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* **24** 829–864. MR2428851 https://doi.org/10.1017/S0266466608080341

[62] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500

[63] WANG, X., PAN, W., HU, W., TIAN, Y. and ZHANG, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734. MR3449068 https://doi.org/10.1080/01621459.2014.993081

[64] WILLIAMSON, B. D., GILBERT, P. B., CARONE, M. and SIMON, N. (2021). Nonparametric variable importance assessment using machine learning techniques. *Biometrics* **77** 9–22. MR4229718 https://doi.org/10.1111/biom.13392

[65] YAO, Q. and TRITCHLER, D. (1993). An exact analysis of conditional independence in several 2×2 contingency tables. *Biometrics* **49** 233–236. MR1221407 https://doi.org/10.2307/2532617

[66] ZHANG, H., ZHOU, S. and GUAN, J. (2018). Measuring conditional independence by independent residuals: Theoretical results and application in causal discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence* **32**.

[67] ZHANG, K., PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2012). Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Annual Conference on Uncertainty in Artificial Intelligence* 804–813.

[68] ZHOU, Y., LIU, J. and ZHU, L. (2020). Test for conditional independence with application to conditional screening. *J. Multivariate Anal.* **175** 104557, 18. MR4017960 https://doi.org/10.1016/j.jmva.2019.104557

The Annals of Statistics
2022, Vol. 50, No. 6, 3415–3438
https://doi.org/10.1214/22-AOS2234

ASYMPTOTIC PROPERTIES OF HIGH-DIMENSIONAL RANDOM FORESTS

BY CHIEN-MING CHI[1,a], PATRICK VOSSLER[2,b], YINGYING FAN[2,c] AND JINCHI LV[2,d]

[1]*Institute of Statistical Science, Academia Sinica,* [a]*xbbchi@stat.sinica.edu.tw*
[2]*Data Sciences and Operations Department, Marshall School of Business, University of Southern California,* [b]*pvossler@marshall.usc.edu,* [c]*fanyingy@marshall.usc.edu,* [d]*jinchilv@marshall.usc.edu*

As a flexible nonparametric learning tool, the random forests algorithm has been widely applied to various real applications with appealing empirical performance, even in the presence of high-dimensional feature space. Unveiling the underlying mechanisms has led to some important recent theoretical results on the consistency of the random forests algorithm and its variants. However, to our knowledge, almost all existing works concerning random forests consistency in a high-dimensional setting were established for various modified random forests models where the splitting rules are independent of the response; a few exceptions assume simple data generating models with binary features. In light of this, in this paper we derive the consistency rates for the random forests algorithm associated with the sample CART splitting criterion, which is the one used in the original version of the algorithm (*Mach. Learn.* **45** (2001) 5–32), in a general high-dimensional nonparametric regression setting through a bias-variance decomposition analysis. Our new theoretical results show that random forests can indeed adapt to high dimensionality and allow for discontinuous regression function. Our bias analysis characterizes explicitly how the random forests bias depends on the sample size, tree height and column subsampling parameter. Some limitations on our current results are also discussed.

REFERENCES

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

[2] BAI, Z.-D., DEVROYE, L., HWANG, H.-K. and TSAI, T.-H. (2005). Maxima in hypercubes. *Random Structures Algorithms* **27** 290–309. MR2162600 https://doi.org/10.1002/rsa.20053

[3] BIAU, G. (2012). Analysis of a random forests model. *J. Mach. Learn. Res.* **13** 1063–1095. MR2930634

[4] BIAU, G., DEVROYE, L. and LUGOSI, G. (2008). Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.* **9** 2015–2033. MR2447310

[5] BIAU, G. and SCORNET, E. (2016). A random forest guided tour. *TEST* **25** 197–227. MR3493512 https://doi.org/10.1007/s11749-016-0481-7

[6] BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.

[7] BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

[8] BREIMAN, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA* **1** 58.

[9] CHI, C.-M, VOSSLER, P., FAN, Y. and LV, J. (2022). Supplement to "Asymptotic properties of high-dimensional random forests." https://doi.org/10.1214/22-AOS2234SUPP

[10] DÍAZ-URIARTE, R. and DE ANDRES, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **7** 3.

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

[12] FAN, J., FENG, Y. and SONG, R. (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *J. Amer. Statist. Assoc.* **106** 544–557. MR2847969 https://doi.org/10.1198/jasa.2011.tm09779

MSC2020 subject classifications. Primary 62G08, 62G05; secondary 62H12, 62G20.
Key words and phrases. Random forests, nonparametric learning, high dimensionality, consistency, rate of convergence, sparsity.

[13] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x
[14] FAN, J. and LV, J. (2018). Sure independence screening (invited review article). *Wiley StatsRef: Statistics Reference Online* 1–8.
[15] GENUER, R. (2012). Variance reduction in purely random forests. *J. Nonparametr. Stat.* **24** 543–562. MR2968888 https://doi.org/10.1080/10485252.2012.677843
[16] GISLASON, P. O., BENEDIKTSSON, J. A. and SVEINSSON, J. R. (2006). Random forests for land cover classification. *Pattern Recogn. Lett.* **27** 294–300.
[17] GOLDSTEIN, B. A., POLLEY, E. C. and BRIGGS, F. B. S. (2011). Random forests for genetic association studies. *Stat. Appl. Genet. Mol. Biol.* **10** Art. 32. MR2823518 https://doi.org/10.2202/1544-6115.1691
[18] HOWARD, J. and BOWLES, M. (2012). The two most important algorithms in predictive modeling today. In *Strata Conference Presentation* **28**.
[19] ISHWARAN, H. and KOGALUR, U. B. (2010). Consistency of random survival forests. *Statist. Probab. Lett.* **80** 1056–1064. MR2651045 https://doi.org/10.1016/j.spl.2010.02.020
[20] ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H. and LAUER, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860. MR2516796 https://doi.org/10.1214/08-AOAS169
[21] KHAIDEM, L., SAHA, S. and DEY, S. R. (2016). Predicting the direction of stock market prices using random forest. ArXiv Preprint. Available at arXiv:1605.00003.
[22] KLUSOWSKI, J. and TIAN, P. (2022). Large scale prediction with decision trees. *J. Amer. Statist. Assoc.* To appear.
[23] KLUSOWSKI, J. M. (2019). Analyzing CART. ArXiv Preprint. Available at arXiv:1906.10086.
[24] KLUSOWSKI, J. M. (2021). Sharp analysis of a simple model for random forests. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 757–765.
[25] LIAW, A. and WIENER, M. (2002). Classification and regression by randomForest. *R News* **2** 18–22.
[26] LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176 https://doi.org/10.1198/016214505000001230
[27] LOUPPE, G., WEHENKEL, L., SUTERA, A. and GEURTS, P. (2013). Understanding variable importances in forests of randomized trees. *Adv. Neural Inf. Process. Syst.* **26** 431–439.
[28] MENTCH, L. and HOOKER, G. (2014). Ensemble trees and CLTs: Statistical inference for supervised learning. ArXiv preprint. Available at arXiv:1404.6473.
[29] MOURTADA, J., GAÏFFAS, S. and SCORNET, E. (2020). Minimax optimal rates for Mondrian trees and forests. *Ann. Statist.* **48** 2253–2276. MR4134794 https://doi.org/10.1214/19-AOS1886
[30] NOBEL, A. (1996). Histogram regression estimation using data-dependent partitions. *Ann. Statist.* **24** 1084–1105. MR1401839 https://doi.org/10.1214/aos/1032526958
[31] QI, Y. (2012). Random forest for bioinformatics. In *Ensemble Machine Learning* 307–323. Springer, Berlin.
[32] SCORNET, E. (2020). Trees, forests, and impurity-based variable importance. ArXiv Preprint. Available at arXiv:2001.04295.
[33] SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876 https://doi.org/10.1214/15-AOS1321
[34] STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620. MR0443204
[35] SYRGKANIS, V. and ZAMPETAKIS, M. (2020). Estimation and inference with trees and forests in high dimensions. In *Conference on Learning Theory* 3453–3454. PMLR.
[36] VARIAN, H. R. (2014). Big data: New tricks for econometrics. *J. Econ. Perspect.* **28** 3–28.
[37] WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839
[38] WAGER, S., HASTIE, T. and EFRON, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15** 1625–1651. MR3225243
[39] ZHU, R., ZENG, D. and KOSOROK, M. R. (2015). Reinforcement learning trees. *J. Amer. Statist. Assoc.* **110** 1770–1784. MR3449072 https://doi.org/10.1080/01621459.2015.1036994

The Annals of Statistics
2022, Vol. 50, No. 6, 3439–3465
https://doi.org/10.1214/22-AOS2235

RERANDOMIZATION WITH DIMINISHING COVARIATE IMBALANCE AND DIVERGING NUMBER OF COVARIATES

BY YUHAO WANG[1,a] AND XINRAN LI[2,b]

[1]*Institute for Interdisciplinary Information Sciences, Tsinghua University,* [a]*yuhaow@tsinghua.edu.cn*
[2]*Department of Statistics, University of Illinois at Urbana-Champaign,* [b]*xinranli@illinois.edu*

Completely randomized experiments have been the gold standard for drawing causal inference because they can balance all potential confounding on average. However, they may suffer from unbalanced covariates for realized treatment assignments. Rerandomization, a design that rerandomizes the treatment assignment until a prespecified covariate balance criterion is met, has recently got attention due to its easy implementation, improved covariate balance and more efficient inference. Researchers have then suggested to use the treatment assignments that minimize the covariate imbalance, namely the optimally balanced design. This has caused again the long-time controversy between two philosophies for designing experiments: randomization versus optimal, and thus almost deterministic designs. Existing literature argued that rerandomization with overly balanced observed covariates can lead to highly imbalanced unobserved covariates, making it vulnerable to model misspecification. On the contrary, rerandomization with properly balanced covariates can provide robust inference for treatment effects while sacrificing some efficiency compared to the ideally optimal design. In this paper, we show it is possible that, by making the covariate imbalance diminishing at a proper rate as the sample size increases, rerandomization can achieve its ideally optimal precision that one can expect with perfectly balanced covariates, while still maintaining its robustness. We further investigate conditions on the number of covariates for achieving the desired optimality. Our results rely on a more delicate asymptotic analysis for rerandomization, allowing both diminishing covariate imbalance threshold (or equivalently the acceptance probability) and diverging number of covariates. The derived theory for rerandomization provides a deeper understanding of its large-sample property and can better guide its practical implementation. Furthermore, it also helps reconcile the controversy between randomized and optimal designs in an asymptotic sense.

REFERENCES

BANERJEE, A. V., CHASSANG, S., MONTERO, S. and SNOWBERG, E. (2020). A theory of experimenters: Robustness, randomization, and balance. *Am. Econ. Rev.* **110** 1206–1230.

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

BENTKUS, V. (2004). A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen.* **49** 400–410. MR2144310 https://doi.org/10.1137/S0040585X97981123

BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498 https://doi.org/10.2307/1990053

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482. MR0740906 https://doi.org/10.1214/aos/1176346500

BIKELIS, A. (1969). The estimation of the remainder term in the central limit theorem for samples taken from finite sets. *Studia Sci. Math. Hungar.* **4** 345–354. MR0254902

MSC2020 subject classifications. Primary 62K99; secondary 62K05.
Key words and phrases. Causal inference, optimal rerandomization, Mahalanobis distance, high-dimensional covariates, Berry–Esseen bound.

BLONIARZ, A., LIU, H., ZHANG, C.-H., SEKHON, J. S. and YU, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proc. Natl. Acad. Sci. USA* **113** 7383–7390. MR3531136 https://doi.org/10.1073/pnas.1510506113

BOLTHAUSEN, E. and GÖTZE, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21** 1692–1710. MR1245764 https://doi.org/10.1214/aos/1176349393

BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2140250

BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education high school study. *Ann. Appl. Stat.* **10** 1958–1976. MR3592044 https://doi.org/10.1214/16-AOAS959

BRANSON, Z. and SHAO, S. (2021). Ridge rerandomization: An experimental design strategy in the presence of covariate collinearity. *J. Statist. Plann. Inference* **211** 287–314. MR4125927 https://doi.org/10.1016/j.jspi.2020.07.002

BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1** 200–232.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Preprint. Available at arXiv:2012.09513.

COX, D. R. (1982). Randomization and concomitant variables in the design of experiments. In *Statistics and Probability: Essays in Honor of C. R. Rao* 197–202. North-Holland, Amsterdam. MR0659470

COX, D. R. (2007). Applied statistics: A review. *Ann. Appl. Stat.* **1** 1–16. MR2393838 https://doi.org/10.1214/07-AOAS113

ESSEEN, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **28A** 19. MR0011909

FAN, C. T., MULLER, M. E. and REZUCHA, I. (1962). Development of sampling plans by using sequential (item by item) selection techniques and digital computers. *J. Amer. Statist. Assoc.* **57** 387–402. MR0139243

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

FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, Edinburgh, London.

FREEDMAN, D. A. (2008). Editorial: Oasis or mirage? *Chance* **21** 59–61. MR2422783 https://doi.org/10.1007/s00144-008-0047-x

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 361–374. MR0125612

HECKMAN, J. J. and KARAPAKULA, G. (2021). Using a satisficing model of experimenter decision-making to guide finite-sample inference for compromised experiments. *Econom. J.* **24** C1–C39. MR4281220 https://doi.org/10.1093/ectj/utab009

HÖGLUND, T. (1978). Sampling from a finite population: A remainder term estimate. *Scand. J. Stat.* **5** 69–71. MR0471130

JOHANSSON, P., RUBIN, D. B. and SCHULTZBERG, M. (2021). On optimal rerandomization designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 395–403. MR4250281 https://doi.org/10.1111/rssb.12417

JOHANSSON, P. and SCHULTZBERG, M. (2022). Rerandomization: A complement or substitute for stratification in randomized experiments? *J. Statist. Plann. Inference* **218** 43–58. MR4327401 https://doi.org/10.1016/j.jspi.2021.09.002

KALLUS, N. (2018). Optimal *a priori* balance in the design of controlled experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 85–112. MR3744713 https://doi.org/10.1111/rssb.12240

KAPELNER, A., KRIEGER, A. M., SKLAR, M., SHALIT, U. and AZRIEL, D. (2021). Harmonizing optimized designs with classic randomization in experiments. *Amer. Statist.* **75** 195–206. MR4256124 https://doi.org/10.1080/00031305.2020.1717619

KASY, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. *Polit. Anal.* **24** 324–338.

KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319. MR0113263

LEI, L. and DING, P. (2021). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika* **108** 815–828. MR4341353 https://doi.org/10.1093/biomet/asaa103

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LI, X. and DING, P. (2020). Rerandomization and regression adjustment. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 241–268. MR4060984

LI, X., DING, P. and RUBIN, D. B. (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proc. Natl. Acad. Sci. USA* **115** 9157–9162. MR3856113 https://doi.org/10.1073/pnas.1808191115

LI, X., DING, P. and RUBIN, D. B. (2020). Rerandomization in 2^K factorial experiments. *Ann. Statist.* **48** 43–63. MR4065152 https://doi.org/10.1214/18-AOS1790

LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

LIU, H., REN, J. and YANG, Y. (2022). Randomization-based Joint Central Limit Theorem and Efficient Covariate Adjustment in Randomized Block 2^K Factorial Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2022.2102985

MACKINNON, J. G. (2013). Thirty years of heteroskedasticity-robust inference. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis* 437–461. Springer, Berlin.

MENG, X. (2013). Scalable simple random sampling and stratified sampling. In *International Conference on Machine Learning. PMLR* 531–539.

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40** 1263–1282. MR2985950 https://doi.org/10.1214/12-AOS1008

MORGAN, K. L. and RUBIN, D. B. (2015). Rerandomization to balance tiers of covariates. *J. Amer. Statist. Assoc.* **110** 1412–1421. MR3449036 https://doi.org/10.1080/01621459.2015.1079528

RAIČ, M. (2015). Multivariate normal approximation: Permutation statistics, local dependence and beyond.

RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

ROSENBAUM, P. R. (2010). *Design of Observational Studies. Springer Series in Statistics*. Springer, New York. MR2561612 https://doi.org/10.1007/978-1-4419-1213-8

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

SAVAGE, L. J. (1962). *The Foundations of Statistical Inference*. Methuen, London.

SCHULTZBERG, M. and JOHANSSON, P. (2020). Asymptotic inference for optimal rerandomization designs. *Open Stat.* **1** 49–58. https://doi.org/10.1515/stat-2020-0102

SHI, L. and DING, P. (2022). Berry–Esseen bounds for design-based causal inference with possibly diverging treatment levels and varying group sizes. Preprint. Available at arXiv:2209.12345.

SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. MR1092986

STUDENT (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29** 363–378.

TAVES, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.* **15** 443–453.

WAGER, S., DU, W., TAYLOR, J. and TIBSHIRANI, R. J. (2016). High-dimensional regression adjustments in randomized experiments. *Proc. Natl. Acad. Sci. USA* **113** 12673–12678. MR3576188 https://doi.org/10.1073/pnas.1614732113

WANG, Y. and LI, X. (2022). Supplement to "Rerandomization with diminishing covariate imbalance and diverging number of covariates." https://doi.org/10.1214/22-AOS2235SUPP

WANG, X., WANG, T. and LIU, H. (2021). Rerandomization in Stratified Randomized Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2021.1990767

YANG, Z., QU, T. and LI, X. (2021). Rejective Sampling, Rerandomization, and Regression Adjustment in Survey Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2021.1984926

ZHANG, H., YIN, G. and RUBIN, D. B. (2021). PCA Rerandomization. Preprint. Available at arXiv:2102.12262.

ZHOU, Q., ERNST, P. A., MORGAN, K. L., RUBIN, D. B. and ZHANG, A. (2018). Sequential rerandomization. *Biometrika* **105** 745–752. MR3842897 https://doi.org/10.1093/biomet/asy031

The Annals of Statistics
2022, Vol. 50, No. 6, 3466–3483
https://doi.org/10.1214/22-AOS2236

CHOOSING BETWEEN PERSISTENT AND STATIONARY VOLATILITY

BY ILIAS CHRONOPOULOS[1,a], LIUDAS GIRAITIS[2,b] AND GEORGE KAPETANIOS[3,c]

[1]*Essex Business School, University of Essex,* [a]*ilias.chronopoulos@essex.ac.uk*
[2]*School of Economics and Finance, Queen Mary University of London,* [b]*l.giraitis@qmul.ac.uk*
[3]*King's Business School, King's College London,* [c]*george.kapetanios@kcl.ac.uk*

This paper suggests a multiplicative volatility model where volatility is decomposed into a stationary and a nonstationary persistent part. We provide a testing procedure to determine which type of volatility is prevalent in the data. The persistent part of volatility is associated with a nonstationary persistent process satisfying some smoothness and moment conditions. The stationary part is related to stationary conditional heteroskedasticity. We outline theory and conditions that allow the extraction of the persistent part from the data and enable standard conditional heteroskedasticity tests to detect stationary volatility after persistent volatility is taken into account. Monte Carlo results support the testing strategy in small samples. The empirical application of the theory supports the persistent volatility paradigm, suggesting that stationary conditional heteroskedasticity is considerably less pronounced than previously thought.

REFERENCES

[1] AMADO, C. and TERÄSVIRTA, T. (2013). Modelling volatility by variance decomposition. *J. Econometrics* **175** 142–153. MR3061936 https://doi.org/10.1016/j.jeconom.2013.03.006

[2] AMADO, C. and TERÄSVIRTA, T. (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Rev.* **36** 421–446. MR3608383 https://doi.org/10.1080/07474938.2014.977064

[3] BERA, A. K. and HIGGINS, M. L. (1993). ARCH models: Properties, estimation and testing. *J. Econ. Surv.* **7** 305–366.

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

[5] BOLLERSLEV, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Rev. Econ. Stat.* **69** 542–547.

[6] BROWNLEES, C. T. and GALLO, G. M. (2009). Comparison of volatility measures: A risk management perspective. *J. Financ. Econom.* **8** 29–56.

[7] CHRONOPOULOS, I., GIRAITIS, L. and KAPETANIOS, G. (2022). Supplement to "Choosing between persistent and stationary volatility." https://doi.org/10.1214/22-AOS2236SUPP

[8] CHRONOPOULOS, I., KAPETANIOS, G. and PETROVA, K. (2021). Kernel-based volatility generalised least squares. *Econom. Stat.* **20** 2–11. MR4302583 https://doi.org/10.1016/j.ecosta.2019.11.001

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

[10] DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.* **34** 2790–2824. MR2329468 https://doi.org/10.1214/009053606000000867

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

[12] DALLA, V., GIRAITIS, L. and PHILLIPS, P. C. B. (2015). Testing mean stability of heteroskedastic time series. Cowles Foundation Discussion Paper No. 2006.

[13] DALLA, V., GIRAITIS, L. and PHILLIPS, P. C. B. (2022). Robust tests for white noise and cross-correlation. *Econometric Theory* **38** 913–941. MR4513134 https://doi.org/10.1017/S0266466620000341

MSC2020 subject classifications. Primary 62M10, 62G08; secondary 62G05, 62F05.
Key words and phrases. ARCH effect, persistence, volatility, time-varying coefficient models, nonparametric estimation.

[14] DENDRAMIS, Y., GIRAITIS, L. and KAPETANIOS, G. (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory* **37** 1100–1134. MR4348398 https://doi.org/10.1017/S0266466620000535

[15] DING, Z., GRANGER, C. W. and ENGLE, R. F. (1993). A long memory property of stock market returns and a new model. *J. Empir. Finance* **1** 83–106.

[16] ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121 https://doi.org/10.2307/1912773

[17] ENGLE, R. F. and RANGEL, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Rev. Financ. Stud.* **21** 1187–1222.

[18] ERAKER, B., JOHANNES, M. and POLSON, N. (2003). The impact of jumps in volatility and returns. *J. Finance* **58** 1269–1300.

[19] GALLANT, A. R. (1984). The Fourier flexible form. *Amer. J. Agric. Econ.* **66** 204–208.

[20] GIRAITIS, L., KAPETANIOS, G., WETHERILT, A. and ŽIKEŠ, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *J. Appl. Econometrics* **31** 58–84. MR3464851 https://doi.org/10.1002/jae.2457

[21] GIRAITIS, L., KAPETANIOS, G. and YATES, T. (2014). Inference on stochastic time-varying coefficient models. *J. Econometrics* **179** 46–65. MR3153648 https://doi.org/10.1016/j.jeconom.2013.10.009

[22] GIRAITIS, L., KAPETANIOS, G. and YATES, T. (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *J. Time Series Anal.* **39** 129–149. MR3767057 https://doi.org/10.1111/jtsa.12271

[23] GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London. MR2977317 https://doi.org/10.1142/p591

[24] GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801.

[25] HEBER, G., LUNDE, A., SHEPHARD, N. and SHEPPARD, K. (2009). Oxford-Man Institute's realized library, version 0.1. Available at https://realized.oxford-man.ox.ac.uk/data/download.

[26] JIANG, F., LI, D. and ZHU, K. (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *J. Econometrics* **224** 306–329. MR4303783 https://doi.org/10.1016/j.jeconom.2020.10.007

[27] KAPETANIOS, G. and YATES, T. (2014). Evolving UK and US macroeconomic dynamics through the lens of a model of deterministic structural change. *Empir. Econ.* **47** 305–345.

[28] MAZUR, B. and PIPIEŃ, M. (2012). On the empirical importance of periodicity in the volatility of financial returns-time varying GARCH as a second order APC (2) process. *Cent. Eur. J. Econ. Model. Econom.* **4** 95–116.

[29] MERCURIO, D. and SPOKOINY, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32** 577–602. MR2060170 https://doi.org/10.1214/009053604000000102

[30] MIKOSCH, T. and STĂRICĂ, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Rev. Econ. Stat.* **86** 378–390.

[31] PATTON, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *J. Econometrics* **160** 246–256. MR2745881 https://doi.org/10.1016/j.jeconom.2010.03.034

[32] PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27** 204–237. MR0199886

[33] RUIZ, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *J. Econometrics* **63** 289–306.

[34] SHEPHARD, N. and SHEPPARD, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (heavy) models. *J. Appl. Econometrics* **25** 197–231. MR2758633 https://doi.org/10.1002/jae.1158

[35] TRAPANI, L. (2016). Testing for (in)finite moments. *J. Econometrics* **191** 57–68. MR3434435 https://doi.org/10.1016/j.jeconom.2015.08.006

[36] VAN BELLEGEM, S. and VON SACHS, R. (2004). Forecasting economic time series with unconditional time-varying variance. *Int. J. Forecast.* **20** 611–627.

The Annals of Statistics
2022, Vol. 50, No. 6, 3484–3509
https://doi.org/10.1214/22-AOS2237

SUP-NORM ADAPTIVE DRIFT ESTIMATION FOR MULTIVARIATE NONREVERSIBLE DIFFUSIONS

BY CATHRINE AECKERLE-WILLEMS[1,a] AND CLAUDIA STRAUCH[2,b]

[1]*Department of Economics, University of Mannheim,* [a]*aeckerle@uni-mannheim.de*
[2]*Department of Mathematics, Aarhus University,* [b]*strauch@math.au.dk*

We consider the question of estimating the drift for a large class of ergodic multivariate and possibly nonreversible diffusion processes, based on continuous observations, in sup-norm loss. Nonparametric classes of smooth functions of unknown order are considered, and we suggest an adaptive approach which allows to construct drift estimators attaining optimal sup-norm rates of convergence. Reversibility structures and related functional inequalities are known to be key tools for these estimation problems. We can discard such restrictions by making use of mixing properties which are satisfied for the very general class of processes under consideration. Analysing diffusions, the scalar case is very distinct from the general multivariate setting. Therefore, we treat scalar and multivariate processes separately which leads to in several aspects improved univariate results. While we consider drift estimation on bounded domains for exponentially β-mixing multivariate processes, for scalar diffusion processes we work under minimal assumptions that allow estimation of unbounded drift terms over the entire real line, and we provide classical minimax results (including lower bounds) which cannot be obtained under state-of-the-art conditions in the multivariate case. In addition, we prove a Donsker theorem for the classical kernel estimator of the invariant density in the scalar setting and establish its semiparametric efficiency.

REFERENCES

AECKERLE-WILLEMS, C. and STRAUCH, C. (2018). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. Available at arXiv:1808.10660.

AECKERLE-WILLEMS, C. and STRAUCH, C. (2021). Concentration of scalar ergodic diffusions and some statistical implications. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1857–1887. MR4328556 https://doi.org/10.1214/20-aihp1144

AECKERLE-WILLEMS, C. and STRAUCH, C. (2022). Supplement to "Sup-norm adaptive drift estimation for multivariate nonreversible diffusions." https://doi.org/10.1214/22-AOS2237SUPP

AMORINO, C. (2021). Rate of estimation for the stationary distribution of jump-processes over anisotropic Holder classes. *Electron. J. Stat.* **15** 5067–5116. MR4347384 https://doi.org/10.1214/21-ejs1913

AMORINO, C. and GLOTER, A. (2021a). Invariant density adaptive estimation for ergodic jump-diffusion processes over anisotropic classes. *J. Statist. Plann. Inference* **213** 106–129. MR4184676 https://doi.org/10.1016/j.jspi.2020.11.006

AMORINO, C. and GLOTER, A. (2021b). Minimax rate of estimation for invariant densities associated to continuous stochastic differential equations over anisotropic Holder classes. Available at arXiv:2110.02774.

BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160 https://doi.org/10.1016/j.jfa.2007.11.002

BARLOW, M. T. and YOR, M. (1982). Semimartingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times. *J. Funct. Anal.* **49** 198–229. MR0680660 https://doi.org/10.1016/0022-1236(82)90080-5

BERTIN, K. (2004). Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes. *Bernoulli* **10** 873–888. MR2093615 https://doi.org/10.3150/bj/1099579160

BICKEL, P. J. and RITOV, Y. (2003). Nonparametric estimators which can be "plugged-in". *Ann. Statist.* **31** 1033–1053. MR2001641 https://doi.org/10.1214/aos/1059655904

MSC2020 subject classifications. 62M05, 62G20, 60F05, 60J60.
Key words and phrases. Multivariate ergodic diffusion, adaptive estimation, minimax optimality.

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

COMTE, F. and GENON-CATALOT, V. (2021). Drift estimation on non compact support for diffusion models. *Stochastic Process. Appl.* **134** 174–207. MR4206844 https://doi.org/10.1016/j.spa.2021.01.001

DALALYAN, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507–2528. MR2253093 https://doi.org/10.1214/009053605000000615

DEXHEIMER, N., STRAUCH, C. and TROTTNER, L. (2020). Mixing it up: A general framework for Markovian statistics. Available at arXiv:2011.00308v1.

DEXHEIMER, N., STRAUCH, C. and TROTTNER, L. (2022). Adaptive invariant density estimation for continuous-time mixing Markov processes under sup-norm risk. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 2029–2064. MR4492970 https://doi.org/10.1214/21-aihp1235

DIRKSEN, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** no. 53, 29. MR3354613 https://doi.org/10.1214/EJP.v20-3760

GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456 https://doi.org/10.2307/3318681

GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673 https://doi.org/10.1007/s00440-008-0137-y

GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 https://doi.org/10.1214/009053604000000797

HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser/Springer, Basel. MR2857021 https://doi.org/10.1007/978-3-0348-0021-1_7

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

KOROSTELËV, A. P. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. *Teor. Veroyatn. Primen.* **38** 875–882. MR1318004 https://doi.org/10.1137/1138075

KUTOYANTS, YU. A. (1998). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1** 131–155. MR2797140 https://doi.org/10.1023/A:1009919612081

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

KUTOYANTS, Y. A. and YOSHIDA, N. (2007). Moment estimation for ergodic diffusion processes. *Bernoulli* **13** 933–951. MR2364220 https://doi.org/10.3150/07-BEJ1040

LEPSKIĬ, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation. Adv. Soviet Math.* **12** 87–106. Amer. Math. Soc., Providence, RI. MR1191692 https://doi.org/10.1137/1136085

MEYER, P. A. (1976). Un cours sur les intégrales stochastiques. In *Séminaire de Probabilités, X (Seconde Partie: Théorie des Intégrales Stochastiques, Univ. Strasbourg, Strasbourg, Année Universitaire* 1974/1975). *Lecture Notes in Math.* **511** 245–400. Springer, Berlin. MR0501332

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. In *Eleventh Saint Flour Probability Summer School*—1981 (*Saint Flour*, 1981). *Lecture Notes in Math.* **976** 75–265. Springer, Berlin. MR0722983 https://doi.org/10.1007/BFb0067986

NEGRI, I. (2001). On efficient estimation of invariant density for ergodic diffusion processes. *Statist. Probab. Lett.* **51** 79–85. MR1820148 https://doi.org/10.1016/S0167-7152(00)00147-4

NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.* **48** 1383–1408. MR4124327 https://doi.org/10.1214/19-AOS1851

NICKL, R. and SÖHL, J. (2019). Bernstein-von Mises theorems for statistical inverse problems II: Compound Poisson processes. *Electron. J. Stat.* **13** 3513–3571. MR4013745 https://doi.org/10.1214/19-ejs1609

OGA, A. and KOIKE, Y. (2021). Drift estimation for a multi-dimensional diffusion process using deep neural networks. Available at arXiv:2112.13332.

PARDOUX, E. and VERETENNIKOV, A. YU. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.* **123** 603–628. MR3003365 https://doi.org/10.1016/j.spa.2012.08.010

QIAN, Z. and ZHENG, W. (2004). A representation formula for transition probability densities of diffusions and applications. *Stochastic Process. Appl.* **111** 57–76. MR2049569 https://doi.org/10.1016/j.spa.2003.12.004

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

SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462. MR3581829 https://doi.org/10.1051/ps/2016017

SPOKOINY, V. G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.* **28** 815–836. MR1792788 https://doi.org/10.1214/aos/1015951999

STRAUCH, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Process. Appl.* **125** 2562–2602. MR3332848 https://doi.org/10.1016/j.spa.2015.02.003

STRAUCH, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields* **164** 361–400. MR3449393 https://doi.org/10.1007/s00440-014-0614-4

TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239 https://doi.org/10.1214/aos/1024691478

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

VAN DER MEULEN, F. H., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** 863–888. MR2265666 https://doi.org/10.3150/bj/1161614950

VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485 https://doi.org/10.3150/11-BEJ385

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

VAN DER VAART, A. and VAN ZANTEN, H. (2005). Donsker theorems for diffusions: Necessary and sufficient conditions. *Ann. Probab.* **33** 1422–1451. MR2150194 https://doi.org/10.1214/009117905000000152

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

VAN WAAIJ, J. and VAN ZANTEN, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.* **10** 628–645. MR3471991 https://doi.org/10.1214/16-EJS1117

The Annals of Statistics
2022, Vol. 50, No. 6, 3510–3537
https://doi.org/10.1214/22-AOS2238

TESTING FOR THE RANK OF A COVARIANCE OPERATOR

BY ANIRVAN CHARKABORTY[1,a] AND VICTOR M. PANARETOS[2,b]

[1]*Department of Mathematics & Statistics, Indian Institute of Science Education and Research Kolkata,* [a]*anirvan.c@iiserkol.ac.in*

[2]*Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne,* [b]*victor.panaretos@epfl.ch*

How can we discern whether the covariance operator of a stochastic process is of reduced rank, and if so, what its precise rank is? And how can we do so at a given level of confidence? This question is central to a great deal of methods for functional data, which require low-dimensional representations whether by functional PCA or other methods. The difficulty is that the determination is to be made on the basis of i.i.d. replications of the process observed discretely and with measurement error contamination. This adds a ridge to the empirical covariance, obfuscating the underlying dimension. We build a matrix-completion inspired test statistic that circumvents this issue by measuring the best possible least square fit of the empirical covariance's off-diagonal elements, optimised over covariances of given finite rank. For a fixed grid of sufficiently large size, we determine the statistic's asymptotic null distribution as the number of replications grows. We then use it to construct a bootstrap implementation of a stepwise testing procedure controlling the familywise error rate corresponding to the collection of hypotheses formalising the question at hand. Under minimal regularity assumptions, we prove that the procedure is consistent and that its bootstrap implementation is valid. The procedure circumvents smoothing and associated smoothing parameters, is indifferent to measurement error heteroskedasticity, and does not assume a low-noise regime. An extensive simulation study reveals an excellent practical performance, stably across a wide range of settings and the procedure is further illustrated by means of two data analyses.

REFERENCES

AMINI, A. A. and WAINWRIGHT, M. J. (2012). Sampled forms of functional PCA in reproducing kernel Hilbert spaces. *Ann. Statist.* **40** 2483–2510. MR3097610 https://doi.org/10.1214/12-AOS1033

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

BATHIA, N., YAO, Q. and ZIEGELMANN, F. (2010). Identifying the finite dimensionality of curve time series. *Ann. Statist.* **38** 3352–3386. MR2766855 https://doi.org/10.1214/10-AOS819

BRESLER, Y. and MACOVSKI, A. (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust. Speech Signal Process.* **34** 1081–1089.

CAREY, J., LIEDO, P., MÜLLER, H., WANG, J. and CHIOU, J. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J. Gerontol., Ser. A, Biol. Sci. Med. Sci.* **53** 245–251.

CHARKABORTY, A. and PANARETOS, V. M. (2022). Supplement to "Testing for the rank of a covariance operator." https://doi.org/10.1214/22-AOS2238SUPP

CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. Tech. report arXiv:1509.03025.

CHOI, Y., TAYLOR, J. and TIBSHIRANI, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *Ann. Statist.* **45** 2590–2617. MR3737903 https://doi.org/10.1214/16-AOS1536

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

MSC2020 subject classifications. Primary 62G05, 62M40; secondary 15A99.

Key words and phrases. Bootstrap, functional data analysis, functional PCA, Karhunen–Loève expansion, matrix completion, measurement error, scree-plot.

FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. R. Soc. A, Math. Phys. Eng. Sci.* **125** 54–59.

HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 689–705. MR2301015 https://doi.org/10.1111/j.1467-9868.2006.00562.x

HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30** 179–185.

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

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

JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2036084

KNEIP, A. (1994). Nonparametric estimation of common regressors for similar curve data. *Ann. Statist.* **22** 1386–1427. MR1311981 https://doi.org/10.1214/aos/1176325634

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663 https://doi.org/10.1214/12-AOS970

LI, Y., WANG, N. and CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* **108** 1284–1294. MR3174708 https://doi.org/10.1080/01621459.2013.788980

LUO, W. and LI, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika* **103** 875–887. MR3620445 https://doi.org/10.1093/biomet/asw051

LYNCH, G., GUO, W., SARKAR, S. K. and FINNER, H. (2017). The control of the false discovery rate in fixed sequence multiple testing. *Electron. J. Stat.* **11** 4649–4673. MR3724971 https://doi.org/10.1214/17-EJS1359

MAURER, W., HOTHORN, L. and LEHMACHER, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: A-priori ordered hypotheses. In *Biometrie in der Chemische-pharmazeutichen Industrie* **6** (J. Vollman, ed.) Fischer Verlag, Stuttgart.

ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.

PATIL, A., RASTOGI, P. and RAPHAEL, B. (2005). Phase-shifting interferometry by a covariance-based method. *Appl. Opt.* **44** 5778–5785.

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

PERES-NETO, P. R., JACKSON, D. A. and SOMERS, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput. Statist. Data Anal.* **49** 974–997. MR2143053 https://doi.org/10.1016/j.csda.2004.06.015

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

STOICA, P. and MOSES, R. L. (2005). *Spectral Analysis of Signals*. Pearson/Prentice Hall, Inc.

UMESH, S. and TUFTS, D. (1996). Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to nmr spectroscopy data. *IEEE Trans. Signal Process.* **44** 2245–2259.

VELICER, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika* **41** 321–327.

XU, G., LIU, H., TONG, L. and KAILATH, T. (1995). A least-squares approach to blind channel identification. *IEEE Trans. Signal Process.* **43** 2982–2993.

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

The Annals of Statistics
2022, Vol. 50, No. 6, 3538–3562
https://doi.org/10.1214/22-AOS2239

COVARIANCE ESTIMATION UNDER ONE-BIT QUANTIZATION

BY SJOERD DIRKSEN[1,a], JOHANNES MALY[2,b] AND HOLGER RAUHUT[3,c]

[1]*Mathematical Institute, Utrecht University,* [a]*s.dirksen@uu.nl*
[2]*Department of Mathematics, Ludwig Maximilian University of Munich,* [b]*maly@math.lmu.de*
[3]*Chair for Mathematics of Information Processing, RWTH Aachen University,* [c]*rauhut@mathc.rwth-aachen.de*

We consider the classical problem of estimating the covariance matrix of a sub-Gaussian distribution from i.i.d. samples in the novel context of coarse quantization, that is, instead of having full knowledge of the samples, they are quantized to one or two bits per entry. This problem occurs naturally in signal processing applications. We introduce new estimators in two different quantization scenarios and derive nonasymptotic estimation error bounds in terms of the operator norm. In the first scenario, we consider a simple, scale-invariant one-bit quantizer and derive an estimation result for the correlation matrix of a centered Gaussian distribution. In the second scenario, we add random dithering to the quantizer. In this case, we can accurately estimate the full covariance matrix of a general sub-Gaussian distribution by collecting two bits per entry of each sample. In both scenarios, our bounds apply to masked covariance estimation. We demonstrate the near optimality of our error bounds by deriving corresponding (minimax) lower bounds and using numerical simulations.

REFERENCES

[1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** 535–561. MR2601042 https://doi.org/10.1090/S0894-0347-09-00650-X

[2] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2011). Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris* **349** 195–200. MR2769907 https://doi.org/10.1016/j.crma.2010.12.014

[3] APOSTOL, T. M. (2013). *Introduction to Analytic Number Theory*. Springer, New York. MR0434929

[4] BAR-SHALOM, O. and WEISS, A. J. (2002). DOA estimation using one-bit quantized measurements. *IEEE Trans. Aerosp. Electron. Syst.* **38** 868–884.

[5] BARANIUK, R. G., FOUCART, S., NEEDELL, D., PLAN, Y. and WOOTTERS, M. (2017). Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Trans. Inf. Theory* **63** 3368–3385. MR3677739 https://doi.org/10.1109/TIT.2017.2688381

[6] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

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

[8] BOURGAIN, J. and TZAFRIRI, L. (1987). Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.* **57** 137–224. MR0890420 https://doi.org/10.1007/BF02772174

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

[10] CAI, T. T., REN, Z. and ZHOU, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probab. Theory Related Fields* **156** 101–143. MR3055254 https://doi.org/10.1007/s00440-012-0422-7

[11] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

MSC2020 subject classifications. Primary 62H12; secondary 62F12.
Key words and phrases. Covariance estimation, quantization.

[12] CHEN, R. Y., GITTENS, A. and TROPP, J. A. (2012). The masked sample covariance estimator: An analysis using matrix concentration inequalities. *Inf. Inference* **1** 2–20. MR3311439 https://doi.org/10.1093/imaiai/ias001

[13] CHOI, J., MO, J. and HEATH, R. W. (2016). Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs. *IEEE Trans. Commun.* **64** 2005–2018.

[14] DAHMEN, J., KEYSERS, D., PITZ, M. and NEY, H. (2000). Structured covariance matrices for statistical image object recognition. In *Mustererkennung* 2000 99–106. Springer.

[15] DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** 1877–1892. MR1188046

[16] DIRKSEN, S. (2011). Noncommutative and vector-valued Rosenthal inequalities. PhD Thesis, Delft Univ. Technology.

[17] DIRKSEN, S. (2014). Itô isomorphisms for L^p-valued Poisson stochastic integrals. *Ann. Probab.* **42** 2595–2643. MR3265175 https://doi.org/10.1214/13-AOP906

[18] DIRKSEN, S. (2019). Quantized compressed sensing: A survey. In *Compressed Sensing and Its Applications. Appl. Numer. Harmon. Anal.* 67–95. Birkhäuser/Springer, Cham. MR4182528 https://doi.org/10.1007/978-3-319-73074-5_2

[19] DIRKSEN, S., MALY, J. and RAUHUT, H. (2022). Supplement to "Covariance estimation under one-bit quantization." https://doi.org/10.1214/22-AOS2239SUPP

[20] DIRKSEN, S. and MENDELSON, S. (2018). Robust one-bit compressed sensing with partial circulant matrices. arXiv:1812.06719.

[21] DIRKSEN, S. and MENDELSON, S. (2021). Non-Gaussian hyperplane tessellations and robust one-bit compressed sensing. *J. Eur. Math. Soc.* (*JEMS*) **23** 2913–2947. MR4275468 https://doi.org/10.4171/jems/1066

[22] DIRKSEN, S., MENDELSON, S. and STOLLENWERK, A. (2022). Sharp estimates on random hyperplane tessellations. arXiv:2201.05204.

[23] EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011 https://doi.org/10.1214/07-AOS559

[24] FURRER, R. and BENGTSSON, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.* **98** 227–255. MR2301751 https://doi.org/10.1016/j.jmva.2006.08.003

[25] GENZEL, M. and STOLLENWERK, A. (2020). Robust 1-bit compressed sensing via hinge loss minimization. *Inf. Inference* **9** 361–422. MR4108973 https://doi.org/10.1093/imaiai/iaz010

[26] GRAY, R. M. and NEUHOFF, D. L. (1998). Quantization. *IEEE Trans. Inf. Theory* **44** 2325–2383. MR1658787 https://doi.org/10.1109/18.720541

[27] GRAY, R. M. and STOCKHAM, T. G. (1993). Dithered quantizers. *IEEE Trans. Inf. Theory* **39** 805–812.

[28] HAGHIGHATSHOAR, S. and CAIRE, G. (2018). Low-complexity massive MIMO subspace estimation and tracking from low-dimensional projections. *IEEE Trans. Signal Process.* **66** 1832–1844. MR3797838 https://doi.org/10.1109/TSP.2018.2795560

[29] JACOVITTI, G. and NERI, A. (1994). Estimation of the autocorrelation function of complex Gaussian stationary processes by amplitude clipped signals. *IEEE Trans. Inf. Theory* **40** 239–245.

[30] JACQUES, L., LASKA, J. N., BOUFOUNOS, P. T. and BARANIUK, R. G. (2013). Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Trans. Inf. Theory* **59** 2082–2102. MR3043783 https://doi.org/10.1109/TIT.2012.2234823

[31] JOHNSON, C. R. (1990). *Matrix Theory and Applications* **40**. American Mathematical Soc.

[32] JUNG, H. C., MALY, J., PALZER, L. and STOLLENWERK, A. (2021). Quantized compressed sensing by rectified linear units. *IEEE Trans. Inf. Theory* **67** 4125–4149. MR4289368 https://doi.org/10.1109/TIT.2021.3070789

[33] KABANAVA, M. and RAUHUT, H. (2017). Masked Toeplitz covariance estimation. arXiv:1709.09377.

[34] KE, Y., MINSKER, S., REN, Z., SUN, Q. and ZHOU, W.-X. (2019). User-friendly covariance estimation for heavy-tailed distributions. *Statist. Sci.* **34** 454–471. MR4017523 https://doi.org/10.1214/19-STS711

[35] KNUDSON, K., SAAB, R. and WARD, R. (2016). One-bit compressive sensing with norm estimation. *IEEE Trans. Inf. Theory* **62** 2748–2758. MR3493876 https://doi.org/10.1109/TIT.2016.2527637

[36] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[37] KRIM, H. and VIBERG, M. (1996). Two decades of array signal processing research: The parametric approach. *IEEE Signal Process. Mag.* **13** 67–94.

[38] LEDOIT, O. and WOLF, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Finance* **10** 603–621.

[39] LEVINA, E. and VERSHYNIN, R. (2012). Partial estimation of covariance matrices. *Probab. Theory Related Fields* **153** 405–419. MR2948681 https://doi.org/10.1007/s00440-011-0349-4

[40] LI, Y., TAO, C., SECO-GRANADOS, G., MEZGHANI, A., SWINDLEHURST, A. L. and LIU, L. (2017). Channel estimation and performance analysis of one-bit massive MIMO systems. *IEEE Trans*. *Signal Process*. **65** 4075–4089. MR3684050 https://doi.org/10.1109/TSP.2017.2706179
[41] 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
[42] PLAN, Y. and VERSHYNIN, R. (2013). Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Trans*. *Inf*. *Theory* **59** 482–494. MR3008160 https://doi.org/10.1109/TIT.2012.2207945
[43] PLAN, Y. and VERSHYNIN, R. (2013). One-bit compressed sensing by linear programming. *Comm*. *Pure Appl*. *Math*. **66** 1275–1297. MR3069959 https://doi.org/10.1002/cpa.21442
[44] ROBERTS, L. (1962). Picture coding using pseudo-random noise. *IRE Trans*. *Inf*. *Theory* **8** 145–154.
[45] ROTH, K., MUNIR, J., MEZGHANI, A. and NOSSEK, J. A. (2015). Covariance based signal parameter estimation of coarse quantized signals. In 2015 *IEEE International Conference on Digital Signal Processing* (*DSP*) 19–23. IEEE.
[46] SCHOENBERG, I. J. (1942). Positive definite functions on spheres. *Duke Math*. *J*. **9** 96–108. MR0005922
[47] SRIVASTAVA, N. and VERSHYNIN, R. (2013). Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann*. *Probab*. **41** 3081–3111. MR3127875 https://doi.org/10.1214/12-AOP760
[48] TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found*. *Comput*. *Math*. **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z
[49] VAN VLECK, J. H. and MIDDLETON, D. (1966). The spectrum of clipped noise. *Proc*. *IEEE* **54** 2–19.
[50] VERSHYNIN, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *J*. *Theoret*. *Probab*. **25** 655–686. MR2956207 https://doi.org/10.1007/s10959-010-0338-z
[51] VERSHYNIN, R. (2018). *High-Dimensional Probability*: *An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

The Annals of Statistics
2022, Vol. 50, No. 6, 3563–3591
https://doi.org/10.1214/22-AOS2240

THE INTEGRATED COPULA SPECTRUM

BY YUICHI GOTO[1,a], TOBIAS KLEY[2,b], RIA VAN HECKE[3,c], STANISLAV VOLGUSHEV[4,e], HOLGER DETTE[3,d] AND MARC HALLIN[5,f]

[1]*Department of Mathematical Sciences, Kyushu University,* [a]*yuichi.goto@math.kyushu-u.ac.jp*
[2]*Institut für Mathematische Stochastik, Georg-August-Universität Göttingen,* [b]*tobias.kley@uni-goettingen.de*
[3]*Fakultät für Mathematik, Ruhr-Universität Bochum,* [c]*ria.vanhecke@rub.de,* [d]*holger.dette@rub.de*
[4]*Department of Statistical Sciences, University of Toronto,* [e]*stanislav.volgushev@utoroto.ca*
[5]*ECARES and Département de Mathématique, Université libre de Bruxelles,* [f]*mhallin@ulb.be*

Frequency domain methods form a ubiquitous part of the statistical toolbox for time-series analysis. In recent years, considerable interest has been given to the development of new spectral methodology and tools capturing dynamics in the entire joint distributions, and thus avoiding the limitations of classical, L^2-based spectral methods. Most of the spectral concepts proposed in that literature suffer from one major drawback, though: their estimation requires the choice of a smoothing parameter, which has a considerable impact on estimation quality and poses challenges for statistical inference. In this paper, associated with the concept of a copula-based spectrum, we introduce the notion of a *copula spectral distribution function* or *integrated copula spectrum*. This integrated copula spectrum retains the advantages of copula-based spectra but can be estimated without the need for smoothing parameters. We provide such estimators, along with a thorough theoretical analysis, based on a functional central limit theorem, of their asymptotic properties. We leverage these results to test various hypotheses that cannot be addressed by classical spectral methods, such as the lack of time reversibility or asymmetry in tail dynamics.

REFERENCES

ANDERSON, T. W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.* **21** 830–847. MR1232521 https://doi.org/10.1214/aos/1176349153

BARUNÍK, J. and KLEY, T. (2019). Quantile coherency: A general measure for dependence between cyclical economic variables. *Econom. J.* **22** 131–152. MR4021117 https://doi.org/10.1093/ectj/utz002

BEARE, B. K. and SEO, J. (2014). Time irreversible copula-based Markov models. *Econometric Theory* **30** 923–960. MR3252030 https://doi.org/10.1017/S0266466614000115

BIRR, S., KLEY, T. and VOLGUSHEV, S. (2019). Model assessment for time series dynamics using copula spectral densities: A graphical tool. *J. Multivariate Anal.* **172** 122–146. MR3958191 https://doi.org/10.1016/j.jmva.2019.03.003

BIRR, S., VOLGUSHEV, S., KLEY, T., DETTE, H. and HALLIN, M. (2017). Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1619–1643. MR3731679 https://doi.org/10.1111/rssb.12231

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375–390. MR0246478 https://doi.org/10.1093/biomet/56.2.375

BRILLINGER, D. R. and ROSENBLATT, M. (1967). Computation and interpretation of k-th order spectra. In *Spectral Analysis Time Series* (*Proc. Advanced Sem., Madison, Wis.*, 1966) 189–232. John Wiley, New York. MR0211567

CHEN, Y.-T., CHOU, R. Y. and KUAN, C.-M. (2000). Testing time reversibility without moment restrictions. *J. Econometrics* **95** 199–218.

DAHLHAUS, R. (1985). Asymptotic normality of spectral estimates. *J. Multivariate Anal.* **16** 412–431. MR0793500 https://doi.org/10.1016/0047-259X(85)90028-4

MSC2020 subject classifications. Primary 62M15, 62G30; secondary 62G10.
Key words and phrases. Copula, ranks, time series, frequency domain, time reversibility.

DAHLHAUS, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Process. Appl.* **30** 69–83. MR0968166 https://doi.org/10.1016/0304-4149(88)90076-2

DAVIS, R. A. and MIKOSCH, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15** 977–1009. MR2597580 https://doi.org/10.3150/09-BEJ213

DAVIS, R. A., MIKOSCH, T. and ZHAO, Y. (2013). Measures of serial extremal dependence and their estimation. *Stochastic Process. Appl.* **123** 2575–2602. MR3054537 https://doi.org/10.1016/j.spa.2013.03.014

DE GOOIJER, J. G. (2017). *Elements of Nonlinear Time Series Analysis and Forecasting. Springer Series in Statistics*. Springer, Cham. MR3642648 https://doi.org/10.1007/978-3-319-43252-6

DETTE, H., HALLIN, M., KLEY, T. and VOLGUSHEV, S. (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli* **21** 781–831. MR3338647 https://doi.org/10.3150/13-BEJ587

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

FUSION MEDIA LIMITED (2022). Investing.com—stock market quotes & financial news. Available at https://www.investing.com/, Accessed: 2022-06-18.

GHIL, M., ALLEN, M., DETTINGER, M., IDE, K., KONDRASHOV, D., MANN, M., ROBERTSON, A., SAUNDERS, A., TIAN, Y. et al. (2002). Advanced spectral methods for climate time series. *Rev. Geophys.* **2002** 1003–1043.

GOTO, Y., KLEY, T., VAN HECKE, R., VOLGUSHEV, S., DETTE, H. and HALLIN, M. (2022). Supplement to "The integrated copula spectrum." https://doi.org/10.1214/22-AOS2240SUPP

GRANGER, C. W. J. and HATANAKA, M. (2015). *Spectral Analysis of Economic Time Series*. Princeton Univ. Press, Princeton, NJ.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York. MR0084975

HAGEMANN, A. (2013). Robust Spectral Analysis. ArXiv E-prints. Available at arXiv:1111.1965v2.

HALLIN, M., LEFÈVRE, C. and PURI, M. L. (1988). On time-reversibility and the uniqueness of moving average representations for non-Gaussian stationary time series. *Biometrika* **75** 170–171. MR0932834 https://doi.org/10.1093/biomet/75.1.170

HAN, H., LINTON, O., OKA, T. and WHANG, Y.-J. (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. *J. Econometrics* **193** 251–270. MR3500187 https://doi.org/10.1016/j.jeconom.2016.03.001

HONG, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *J. Amer. Statist. Assoc.* **94** 1201–1220. MR1731483 https://doi.org/10.2307/2669935

HONG, Y. (2000). Generalized spectral tests for serial dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 557–574. MR1772415 https://doi.org/10.1111/1467-9868.00250

IBRAGIMOV, I. A. (1963). An estimate for the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8** 366–401.

JONDEAU, E. and ROCKINGER, M. (2003). Testing for differences in the tails of stock-market returns. *J. Empir. Finance* **10** 559–581.

KLEY, T. (2016). Quantile-based spectral analysis in an object-oriented framework and a reference implementation in R: The quantspec package. *J. Stat. Softw.* **70** 1–27.

KLEY, T., VOLGUSHEV, S., DETTE, H. and HALLIN, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli* **22** 1770–1807. MR3474833 https://doi.org/10.3150/15-BEJ711

KLÜPPELBERG, C. and MIKOSCH, T. (1996). The integrated periodogram for stable processes. *Ann. Statist.* **24** 1855–1879. MR1421152 https://doi.org/10.1214/aos/1069362301

KOENKER, R. and XIAO, Z. (2006). Quantile autoregression. *J. Amer. Statist. Assoc.* **101** 980–990. MR2324109 https://doi.org/10.1198/016214506000000672

KOKOSZKA, P. and MIKOSCH, T. (1997). The integrated periodogram for long-memory processes with finite or infinite variance. *Stochastic Process. Appl.* **66** 55–78. MR1431870 https://doi.org/10.1016/S0304-4149(96)00124-X

KRUPSKII, P. and JOE, H. (2019). Nonparametric estimation of multivariate tail probabilities and tail dependence coefficients. *J. Multivariate Anal.* **172** 147–161. MR3958192 https://doi.org/10.1016/j.jmva.2019.02.013

LANGE, H., BRUNTON, S. and KUTZ, N. (2019). Spectral methods for time series prediction with application to fluid flows. In *APS Division of Fluid Dynamics Meeting Abstracts* Q41.008. APS Meeting Abstracts.

LEE, J. and RAO, S. S. (2012). The probabilistic spectral density. Personal communication.

LI, T.-H. (2008). Laplace periodogram for time series analysis. *J. Amer. Statist. Assoc.* **103** 757–768. MR2435471 https://doi.org/10.1198/016214508000000265

LI, T.-H. (2012). Quantile periodograms. *J. Amer. Statist. Assoc.* **107** 765–776. MR2980083 https://doi.org/10.1080/01621459.2012.682815

LI, T.-H. (2014). *Time Series with Mixed Spectra*. CRC Press, Boca Raton, FL. MR3098581

LI, T.-H. (2021). Quantile-frequency analysis and spectral measures for diagnostic checks of time series with nonlinear dynamics. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 270–290. MR4226668 https://doi.org/10.1111/rssc.12458

LI, B. and GENTON, M. G. (2013). Nonparametric identification of copula structures. *J. Amer. Statist. Assoc.* **108** 666–675. MR3174650 https://doi.org/10.1080/01621459.2013.787083

LIKKASON, O. (2011). *Spectral Analysis of Geophysical Data*.

LIM, Y. and OH, H.-S. (2021). Quantile spectral analysis of long-memory processes. *Empir. Econ.* 1–22.

LINTON, O. and WHANG, Y.-J. (2007). The quantilogram: With an application to evaluating directional predictability. *J. Econometrics* **141** 250–282. MR2411744 https://doi.org/10.1016/j.jeconom.2007.01.004

MANGOLD, B. (2017). New concepts of symmetry for copulas. Technical report, FAU Discussion Papers in Economics.

MIKOSCH, T. and NORVAIŠA, R. (1997). Uniform convergence of the empirical spectral distribution function. *Stochastic Process. Appl.* **70** 85–114. MR1472960 https://doi.org/10.1016/S0304-4149(97)00053-7

NELSEN, R. B. (1993). Some concepts of bivariate symmetry. *J. Nonparametr. Stat.* **3** 95–101. MR1272164 https://doi.org/10.1080/10485259308832574

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

NEUMANN, M. H. and PAPARODITIS, E. (2008). Simultaneous confidence bands in spectral density estimation. *Biometrika* **95** 381–397. MR2422696 https://doi.org/10.1093/biomet/asn005

PAPARODITIS, E. and POLITIS, D. N. (2002). The local bootstrap for Markov processes. *J. Statist. Plann. Inference* **108** 301–328.

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series. Vol.* 1. *Probability and Mathematical Statistics*. Academic Press, London. MR0628735

RAMSEY, J. B. and ROTHMAN, P. (1996). Time irreversibility and business cycle asymmetry. *J. Money Credit Bank.* **28** 1–21.

ROSCO, J. F. and JOE, H. (2013). Measures of tail asymmetry for bivariate copulas. *Statist. Papers* **54** 709–726. MR3072896 https://doi.org/10.1007/s00362-012-0457-y

SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 https://doi.org/10.3150/11-BEJ387

SO, M. K. P. and CHAN, R. K. S. (2014). Bayesian analysis of tail asymmetry based on a threshold extreme value model. *Comput. Statist. Data Anal.* **71** 568–587. MR3131990 https://doi.org/10.1016/j.csda.2013.02.008

SU, X., ZHAN, W. and LI, Y. (2021). Quantile dependence between investor attention and cryptocurrency returns: Evidence from time and frequency domain analyses. *Appl. Econ.*

VON SACHS, R. (2020). Nonparametric spectral analysis of multivariate time series. *Annu. Rev. Stat. Appl.* **7** 361–386. MR4104197 https://doi.org/10.1146/annurev-statistics-031219-041138

WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 https://doi.org/10.1239/jap/1082999076

The Annals of Statistics
2022, Vol. 50, No. 6, 3592–3618
https://doi.org/10.1214/22-AOS2241

COMPARISON OF MARKOV CHAINS VIA WEAK POINCARÉ INEQUALITIES WITH APPLICATION TO PSEUDO-MARGINAL MCMC

BY CHRISTOPHE ANDRIEU[a], ANTHONY LEE[b], SAM POWER[c] AND ANDI Q. WANG[d]

School of Mathematics, University of Bristol, [a]*C.Andrieu@bristol.ac.uk,* [b]*anthony.lee@bristol.ac.uk,* [c]*sam.power@bristol.ac.uk,* [d]*andi.wang@bristol.ac.uk*

We investigate the use of a certain class of functional inequalities known as weak Poincaré inequalities to bound convergence of Markov chains to equilibrium. We show that this enables the straightforward and transparent derivation of subgeometric convergence bounds for methods such as the Independent Metropolis–Hastings sampler and pseudo-marginal methods for intractable likelihoods, the latter being subgeometric in many practical settings. These results rely on novel quantitative comparison theorems between Markov chains. Associated proofs are simpler than those relying on drift/minorisation conditions and the tools developed allow us to recover and further extend known results as particular cases. We are then able to provide new insights into the practical use of pseudo-marginal algorithms, analyse the effect of averaging in Approximate Bayesian Computation (ABC) and the use of products of independent averages and also to study the case of log-normal weights relevant to particle marginal Metropolis–Hastings (PMMH).

REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

[2] ANDRIEU, C., LEE, A., POWER, S. and WANG, A. Q. (2022). Supplement to "Comparison of Markov chains via weak Poincaré inequalities with application to pseudo-marginal MCMC." https://doi.org/10.1214/22-AOS2241SUPP

[3] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

[4] ANDRIEU, C. and VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762 https://doi.org/10.1214/14-AAP1022

[5] BAKRY, D., BOLLEY, F. and GENTIL, I. (2010). Around Nash inequalities. In *Journées équations aux dérivées partielles* 1–16. Available at www.numdam.org/item/10.5802/jedp.59.pdf. https://doi.org/10.5802/jedp.59

[6] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160 https://doi.org/10.1016/j.jfa.2007.11.002

[7] BÉRARD, J., DEL MORAL, P. and DOUCET, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.* **19** no. 94, 28. MR3272327 https://doi.org/10.1214/EJP.v19-3428

[8] BOBKOV, S. G. and LEDOUX, M. (2009). Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* **37** 403–427. MR2510011 https://doi.org/10.1214/08-AOP407

[9] BORNN, L., PILLAI, N., SMITH, A. and WOODARD, D. (2014). One pseudo-sample is enough in approximate Bayesian computation MCMC. *Biometrika* **99** 1–10.

[10] CARACCIOLO, S., PELISSETTO, A. and SOKAL, A. D. (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.* **60** 1–53. MR1063212 https://doi.org/10.1007/BF01013668

MSC2020 subject classifications. Primary 65C40, 65C05; secondary 62J10.
Key words and phrases. Markov chain Monte Carlo, weak Poincaré inequality, subgeometric convergence, pseudo-marginal.

[11] CATTIAUX, P., GOZLAN, N., GUILLIN, A. and ROBERTO, C. (2010). Functional inequalities for heavy tailed distributions and application to isoperimetry. *Electron. J. Probab.* **15** 346–385. MR2609591 https://doi.org/10.1214/EJP.v15-754

[12] DELIGIANNIDIS, G. and LEE, A. (2018). Which ergodic averages have finite asymptotic variance? *Ann. Appl. Probab.* **28** 2309–2334. MR3843830 https://doi.org/10.1214/17-AAP1358

[13] DIACONIS, P. and SALOFF-COSTE, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Probab.* **9** 459–510. MR1385408 https://doi.org/10.1007/BF02214660

[14] DONOHUE, M. C., OVERHOLSER, R., XU, R. and VAIDA, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika* **98** 685–700. MR2836414 https://doi.org/10.1093/biomet/asr023

[15] DOUC, R., MOULINES, E., PRIOURET, P. and SOULIER, P. (2018). *Markov Chains. Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

[16] DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

[17] FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87. MR1097464

[18] GÅSEMYR, J. (2006). The spectrum of the independent Metropolis–Hastings algorithm. *J. Theoret. Probab.* **19** 152–165. MR2256484 https://doi.org/10.1007/s10959-006-0009-2

[19] GREENBERG, E. R., BARON, J. A., STUKEL, T. A., STEVENS, M. M., MANDEL, J. S., SPENCER, S. K. et al. and SKIN CANCER PREVENTION STUDY GROUP (1990). A clinical trial of beta carotene to prevent basal-cell and squamous-cell cancers of the skin. *N. Engl. J. Med.* **323** 789–795.

[20] JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063 https://doi.org/10.1214/aoap/1015961162

[21] JARNER, S. F. and ROBERTS, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939 https://doi.org/10.1111/j.1467-9469.2007.00557.x

[22] KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Ann. Statist.* **39** 2585–2606. MR2906879 https://doi.org/10.1214/11-AOS916

[23] KONTOYIANNIS, I. and MEYN, S. P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13** 304–362. MR1952001 https://doi.org/10.1214/aoap/1042765670

[24] LEE, A. and ŁATUSZYŃSKI, K. (2014). Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika* **101** 655–671. MR3254907 https://doi.org/10.1093/biomet/asu027

[25] LEE, A., TIBERI, S. and ZANELLA, G. (2019). Unbiased approximations of products of expectations. *Biometrika* **106** 708–715. MR3992400 https://doi.org/10.1093/biomet/asz008

[26] MEYN, S. P. and TWEEDIE, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

[27] PECCOUD, J. and YCART, B. (1995). Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48** 222–234.

[28] ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

[29] RÖCKNER, M. and WANG, F.-Y. (2001). Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal.* **185** 564–603. MR1856277 https://doi.org/10.1006/jfan.2001.3776

[30] ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[31] SHERLOCK, C., THIERY, A. H. and LEE, A. (2017). Pseudo-marginal Metropolis–Hastings sampling using averages of unbiased estimators. *Biometrika* **104** 727–734. MR3694593 https://doi.org/10.1093/biomet/asx031

[32] SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. and ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278

[33] TIBERI, S., WALSH, M., CAVALLARO, M., HEBENSTREIT, D. and FINKENSTÄDT, B. (2018). Bayesian inference on stochastic gene transcription from flow cytometry data. *Bioinformatics* **34** i647–i655.

[34] TRAN, M. N., KOHN, R., QUIROZ, M. and VILLANI, M. (2016). The block pseudo-marginal sampler. arXiv preprint arXiv:1603.02485.

The Annals of Statistics
2022, Vol. 50, No. 6, 3619–3646
https://doi.org/10.1214/22-AOS2242

ON OPTIMAL BLOCK RESAMPLING FOR GAUSSIAN-SUBORDINATED LONG-RANGE DEPENDENT PROCESSES

BY QIHAO ZHANG[1,a], SOUMENDRA N. LAHIRI[2,c] AND DANIEL J. NORDMAN[1,b]

[1]*Department of Statistics, Iowa State University,* [a]*qihaoz@iastate.edu,* [b]*dnordman@iastate.edu*
[2]*Department of Mathematics and Statistics, Washington University in St. Louis,* [c]*s.lahiri@wustl.edu*

Block-based resampling estimators have been intensively investigated for weakly dependent time processes, which has helped to inform implementation (e.g., best block sizes). However, little is known about resampling performance and block sizes under strong or long-range dependence. To establish guideposts in block selection, we consider a broad class of strongly dependent time processes, formed by a transformation of a stationary long-memory Gaussian series, and examine block-based resampling estimators for the variance of the prototypical sample mean; extensions to general statistical functionals are also considered. Unlike weak dependence, the properties of resampling estimators under strong dependence are shown to depend intricately on the nature of nonlinearity in the time series (beyond Hermite ranks) in addition to the long-memory coefficient and block size. Additionally, the intuition has often been that optimal block sizes should be larger under strong dependence (say $O(n^{1/2})$ for a sample size n) than the optimal order $O(n^{1/3})$ known under weak dependence. This intuition turns out to be largely incorrect, though a block order $O(n^{1/2})$ may be reasonable (and even optimal) in many cases, owing to nonlinearity in a long-memory time series. While optimal block sizes are more complex under long-range dependence compared to short-range, we provide a consistent data-driven rule for block selection. Numerical studies illustrate that the guides for block selection perform well in other block-based problems with long-memory time series, such as distribution estimation and strategies for testing Hermite rank.

REFERENCES

[1] ABADIR, K. M., DISTASO, W. and GIRAITIS, L. (2009). Two estimators of the long-run variance: Beyond short memory. *J. Econometrics* **150** 56–70. MR2525994 https://doi.org/10.1016/j.jeconom.2009.02.010

[2] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Stat.* **23** 193–212. MR0050238 https://doi.org/10.1214/aoms/1177729437

[3] ANDREWS, D. W. K. and SUN, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* **72** 569–614. MR2036733 https://doi.org/10.1111/j.1468-0262.2004.00501.x

[4] BAI, S. and TAQQU, M. S. (2017). On the validity of resampling methods under long memory. *Ann. Statist.* **45** 2365–2399. MR3737895 https://doi.org/10.1214/16-AOS1524

[5] BAI, S. and TAQQU, M. S. (2018). How the instability of ranks under long memory affects large-sample inference. *Statist. Sci.* **33** 96–116. MR3757507 https://doi.org/10.1214/17-STS633

[6] BAI, S., TAQQU, M. S. and ZHANG, T. (2016). A unified approach to self-normalized block sampling. *Stochastic Process. Appl.* **126** 2465–2493. MR3505234 https://doi.org/10.1016/j.spa.2016.02.007

[7] BERAN, J. (1991). *M* estimators of location for Gaussian and related processes with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86** 704–708. MR1147095

[8] BERAN, J. (2010). Long-range dependence. *Wiley Ser. Comput. Stat.* **2** 26–35.

[9] BERAN, J., MÖHRLE, S. and GHOSH, S. (2016). Testing for Hermite rank in Gaussian subordination processes. *J. Comput. Graph. Statist.* **25** 917–934. MR3533645 https://doi.org/10.1080/10618600.2015.1056345

MSC2020 subject classifications. Primary 62G09; secondary 62G20, 62M10.
Key words and phrases. Block length, block bootstrap, long-memory, subsampling, variance estimation.

[10] BETKEN, A. and WENDLER, M. (2018). Subsampling for general statistics under long range dependence with application to change point analysis. *Statist. Sinica* **28** 1199–1224. MR3821001

[11] BREIDT, F. J., CRATO, N. and DE LIMA, P. (1998). The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83** 325–348. MR1613867 https://doi.org/10.1016/S0304-4076(97)00072-9

[12] BÜHLMAN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31** 295–310.

[13] CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813 https://doi.org/10.1214/aos/1176350057

[14] DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312 https://doi.org/10.1214/aos/1176347394

[15] DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52. MR0550122 https://doi.org/10.1007/BF00535673

[16] FERNHOLZ, L. T. (2012). *Von Mises Calculus for Statistical Functionals*. Springer Science & Business Media. MR0713611 https://doi.org/10.1007/978-1-4612-5604-5

[17] FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704 https://doi.org/10.1214/09-AOS735

[18] GIRAITIS, L., ROBINSON, P. M. and SURGAILIS, D. (1999). Variance-type estimation of long memory. *Stochastic Process. Appl.* **80** 1–24. MR1670119 https://doi.org/10.1016/S0304-4149(98)00062-3

[19] GIRAITIS, L. and TAQQU, M. S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *Ann. Statist.* **27** 178–203. MR1701107 https://doi.org/10.1214/aos/1018031107

[20] GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1** 15–29. MR0605572 https://doi.org/10.1111/j.1467-9892.1980.tb00297.x

[21] HALL, P., HOROWITZ, J. L. and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282 https://doi.org/10.1093/biomet/82.3.561

[22] HALL, P. and JING, B. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc. Ser. B* **58** 727–737. MR1410187

[23] HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* **8** 1189–1204. MR1666245

[24] HÖSSJER, O. and MIELNICZUK, J. (1995). Delta method for long-range dependent observations. *J. Nonparametr. Stat.* **5** 75–82. MR1342943 https://doi.org/10.1080/10485259508832635

[25] HURVICH, C. M., DEO, R. and BRODSKY, J. (1998). The mean squared error of Geweke and Porter–Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Series Anal.* **19** 19–46. MR1624096 https://doi.org/10.1111/1467-9892.00075

[26] JACH, A., MCELROY, T. S. and POLITIS, D. N. (2016). Corrigendum to 'Subsampling inference for the mean of heavy-tailed long-memory time series' [MR2877610]. *J. Time Series Anal.* **37** 713–720. MR3536227 https://doi.org/10.1111/jtsa.12191

[27] KIM, Y. M. and NORDMAN, D. J. (2011). Properties of a block bootstrap under long-range dependence. *Sankhya A* **73** 79–109. MR2887088 https://doi.org/10.1007/s13171-011-0003-3

[28] KREISS, J.-P. and LAHIRI, S. N. (2012). Bootstrap methods for time series. In *Handbook of Statist.* **30** 3–26. Elsevier.

[29] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

[30] LAHIRI, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18** 405–413. MR1247453 https://doi.org/10.1016/0167-7152(93)90035-H

[31] LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117 https://doi.org/10.1214/aos/1018031117

[32] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

[33] LAHIRI, S. N., FURUKAWA, K. and LEE, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Stat. Methodol.* **4** 292–321. MR2380557 https://doi.org/10.1016/j.stamet.2006.08.002

[34] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 225–248. Wiley, New York. MR1197787

[35] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. MR0242239 https://doi.org/10.1137/1010093

[36] MONTANARI, A., TAQQU, M. S. and TEVEROVSKY, V. (1999). Estimating long-range dependence in the presence of periodicity: An empirical study. *Math. Comput. Modelling* **29** 217–228.

[37] NORDMAN, D. J. and LAHIRI, S. N. (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* **21** 1087–1111. MR2200986 https://doi.org/10.1017/S0266466605050541

[38] NORDMAN, D. J. and LAHIRI, S. N. (2014). Convergence rates of empirical block length selectors for block bootstrap. *Bernoulli* **20** 958–978. MR3178523 https://doi.org/10.3150/13-BEJ511

[39] PAPARODITIS, E. and POLITIS, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.* **5** 131–148. MR1909304 https://doi.org/10.1111/1368-423X.t01-1-00077

[40] POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.* **18** 219–230. MR2026081 https://doi.org/10.1214/ss/1063994977

[41] POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770

[42] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

[43] POLITIS, D. N. and WHITE, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534 https://doi.org/10.1081/ETC-120028836

[44] ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214 https://doi.org/10.1214/aos/1176324636

[45] ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301 https://doi.org/10.1214/aos/1176324317

[46] ROBINSON, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* **21** 171–180. MR2153861 https://doi.org/10.1017/S0266466605050115

[47] SHAO, J. (2003). *Mathematical Statistics*. Springer.

[48] SONG, W. T. and SCHMEISER, B. W. (1995). Optimal mean-squared-error batch sizes. *Manage. Sci.* **41** 110–123.

[49] TAQQU, M. S. (1974/75). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329 https://doi.org/10.1007/BF00532868

[50] TAQQU, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahrsch. Verw. Gebiete* **40** 203–238. MR0471045 https://doi.org/10.1007/BF00736047

[51] TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83. MR0550123 https://doi.org/10.1007/BF00535674

[52] TAQQU, M. S. (2002). Fractional Brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.). Springer Science & Business Media. MR1956042

[53] TEVEROVSKY, V. and TAQQU, M. (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *J. Time Series Anal.* **18** 279–304. MR1456641 https://doi.org/10.1111/1467-9892.00050

[54] WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027 https://doi.org/10.2307/1912934

[55] ZHANG, Q., LAHIRI, S. N. and NORDMAN, D. J. (2022). Supplement to "On optimal block resampling for Gaussian-subordinated long-range dependent processes." https://doi.org/10.1214/22-AOS2242SUPPA, https://doi.org/10.1214/22-AOS2242SUPPB

[56] ZHANG, T., HO, H.-C., WENDLER, M. and WU, W. B. (2013). Block sampling under strong dependence. *Stochastic Process. Appl.* **123** 2323–2339. MR3038507 https://doi.org/10.1016/j.spa.2013.02.006

The Annals of Statistics
2022, Vol. 50, No. 6, 3647–3676
https://doi.org/10.1214/22-AOS2245

THE STEIN EFFECT FOR FRÉCHET MEANS

BY ANDREW MCCORMACK[a] AND PETER HOFF[b]

Department of Statistical Science, Duke University, [a]*andrew.mccormack@duke.edu,* [b]*peter.hoff@duke.edu*

The Fréchet mean is a useful description of location for a probability distribution on a metric space that is not necessarily a vector space. This article considers simultaneous estimation of multiple Fréchet means from a decision-theoretic perspective, and in particular, the extent to which the unbiased estimator of a Fréchet mean can be dominated by a generalization of the James–Stein shrinkage estimator. It is shown that if the metric space satisfies a nonpositive curvature condition, then this generalized James–Stein estimator asymptotically dominates the unbiased estimator as the dimension of the space grows. These results hold for a large class of distributions on a variety of spaces, including Hilbert spaces and, therefore, partially extend known results on the applicability of the James–Stein estimator to nonnormal distributions on Euclidean spaces. Simulation studies on phylogenetic trees and symmetric positive definite matrices are presented, numerically demonstrating the efficacy of this generalized James–Stein estimator.

REFERENCES

[1] ÅKERBORG, Ö., SENNBLAD, B., ARVESTAD, L. and LAGERGREN, J. (2009). Simultaneous Bayesian gene tree reconstruction and reconciliation analysis. *Proc. Natl. Acad. Sci. USA* **106** 5714–5719.

[2] ALEKSANDROV, A. D. (1951). A theorem on triangles in a metric space and some of its applications. *Tr. Mat. Inst. Steklova* **38** 5–23. MR0049584

[3] ALEXANDER, S., KAPOVITCH, V. and PETRUNIN, A. (2019). *An Invitation to Alexandrov Geometry. CAT*(0) *Spaces. SpringerBriefs in Math.* **22**. Springer, Cham. MR3930625 https://doi.org/10.1007/978-3-030-05312-3

[4] ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. MR2288028 https://doi.org/10.1137/050637996

[5] BAČÁK, M. (2014). *Convex Analysis and Optimization in Hadamard Spaces. De Gruyter Series in Nonlinear Analysis and Applications* **22**. de Gruyter, Berlin. MR3241330 https://doi.org/10.1515/9783110361629

[6] BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Stat.* **41** 642–645. MR0253461 https://doi.org/10.1214/aoms/1177697104

[7] BERAN, R. (2010). The unbearable transparency of Stein estimation. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis*: *A Festschrift in Honor of Professor Jana Jurečková. Inst. Math. Stat.* (*IMS*) *Collect.* **7** 25–34. IMS, Beachwood, OH. MR2808363

[8] BERGER, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328. MR0386080

[9] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611 https://doi.org/10.1007/978-1-4757-4286-2

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

[11] BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

[12] BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1991). Generalizations of James–Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650. MR1126343 https://doi.org/10.1214/aos/1176348267

[13] BRANDWEIN, A. C. and STRAWDERMAN, W. E. (2012). Stein estimation for spherically symmetric distributions: Recent developments. *Statist. Sci.* **27** 11–23. MR2953492 https://doi.org/10.1214/10-STS323

MSC2020 subject classifications. Primary 62R20; secondary 62C15.
Key words and phrases. Admissibility, empirical Bayes, Hadamard space, nonparametric, shrinkage.

[14] BRIDSON, M. R. and HAEFLIGER, A. (1999). *Metric Spaces of Non-positive Curvature*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **319**. Springer, Berlin. MR1744486 https://doi.org/10.1007/978-3-662-12494-9

[15] BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Stat.* **37** 1087–1136. MR0216647 https://doi.org/10.1214/aoms/1177699259

[16] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. MR0286209 https://doi.org/10.1214/aoms/1177693318

[17] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418 https://doi.org/10.1090/gsm/033

[18] CARMICHAEL, O., CHEN, J., PAUL, D. and PENG, J. (2013). Diffusion tensor smoothing through weighted Karcher means. *Electron. J. Stat.* **7** 1913–1956. MR3084676 https://doi.org/10.1214/13-EJS825

[19] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

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

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

[22] DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. MR4084166

[23] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[24] EFRON, B. and MORRIS, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59** 335–347. MR0334386 https://doi.org/10.1093/biomet/59.2.335

[25] EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

[26] FELSENSTEIN, J. and FELENSTEIN, J. (2004). *Inferring Phylogenies* **2**. Sinauer Associates, Sunderland, MA.

[27] FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

[28] FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *J. Multivariate Anal.* **85** 24–39. MR1978175 https://doi.org/10.1016/S0047-259X(02)00023-4

[29] FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (2018). *Shrinkage Estimation*. *Springer Series in Statistics*. Springer, Cham. MR3887633 https://doi.org/10.1007/978-3-030-02185-6

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

[31] GARYFALLIDIS, E., BRETT, M., AMIRBEKIAN, B., ROKEM, A., VAN DER WALT, S., DESCOTEAUX, M., NIMMO-SMITH, I. and DIPY CONTRIBUTORS (2014). DIPY, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **8** 8. https://doi.org/10.3389/fninf.2014.00008

[32] GINESTET, C. E. (2012). Strong consistency of Fréchet sample mean sets for graph-valued random variables. Preprint. Available at arXiv:1204.3183.

[33] GUTMANN, S. (1982). Stein's paradox is impossible in problems with finite sample space. *Ann. Statist.* **10** 1017–1020. MR0663454

[34] GUTMANN, S. (1984). Decisions immune to Stein's effect. *Sankhyā Ser. A* **46** 186–194. MR0778869

[35] HAFF, L. R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19** 1163–1190. MR1126320 https://doi.org/10.1214/aos/1176348244

[36] HOTZ, T., HUCKEMANN, S., LE, H., MARRON, J. S., MATTINGLY, J. C., MILLER, E., NOLEN, J., OWEN, M., PATRANGENARU, V. et al. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23** 2238–2258. MR3127934 https://doi.org/10.1214/12-AAP899

[37] HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484. MR0467991

[38] HUELSENBECK, J. P. (1995). Performance of phylogenetic methods in simulation. *Syst. Biol.* **44** 17–48.

[39] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

[40] JUKES, T. H. and CANTOR, C. R. (1969). Evolution of protein molecules. *Mammalian Prot. Metab.* **3** 21–132.

[41] KNEIP, A. (1994). Ordered linear smoothers. *Ann. Statist.* **22** 835–866. MR1292543 https://doi.org/10.1214/aos/1176325498

[42] KUBOKAWA, T. (1998). The Stein phenomenon in simultaneous estimation: A review. In *Applied Statistical Science, III* 143–173. Nova Sci. Publ., Commack, NY. MR1673649

[43] KUBOKAWA, T. and SRIVASTAVA, M. S. (1999). Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. *Ann. Statist.* **27** 600–609. MR1714715 https://doi.org/10.1214/aos/1018031209

[44] LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

[45] LEE, J. M. (2018). *Introduction to Riemannian Manifolds*. *Graduate Texts in Mathematics* **176**. Springer, Cham. MR3887684

[46] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875

[47] LI, K.-C. and HWANG, J. T. (1984). The data-smoothing aspect of Stein estimates. *Ann. Statist.* **12** 887–897. MR0751280 https://doi.org/10.1214/aos/1176346709

[48] LIU, X., LIU, L. and HU, J. (2017). James-Stein estimation problem for a multivariate normal random matrix and an improved estimator. *Linear Algebra Appl.* **532** 231–256. MR3688639 https://doi.org/10.1016/j.laa.2017.06.032

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

[50] MADDISON, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46** 523–536.

[51] MARCHAND, E. and STRAWDERMAN, W. E. (2004). Estimation in restricted parameter spaces: A review. In *A Festschrift for Herman Rubin*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 21–44. IMS, Beachwood, OH. MR2126884 https://doi.org/10.1214/lnms/1196285377

[52] MCCORMACK, A. and HOFF, P. (2022). Supplement A to "The Stein effect for Fréchet means": Proofs. https://doi.org/10.1214/22-AOS2245SUPPA

[53] MCCORMACK, A. and HOFF, P. (2022). Supplement B to "The Stein effect for Fréchet means": Counterexamples, numerical results and algorithms. https://doi.org/10.1214/22-AOS2245SUPPB

[54] MILLER, E., OWEN, M. and PROVAN, J. S. (2015). Polyhedral computational geometry for averaging metric phylogenetic trees. *Adv. in Appl. Math.* **68** 51–91. MR3345895 https://doi.org/10.1016/j.aam.2015.04.002

[55] PATRANGENARU, V. and ELLINGSON, L. (2016). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL. MR3444169

[56] PENNEC, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vision* **25** 127–154. MR2254442 https://doi.org/10.1007/s10851-006-6228-4

[57] PENNEC, X., FILLARD, P., AYACHE, N. and EPIDAURE, P. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66. https://doi.org/10.1007/s11263-005-3222-z

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

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

[60] RASMUSSEN, M. D. and KELLIS, M. (2011). A Bayesian approach for fast and accurate gene tree reconstruction. *Mol. Biol. Evol.* **28** 273–290. https://doi.org/10.1093/molbev/msq189

[61] SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618

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

[63] SHAO, P. Y.-S. and STRAWDERMAN, W. E. (1994). Improving on the James–Stein positive-part estimator. *Ann. Statist.* **22** 1517–1538. MR1311987 https://doi.org/10.1214/aos/1176325640

[64] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley–Los Angeles, CA. MR0084922

[65] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[66] STURM, K.-T. (2002). Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Ann. Probab.* **30** 1195–1222. MR1920105 https://doi.org/10.1214/aop/1029867125

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

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

[69] SZÖLLŐSI, G. J., TANNIER, E., DAUBIN, V. and BOUSSAU, B. (2015). The inference of gene trees with species trees. *Syst. Biol.* **64** e42–e62. https://doi.org/10.1093/sysbio/syu048

[70] TABELOW, K., POLZEHL, J., SPOKOINY, V. and VOSS, H. U. (2008). Diffusion tensor imaging: Structural adaptive smoothing. *NeuroImage* **39** 1763–1773. https://doi.org/10.1016/j.neuroimage.2007.10.024

[71] TSUKUMA, H. and KUBOKAWA, T. (2007). Methods for improvement in estimation of a normal mean matrix. *J. Multivariate Anal.* **98** 1592–1610. MR2370109 https://doi.org/10.1016/j.jmva.2007.04.009

[72] XIE, X., KOU, S. C. and BROWN, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *J. Amer. Statist. Assoc.* **107** 1465–1479. MR3036408 https://doi.org/10.1080/01621459.2012.728154

[73] YANG, C. and VEMURI, B. C. (2019). Shrinkage estimation on the manifold of symmetric positive-definite matrices with applications to neuroimaging. In *International Conference on Information Processing in Medical Imaging* 566–578. Springer, Berlin.

[74] YANG, C. and VEMURI, B. C. (2020). Shrinkage estimation of the Frechet mean in Lie groups. Preprint. Available at arXiv:2009.13020.

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

VOLUME 50

2022

CONTENTS OF VOLUME 50

Articles

Erratum

Correction note