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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 45, No. 4—August 2017

The Annals of Statistics
2017, Vol. 45, No. 4, 1375–1402
DOI: 10.1214/16-AOS1479

ON THE OPTIMALITY OF BAYESIAN CHANGE-POINT DETECTION

BY DONG HAN[1*], FUGEE TSUNG[2†] AND JINGUO XIAN[1*]

Shanghai Jiao Tong University and*
Hong Kong University of Science and Technology†

By introducing suitable loss random variables of detection, we obtain optimal tests in terms of the stopping time or alarm time for Bayesian change-point detection not only for a general prior distribution of change-points but also for observations being a Markov process. Moreover, the optimal (minimal) average detection delay is proved to be equal to 1 for any (possibly large) average run length to false alarm if the number of possible change-points is finite.

REFERENCES

[1] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1991). *The Theory of Optimal Stopping*. Dover, New York. MR1105325
[2] HAN, D. and TSUNG, F. (2004). A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes. *Ann. Statist.* **32** 316–339. MR2051010
[3] HAN, D., TSUNG, F. and XIAN, J. (2017). Supplement to "On the optimality of Bayesian change-point detection." DOI:10.1214/16-AOS1479SUPP.
[4] LAWLER, G. F. and LIMIC, V. (2010). *Random Walk: A Modern Introduction*. *Cambridge Studies in Advanced Mathematics* **123**. Cambridge Univ. Press, Cambridge. MR2677157
[5] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. MR0309251
[6] MOUSTAKIDES, G. V. (2008). Sequential change detection revisited. *Ann. Statist.* **36** 787–807. MR2396815
[7] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. MR0773162
[8] POLLAK, M. and TARTAKOVSKY, A. G. (2009). Optimality properties of the Shiryaev–Roberts procedure. *Statist. Sinica* **19** 1729–1739. MR2589207
[9] POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2010). On optimality of the Shiryaev–Roberts procedure for detecting a change in distribution. *Ann. Statist.* **38** 3445–3457. MR2766858
[10] POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2012). State-of-the-art in sequential change-point detection. *Methodol. Comput. Appl. Probab.* **14** 649–684. MR2966314
[11] POOR, H. V. and HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge Univ. Press, Cambridge. MR2482527
[12] ROBERTS, W. S. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430. MR0196887
[13] SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **13** 22–46.

MSC2010 subject classifications. Primary 62L10; secondary 62L15.
Key words and phrases. Optimal test, Bayesian change-point detection, Markov process.

[14] SHIRYAEV, A. N. and ZRYUMOV, P. Y. (2009). On the linear and nonlinear generalized Bayesian disorder problem (discrete time case). In *Optimality and Risk—Modern Trends in Mathematical Finance* 227–235. Springer, Berlin. MR2648606

[15] SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York. MR0468067

[16] TARTAKOVSKIĬ, A. G. (2008). Asymptotic optimality in Bayesian quickest detection problems under a global false alarm probability constraint. *Theory Probab. Appl*. **53** 443–466.

[17] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis*: *Hypothesis Testing and Changepoint Detection*. *Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. MR3241619

[18] TARTAKOVSKY, A. G., POLLAK, M. and POLUNCHENKO, A. S. (2012). Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures. *Theory Probab. Appl*. **56** 457–484. MR3136464

[19] TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2005). General asymptotic Bayesian theory of quickest change detection. *Theory Probab. Appl*. **49** 458–497. MR2144868

[20] YAKIR, B. (1994). Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-Point Problems* (*South Hadley*, *MA*, 1992) (E. Carlstein, H. Muller and D. Siegmund, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23** 346–358. IMS, Hayward, CA. MR1477935

The Annals of Statistics
2017, Vol. 45, No. 4, 1403–1430
DOI: 10.1214/16-AOS1488

COMPUTATIONAL AND STATISTICAL BOUNDARIES FOR SUBMATRIX LOCALIZATION IN A LARGE NOISY MATRIX

BY T. TONY CAI[1], TENGYUAN LIANG AND ALEXANDER RAKHLIN[2]

University of Pennsylvania

We study in this paper computational and statistical boundaries for submatrix *localization*. Given one observation of (one or multiple nonoverlapping) signal submatrix (of magnitude λ and size $k_m \times k_n$) embedded in a large noise matrix (of size $m \times n$), the goal is to optimal identify the support of the signal submatrix computationally and statistically.

Two transition thresholds for the signal-to-noise ratio λ/σ are established in terms of m, n, k_m and k_n. The first threshold, SNR_c, corresponds to the computational boundary. We introduce a new linear time spectral algorithm that identifies the submatrix with high probability when the signal strength is above the threshold SNR_c. Below this threshold, it is shown that no polynomial time algorithm can succeed in identifying the submatrix, under the *hidden clique hypothesis*. The second threshold, SNR_s, captures the statistical boundary, below which no method can succeed in localization with probability going to one in the minimax sense. The exhaustive search method successfully finds the submatrix above this threshold. In marked contrast to submatrix detection and sparse PCA, the results show an interesting phenomenon that SNR_c is *always* significantly larger than SNR_s under the sub-Gaussian error model, which implies an essential gap between statistical optimality and computational efficiency for submatrix localization.

REFERENCES

[1] AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947

[2] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847

[3] BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning*.

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

[5] BERTHET, Q. and RIGOLLET, P. (2013). Computational lower bounds for sparse PCA. Preprint. Available at arXiv:1304.0828.

[6] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849

MSC2010 subject classifications. Primary 62C20; secondary 90C27.

Key words and phrases. Computational boundary, computational complexity, detection, planted clique, lower bounds, minimax, signal-to-noise ratio, statistical boundary, submatrix localization.

[7] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803
[8] BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567
[9] BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. (2013). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. Preprint. Available at arXiv:1303.5647.
[10] CAI, T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458
[11] CAI, T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281
[12] CAI, T., MA, Z. and WU, Y. (2017). Supplement to "Computational and statistical boundaries for submatrix localization in a large noisy matrix." DOI:10.1214/16-AOS1488SUPP.
[13] CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** Art. 11, 37. MR2811000
[14] CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. USA* **110** E1181–E1190. MR3047651
[15] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474
[16] CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2009). Sparse and low-rank matrix decompositions. In *47th Annual Allerton Conference on Communication, Control, and Computing* 962–967. IEEE, Allerton, IL.
[17] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** Paper No. 27, 57. MR3491121
[18] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* (3) **84** 066106.
[19] DESHPANDE, Y. and MONTANARI, A. (2015). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378
[20] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984
[21] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195
[22] DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921. MR1635414
[23] DRINEAS, P., KANNAN, R. and MAHONEY, M. W. (2006). Fast Monte Carlo algorithms for matrices. II. Computing a low-rank approximation to a matrix. *SIAM J. Comput.* **36** 158–183. MR2231644
[24] FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2013). Statistical algorithms and a lower bound for detecting planted cliques. In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing* 655–664. ACM, New York. MR3210827
[25] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[26] HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** no. 52, 6. MR2994877
[27] JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2015). Phase transitions in semidefinite relaxations. Preprint. Available at arXiv:1511.08769.
[28] JOHNSTONE, I. M. (2013). Gaussian estimation: Sequence and wavelet models. Unpublished manuscript.

[29] KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.

[30] LATAŁA, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** 1273–1282 (electronic). MR2111932

[31] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698

[32] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742

[33] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory* **62** 1458–1484. MR3472260

[34] NG, A. Y., JORDAN, M. I., WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. *Adv. Neural Inf. Process. Syst.* **2** 849–856.

[35] SHABALIN, A. A., WEIGMAN, V. J., PEROU, C. M. and NOBEL, A. B. (2009). Finding large average submatrices in high dimensional data. *Ann. Appl. Stat.* **3** 985–1012. MR2750383

[36] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*, *Vol.* 11. Springer, New York. MR2724359

[37] TSYBAKOV, A. B. (2014). Aggregation and minimax optimality in high-dimensional estimation.

[38] TUFTS, D. W. and SHAH, A. A. (1993). Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix. *IEEE Trans. Signal Process.* **41** 1716–1721.

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

[40] VU, V. (2008). Random discrete matrices. In *Horizons of Combinatorics*. *Bolyai Soc. Math. Stud.* **17** 257–280. Springer, Berlin. MR2432537

[41] VU, V. (2014). A simple SVD algorithm for finding hidden partitions. Preprint. Available at arXiv:1404.3918.

[42] VU, V. Q. and LEI, J. (2012). Minimax rates of estimation for sparse PCA in high dimensions. Preprint. Available at arXiv:1202.0786.

[43] WAINWRIGHT, M. J. (2014). Structured regularizers for high-dimensional problems: Statistical and computational issues. *Annual Review of Statistics and Its Application* **1** 233–253.

[44] WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438

[45] ZASS, R. and SHASHUA, A. (2006). Nonnegative sparse PCA. In *Advances in Neural Information Processing Systems* 1561–1568.

[46] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. Preprint. Available at arXiv:1402.1918.

The Annals of Statistics
2017, Vol. 45, No. 4, 1431–1461
DOI: 10.1214/16-AOS1495

TESTS FOR SEPARABILITY IN NONPARAMETRIC COVARIANCE OPERATORS OF RANDOM SURFACES

BY JOHN A. D. ASTON[1], DAVIDE PIGOLI AND SHAHIN TAVAKOLI

University of Cambridge

The assumption of separability of the covariance operator for a random image or hypersurface can be of substantial use in applications, especially in situations where the accurate estimation of the full covariance structure is unfeasible, either for computational reasons, or due to a small sample size. However, inferential tools to verify this assumption are somewhat lacking in high-dimensional or functional data analysis settings, where this assumption is most relevant. We propose here to test separability by focusing on K-dimensional projections of the difference between the covariance operator and a nonparametric separable approximation. The subspace we project onto is one generated by the eigenfunctions of the covariance operator estimated under the separability hypothesis, negating the need to ever estimate the full nonseparable covariance. We show that the rescaled difference of the sample covariance operator with its separable approximation is asymptotically Gaussian. As a by-product of this result, we derive asymptotically pivotal tests under Gaussian assumptions, and propose bootstrap methods for approximating the distribution of the test statistics. We probe the finite sample performance through simulations studies, and present an application to log-spectrogram images from a phonetic linguistics dataset.

REFERENCES

ASTON, J. A. D. and KIRCH, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688

ASTON, J. A. D., PIGOLI, D. and TAVAKOLI, S. (2017). Supplement to "Tests for separability in nonparametric covariance operators of random surfaces." DOI:10.1214/16-AOS1495SUPPA, DOI:10.1214/16-AOS1495SUPPB.

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

CHEN, K., DELICADO, P. and MÜLLER, H.-G. (2016). Modeling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear. doi:10.1111/rssb.12160.

CONSTANTINOU, P., KOKOSZKA, P. and REIMHERR, M. (2015). Testing separability of space–time functional processes. Available at arXiv:1509.07017.

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

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

MSC2010 subject classifications. Primary 62G10, 62G20; secondary 62F40.

Key words and phrases. Acoustic phonetic data, bootstrap, dimensional reduction, functional data, partial trace, sparsity.

FUENTES, M. (2006). Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference* **136** 447–466. MR2211349

GARCIA, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Statist. Data Anal.* **54** 1167–1178. MR2580947

GENTON, M. G. (2007). Separable approximations of space–time covariance matrices. *Environmetrics* **18** 681–695. MR2408938

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GNEITING, T., GENTON, M. G. and GUTTORP, P. (2007). Geostatistical space–time models, stationarity, separability, and full symmetry. *Monogr. Statist. Appl. Probab.* **107** 151.

GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. A. (1990). *Classes of Linear Operators. Vol. I. Operator Theory: Advances and Applications* **49**. Birkhäuser, Basel. MR1130394

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

KADISON, R. V. and RINGROSE, J. R. (1997). *Fundamentals of the Theory of Operator Algebras: Elementary Theory. Vol. I. Graduate Studies in Mathematics* **15**. Amer. Math. Soc., Providence, RI. MR1468229

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545

LIU, C., RAY, S. and HOOKER, G. (2014). Functional principal components analysis of spatially correlated data. Available at arXiv:1411.4681.

LU, N. and ZIMMERMAN, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett.* **73** 449–457. MR2187860

MAS, A. (2006). A sufficient condition for the CLT in the space of nuclear operators—Application to covariance of random functions. *Statist. Probab. Lett.* **76** 1503–1509. MR2245571

MITCHELL, M. W., GENTON, M. G. and GUMPERTZ, M. L. (2005). Testing for separability of space–time covariances. *Environmetrics* **16** 819–831. MR2216653

PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356

RABINER, L. R. and SCHAFER, R. W. (1978). *Digital Processing of Speech Signals* **100**. Prentice-hall, Englewood Cliffs.

RAMSAY, J. O., GRAVES, S. and HOOKER, G. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York. MR1910407

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

RINGROSE, J. R. (1971). *Compact Non-self-Adjoint Operators*. Van Nostrand Reinhold, London. MR3075382

SECCHI, P., VANTINI, S. and VITELLI, V. (2015). Analysis of spatio-temporal mobile phone data: A case study in the metropolitan area of Milan. *Stat. Methods Appl.* **24** 279–300. MR3376864

SIMPSON, S. L. (2010). An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data. *Stat. Methodol.* **7** 511–519. MR2719948

SIMPSON, S. L., EDWARDS, L. J., STYNER, M. A. and MULLER, K. E. (2014). Separability tests for high-dimensional, low-sample size multivariate repeated measures data. *J. Appl. Stat.* **41** 2450–2461. MR3256397

TANG, R. and MÜLLER, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217

TAVAKOLI, S. (2016). covsep: Tests for determining if the covariance structure of 2-dimensional data is separable. R package version 1.0.0. Available at https://CRAN.R-project.org/package=covsep.

TIBSHIRANI, R. J. (2014). Past, present, and future of statistical science. In *Praise of Sparsity and Convexity* (X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott and J.-L. Wang, eds.) 497–506. Chapman & Hall, London.

WANG, J.-L., CHIOU, J.-M. and MUELLER, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3** 257–295.

WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C., FRISTON, K. J. and EVANS, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* **4** 58–73.

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

The Annals of Statistics
2017, Vol. 45, No. 4, 1462–1487
DOI: 10.1214/16-AOS1496

IDENTIFICATION OF UNIVERSALLY OPTIMAL CIRCULAR DESIGNS FOR THE INTERFERENCE MODEL

BY WEI ZHENG*,[1] MINGYAO AI†,[2] AND KANG LI†

*Indiana University-Purdue University Indianapolis** *and Peking University*†

Many applications of block designs exhibit neighbor and edge effects. A popular remedy is to use the circular design coupled with the interference model. The search for optimal or efficient designs has been intensively studied in recent years. The circular neighbor balanced designs at distances 1 and 2 (CNBD2), including orthogonal array of type I (OA_I) of strength 2, are the two major designs proposed in literature for the purpose of estimating the direct treatment effects. They are shown to be optimal within some reasonable subclasses of designs. By using benchmark designs in approximate design theory, we show that CNBD2 is highly efficient among all possible designs when the error terms are homoscedastic and uncorrelated. However, when the error terms are correlated, these designs will be outperformed significantly by other designs. Note that CNBD2 fall into the special catalog of pseudo symmetric designs, and they only exist when the number of treatments is larger than the block size and the number of blocks is multiple of some constants. In this paper, we elaborate equivalent conditions for any design, pseudo symmetric or not, to be universally optimal for any size of experiment and any covariance structure of the error terms. This result is novel for circular designs and sheds light on other similar models in the search for optimal or efficient asymmetric designs.

REFERENCES

AI, M., GE, G. and CHAN, L. (2007). Circular neighbor-balanced designs universally optimal for total effects. *Sci. China Ser. A* **50** 821–828. MR2353065

AI, M., YU, Y. and HE, S. (2009). Optimality of circular neighbor-balanced designs for total effects with autoregressive correlated observations. *J. Statist. Plann. Inference* **139** 2293–2304. MR2507991

ALDRED, R. E. L., BAILEY, R. A., MCKAY, B. D. and WANLESS, I. M. (2014). Circular designs balanced for neighbours at distances one and two. *Biometrika* **101** 943–956. MR3286927

AZAÏS, J. M., BAILEY, R. A. and MONOD, H. (1993). A catalogue of efficient neighbour-designs with border plots. *Biometrics* **49** 1252–1261.

AZAÏS, J. M., ONILLON, J. and LEFORT-BUSON, M. (1986). Une méthode d'étude de phénomènes de compétition entre génotypes. Application au colza (*Brassica napus* L.). *Agronomie* **6** 601–614.

BAILEY, R. A. and DRUILHET, P. (2004). Optimality of neighbor-balanced designs for total effects. *Ann. Statist.* **32** 1650–1661. MR2089136

BAILEY, R. A. and DRUILHET, P. (2014). Optimal cross-over designs for full interaction models. *Ann. Statist.* **42** 2282–2300. MR3269980

MSC2010 subject classifications. Primary 62K05; secondary 62J05.

Key words and phrases. Approximate design theory, circular design, interference model, linear equations system, universal optimality.

BAILEY, R. A. and PAYNE, R. W. (1990). Experimental design: Statistical research and its application. In *Institute of Arable Crops Research Report for* 1989 (J. Abbott, ed.) 107–112. Agriculture and Food Research Council, Institute of Arable Crops Research, Harpenden.

BESAG, J. and KEMPTON, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42** 231–251.

BHALLI, M. A., DAY, A. D., TUCKER, H., THOMPSON, R. K. and MASSEY, G. D. (1964). End-border effects in irrigated barley yield trials. *Agron. J.* **56** 346–348.

CLARKE, F. R., BAKER, R. J. and DEPAUW, R. M. (2000). Plot direction and spacing effects on interplot interference in spring wheat cultivar trials. *Crop Sci.* **40** 655–658.

CONNOLLY, T., MCNICOL, J. W., WASTIE, R. L. and STEWART, H. E. (1995). Evaluating between-plant interference in field trials for assessing potato genotypes for resistance to late blight. *Ann. Appl. Biol.* **127** 273–282.

DAVID, O., MONOD, H., LORGEOU, J. and PHILIPPEAU, G. (2001). Control of interplot interference in Grain Maize: A multi-site comparison. *Crop Sci.* **41** 406–414.

DRAPER, N. R. and GUTTMAN, I. (1980). Incorporating overlap effects from neighbouring units into response surface models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **29** 128–134.

DRUILHET, P. (1999). Optimality of neighbour balanced designs. *J. Statist. Plann. Inference* **81** 141–152. MR1718460

DRUILHET, P. and TINSSON, W. (2012). Efficient circular neighbour designs for spatial interference model. *J. Statist. Plann. Inference* **142** 1161–1169. MR2879760

DYKE, G. V. and SHELLEY, C. F. (1976). Serial designs balanced for effects of neighbours on both sides. *J. Agric. Sci.* **87** 303–305.

FILIPIAK, K. (2012). Universally optimal designs under an interference model with equal left- and right-neighbor effects. *Statist. Probab. Lett.* **82** 592–598. MR2887476

FILIPIAK, K. and MARKIEWICZ, A. (2003). Optimality of neighbor balanced designs under mixed effects model. *Statist. Probab. Lett.* **61** 225–234. MR1959129

FILIPIAK, K. and MARKIEWICZ, A. (2004). Optimality of type I orthogonal arrays for general interference model with correlated observations. *Statist. Probab. Lett.* **68** 259–265. MR2083894

FILIPIAK, K. and MARKIEWICZ, A. (2005). Optimality and efficiency of circular neighbor balanced designs for correlated observations. *Metrika* **61** 17–27. MR2126434

FILIPIAK, K. and MARKIEWICZ, A. (2007). Optimal designs for a mixed interference model. *Metrika* **65** 369–386. MR2299558

FILIPIAK, K. and MARKIEWICZ, A. (2012). On universal optimality of circular weakly neighbor balanced designs under an interference model. *Comm. Statist. Theory Methods* **41** 2356–2366.

FILIPIAK, K. and MARKIEWICZ, A. (2014). On the optimality of circular block designs under a mixed interference model. *Comm. Statist. Theory Methods* **43** 4534–4545. MR3275481

FILIPIAK, K. and RÓŻAŃSKI, R. (2013). On the E-optimality of complete block designs under a mixed interference model. *J. Statist. Plann. Inference* **143** 583–592. MR2995117

FILIPIAK, K., RÓŻAŃSKI, R., SAWIKOWSKA, A. and WOJTERA-TYRAKOWSKA, D. (2008). On the E-optimality of complete designs under an interference model. *Statist. Probab. Lett.* **78** 2470–2477. MR2462681

FINNEY, D. J. and OUTHWAITE, A. D. (1956). Serially balanced sequences in bioassay. *Proc. R. Soc. Lond., B, Biol. Sci.* **145** 493–507.

GILL, P. S. (1993). Design and analysis of field experiments incorporating local and remote effects of treatments. *Biom. J.* **35** 343–354. MR1222502

GOLDRINGER, I., BRABANT, P. and KEMPTON, R. A. (1994). Adjustment for competition between genotypes in single-row-plot trials of winter wheat (*Triticum aestivum*). *Plant Breed.* **112** 294–300.

HIDE, G. A. and READ, P. J. (1990). Effect of neighbouring plants on the yield of potatoes from seed tubers affected with gangrene (*Phoma foveata*) or from plants affected with stem canker (*Rhizoctonia solani*). *Ann. Appl. Biol.* **116** 233–243.

JEEVITHA, M. and SANTHARAM, C. (2013). Optimality and efficiency of circular neighbor balanced design. *International Journal of Statistika and Mathematika* **7** 19–23.

KEMPTON, R. A. (1982). Adjustment for competition between varieties in plant breeding trials. *J. Agric. Sci.* **98** 599–611.

KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models* (*Proc. Internat. Sympos.*, *Colorado State Univ.*, *Ft. Collins*, *Colo.*, 1973) (J. N. Srivastava, ed.) 333–353. North-Holland, Amsterdam. MR0395079

KUNERT, J. and MARTIN, R. J. (2000). On the determination of optimal designs for an interference model. *Ann. Statist.* **28** 1728–1742. MR1835039

KUNERT, J. and MERSMANN, S. (2011). Optimal designs for an interference model. *J. Statist. Plann. Inference* **141** 1623–1632. MR2747931

KUSHNER, H. B. (1997). Optimal repeated measurements designs: The linear optimality equations. *Ann. Statist.* **25** 2328–2344. MR1604457

LANGTON, S. (1990). Avoiding edge effects in agroforestry experiments; the use of neighbour-balanced designs and guard areas. *Agrofor. Syst.* **12** 173–185.

LI, K., ZHENG, W. and AI, M. (2015). Optimal designs for the proportional interference model. *Ann. Statist.* **43** 1596–1616. MR3357872

MACDONALD, G. E. and PECK, N. H. (1976). Border effects in a long-term fertility experiment. *Agron. J.* **68** 530–532.

MAGDA, C. G. (1980). Circular balanced repeated measurements designs. *Comm. Statist. Theory Methods* **9** 1901–1918. MR0603164

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York. MR1211416

REES, D. H. (1967). Some designs of use in serology. *Biometrics* **23** 779–791.

SPECKEL, D., VINCOURT, P., AZAÏS, J. M. and KOBILINSKY, A. (1987). Étude de la compétition interparcellaire chez le tournesol. *Biom. Praxim.* **27** 21–43.

ZHENG, W. (2015). Universally optimal designs for two interference models. *Ann. Statist.* **43** 501–518. MR3316188

The Annals of Statistics
2017, Vol. 45, No. 4, 1488–1515
DOI: 10.1214/16-AOS1497

CO-CLUSTERING OF NONSMOOTH GRAPHONS

BY DAVID CHOI

Carnegie Mellon University

Performance bounds are given for exploratory co-clustering/blockmodeling of bipartite graph data, where we assume the rows and columns of the data matrix are samples from an arbitrary population. This is equivalent to assuming that the data is generated from a nonsmooth graphon. It is shown that co-clusters found by any method can be extended to the row and column populations, or equivalently that the estimated blockmodel approximates a blocked version of the generative graphon, with estimation error bounded by $O_P(n^{-1/2})$. Analogous performance bounds are also given for degree-corrected blockmodels and random dot product graphs, with error rates depending on the dimensionality of the latent variable space.

REFERENCES

AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* 692–700.

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2009). Mixed membership stochastic blockmodels. In *Advances in Neural Information Processing Systems* 33–40.

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis*: *A Hitchhiker's Guide*, 3rd ed. Springer, Berlin. MR2378491

BIAU, G., DEVROYE, L. and LUGOSI, G. (2008). On the performance of clustering in Hilbert spaces. *IEEE Trans*. *Inform*. *Theory* **54** 781–790. MR2444554

BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. and LEFEBVRE, E. (2008). Fast unfolding of communities in large networks. *J*. *Stat*. *Mech*. *Theory Exp*. **2008** P10008.

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

CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann*. *Statist*. **43** 1027–1059. MR3346696

CHEN, A., AMINI, A. A., LEVINA, E. and BICKEL, P. J. (2012). Fitting community models to large sparse networks. *Ann*. *Statist*. **41** 2097–2122.

CHEN, H.-C., ZOU, W., TIEN, Y.-J. and CHEN, J. J. (2013). Identification of bicluster regions in a binary matrix and its applications. *PLoS ONE* **8** e71680.

CHOI, D. (2017). Supplement to "Co-clustering of nonsmooth graphons." DOI:10.1214/16-AOS1497SUPP.

CHOI, D. and WOLFE, P. J. (2014). Co-clustering separately exchangeable network data. *Ann*. *Statist*. **42** 29–63. MR3161460

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys*. *Rev*. *E* **84** 066106.

MSC2010 subject classifications. Primary 62G05; secondary 05C80, 60B20.

Key words and phrases. Graphon, bipartite graph, co-clustering, statistical network analysis, stochastic blockmodel, degree-corrected blockmodel, random dot product graph.

DIACONIS, P. and JANSON, S. (2007). Graph limits and exchangeable random graphs. Preprint. Available at arXiv:0712.2749.

GAO, C., LU, Y. and ZHOU, H. H. (2014). Rate-optimal graphon estimation. Preprint. Available at arXiv:1410.5837.

GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2015a). Achieving optimal misclassification proportion in stochastic block model. Preprint. Available at arXiv:1505.03772.

GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2015b). Optimal estimation and completion of matrices with biclustering structures. Preprint. Available at arXiv:1512.00150.

GOH, K.-I., CUSICK, M. E., VALLE, D., CHILDS, B., VIDAL, M. and BARABÁSI, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci. USA* **104** 8685–8690.

HARPAZ, R., PEREZ, H., CHASE, H. S., RABADAN, R., HRIPCSAK, G. and FRIEDMAN, C. (2011). Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clin. Pharmacol. Ther.* **89** 243–250.

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. N. and BARABÁSI, A.-L. (2000). The large-scale organization of metabolic networks. *Nature* **407** 651–654.

JI, P. and JIN, J. (2014). Coauthorship and citation networks for statisticians. Preprint. Available at arXiv:1410.2840.

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107, 10. MR2788206

KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2015). Oracle inequalities for network models and sparse graphon estimation. Preprint. Available at arXiv:1507.04118.

KRZAKALA, F., MOORE, C., MOSSEL, E., NEEMAN, J., SLY, A., ZDEBOROVÁ, L. and ZHANG, P. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** 20935–20940. MR3174850

LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *Ann. Appl. Stat.* **5** 309–336. MR2810399

LOVÁSZ, L. (2012). *Large Networks and Graph Limits* **60**.

MOSSEL, E., NEEMAN, J. and SLY, A. (2013). A proof of the block model threshold conjecture. Preprint. Available at arXiv:1311.4115.

NEWMAN, M. E. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64** 016131.

NEWMAN, M. E. (2013). Spectral community detection in sparse networks. Preprint. Available at arXiv:1308.6494.

OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.

ROHE, K., QIN, T. and YU, B. (2012). Co-clustering for directed graphs: The stochastic co-Blockmodel and spectral algorithm Di-Sim. Preprint. Available at arXiv:1204.2296.

SCHNEIDER, R. (2013). *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge Univ. Press, Cambridge.

SUSSMAN, D. L., TANG, M. and PRIEBE, C. E. (2012). Universally consistent latent position estimation and vertex classification for random dot product graphs. Preprint. Available at arXiv:1207.6745.

SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899

TRAUD, A. L., KELSIC, E. D., MUCHA, P. J. and PORTER, M. A. (2011). Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **53** 526–543. MR2834086

VAN UITERT, M., MEULEMAN, W. and WESSELS, L. (2008). Biclustering sparse binary genomic data. *J. Comput. Biol.* **15** 1329–1345. MR2461979

WASSERMAN, L. (2006). *All of Nonparametric Statistics*. Springer, New York. MR2172729

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083

The Annals of Statistics
2017, Vol. 45, No. 4, 1516–1541
DOI: 10.1214/16-AOS1498

MINIMAX THEORY OF ESTIMATION OF LINEAR FUNCTIONALS OF THE DECONVOLUTION DENSITY WITH OR WITHOUT SPARSITY

BY MARIANNA PENSKY[1]

University of Central Florida

The present paper considers the problem of estimating a linear functional $\Phi = \int_{-\infty}^{\infty} \varphi(x) f(x)\,dx$ of an unknown deconvolution density f on the basis of n i.i.d. observations, $Y_1, \dots, Y_n$ of $Y = \theta + \xi$, where ξ has a known pdf g, and f is the pdf of θ. The objective of the present paper is to develop the a general minimax theory of estimating Φ, and to relate this problem to estimation of functionals $\Phi_n = n^{-1}\sum_{i=1}^n \varphi(\theta_i)$ in indirect observations. In particular, we offer a general, Fourier transform based approach to estimation of Φ (and Φ_n) and derive upper and minimax lower bounds for the risk for an arbitrary square integrable function φ. Furthermore, using technique of inversion formulas, we extend the theory to a number of situations when the Fourier transform of φ does not exist, but Φ can be presented as a functional of the Fourier transform of f and its derivatives. The latter enables us to construct minimax estimators of the functionals that have never been handled before such as the odd absolute moments and the generalized moments of the deconvolution density. Finally, we generalize our results to the situation when the vector $\boldsymbol{\theta}$ is sparse and the objective is estimating Φ (or Φ_n) over the nonzero components only. As a direct application of the proposed theory, we automatically recover multiple recent results and obtain a variety of new ones such as, for example, estimation of the mixing probability density function with classical and Berkson errors and estimation of the $(2M + 1)$-th absolute moment of the deconvolution density.

REFERENCES

BERKSON, J. (1950). Are there two regression problems? *J. Amer. Statist. Assoc.* **45** 164–180.

BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. MR2546799

CAI, T. T., JIN, J. and LOW, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653

CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169

CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041. MR2816346

COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2015). Minimax estimation of linear and quadratic functionals on sparsity classes. Available at arXiv:1502.00665.

DATTNER, I., GOLDENSHLUGER, A. and JUDITSKY, A. (2011). On deconvolution of distribution functions. *Ann. Statist.* **39** 2477–2501. MR2906875

MSC2010 subject classifications. Primary 62G20; secondary 62G05, 62G07.
Key words and phrases. Linear functional, minimax lower bound, sparsity, deconvolution.

DELAIGLE, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canad. J. Statist.* **35** 89–104. MR2345376

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38** 481–482. MR0045992

GRADSHTEIN, I. S. and RYZHIK, I. M. (1980). *Tables of Integrals, Series, and Products*. Academic Press, New York.

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167

PENSKY, M. (2017). Supplement to "Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity." DOI:10.1214/16-AOS1498SUPP.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

ZAYED, A. I. (1996). *Handbook of Function and Generalized Function Transformations*. CRC Press, Boca Raton, FL. MR1392476

The Annals of Statistics
2017, Vol. 45, No. 4, 1542–1578
DOI: 10.1214/16-AOS1499

NONPARAMETRIC CHANGE-POINT ANALYSIS OF VOLATILITY

BY MARKUS BIBINGER[1], MORITZ JIRAK[1] AND MATHIAS VETTER

Philipps-Universität Marburg, Technische Universität Braunschweig and Christian-Albrechts-Universität zu Kiel

In this work, we develop change-point methods for statistics of high-frequency data. The main interest is in the volatility of an Itô semimartingale, the latter being discretely observed over a fixed time horizon. We construct a minimax-optimal test to discriminate continuous paths from paths with volatility jumps, and it is shown that the test can be embedded into a more general theory to infer the smoothness of volatilities. In a high-frequency setting, we prove weak convergence of the test statistic under the hypothesis to an extreme value distribution. Moreover, we develop methods to infer changes in the Hurst parameters of fractional volatility processes. A simulation study is conducted to demonstrate the performance of our methods in finite-sample applications.

REFERENCES

AÏT-SAHALIA, Y. and JACOD, J. (2009). Testing for jumps in a discretely observed process. *Ann. Statist.* **37** 184–222. MR2488349

ALVAREZ, A., PANLOUP, F., PONTIER, M. and SAVY, N. (2012). Estimation of the instantaneous volatility. *Stat. Inference Stoch. Process.* **15** 27–59. MR2892587

ANDERSEN, T. G. and BOLLERSLEV, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Internat. Econom. Rev.* **39** 885–905.

ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

ARCONES, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22** 2242–2274. MR1331224

AUE, A., HÖRMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452

BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280. MR1904704

BIBINGER, M., JIRAK, M. and VETTER, M. (2016a). Supplement to "Nonparametric change-point analysis of volatility." DOI:10.1214/16-AOS1499SUPPA.

BIBINGER, M., JIRAK, M. and VETTER, M. (2016b). Supplement to "Nonparametric change-point analysis of volatility." DOI:10.1214/16-AOS1499SUPPB.

BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2013). Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **41** 2479–2512. MR3112923

MSC2010 subject classifications. Primary 62M10; secondary 62G10.

Key words and phrases. High-frequency data, nonparametric change-point test, minimax-optimal test, stochastic volatility, volatility jumps.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958

COEURJOLLY, J.-F. (2000). Simulation and identification of the fractional Brownian motion: A bibliographical and comparative study. *J. Stat. Softw.* **50** 1–53.

COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8** 291–323. MR1645101

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester. MR2743035

EMBRECHTS, P. and MAEJIMA, M. (2000). An introduction to the theory of self-similar stochastic processes. In *Proceedings of the Summer School on Mathematical Physics* 1999: *Methods of Renormalization Group* (*Tokyo*) **14** 1399–1420. Internat. J. Modern Phys. B, 12-13. MR1791158

GATHERAL, J., JAISSON, T. and ROSENBAUM, M. (2014). Volatility is rough. Preprint. Available at arXiv:1410.3394.

HALL, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields* **89** 447–455. MR1118558

HINKLEY, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika* **58** 509–523. MR0312623

HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872

IACUS, S. M. and YOSHIDA, N. (2012). Estimation for the change point of volatility in a stochastic differential equation. *Stochastic Process. Appl.* **122** 1068–1092. MR2891447

INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446

JACOD, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités*, *XXXI*. *Lecture Notes in Math.* **1655** 232–246. Springer, Berlin. MR1478732

JACOD, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.* **118** 517–559. MR2394762

JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. MR3113818

JACOD, J. and TODOROV, V. (2010). Do price and volatility jump together? *Ann. Appl. Probab.* **20** 1425–1469. MR2676944

MANCINI, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.* **36** 270–296. MR2528985

MARSAGLIA, G., TSANG, WAN, W. and WANG, J. (2003). Evaluating Kolmogorov's distribution. *J. Stat. Softw.* **8** 1–4.

MÓRICZ, F. A., SERFLING, R. J. and STOUT, W. F. (1982). Moment and probability bounds with quasisuperadditive structure for the maximum partial sum. *Ann. Probab.* **10** 1032–1040. MR0672303

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761. MR1165590

MÜLLER, H.-G. and STADTMÜLLER, U. (1999). Discontinuous versus smooth regression. *Ann. Statist.* **27** 299–337. MR1701113

MYKLAND, P. A. (2012). A Gaussian calculus for inference from high frequency data. *Ann. Finance* **8** 235–258. MR2922797

MYKLAND, P. A. and ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77** 1403–1445. MR2561071

PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412

PETTITT, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika* **67** 79–84. MR0570508

PHILLIPS, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301. MR0882096

SPOKOINY, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.* **26** 1356–1378. MR1647669

SPOKOINY, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Ann. Statist.* **37** 1405–1436. MR2509078

TODOROV, V. and TAUCHEN, G. (2011). Volatility jumps. *J. Bus. Econom. Statist.* **29** 356–371. MR2848503

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759

The Annals of Statistics
2017, Vol. 45, No. 4, 1579–1608
DOI: 10.1214/16-AOS1500

A NEW APPROACH TO OPTIMAL DESIGNS FOR CORRELATED OBSERVATIONS[1]

BY HOLGER DETTE*, MARIA KONSTANTINOU*
AND ANATOLY ZHIGLJAVSKY†

Ruhr-Universität Bochum and Cardiff University*†

This paper presents a new and efficient method for the construction of optimal designs for regression models with dependent error processes. In contrast to most of the work in this field, which starts with a model for a finite number of observations and considers the asymptotic properties of estimators and designs as the sample size converges to infinity, our approach is based on a continuous time model. We use results from stochastic analysis to identify the best linear unbiased estimator (BLUE) in this model. Based on the BLUE, we construct an efficient linear estimator and corresponding optimal designs in the model for finite sample size by minimizing the mean squared error between the optimal solution in the continuous time model and its discrete approximation with respect to the weights (of the linear estimator) and the optimal design points, in particular in the multiparameter case.

In contrast to previous work on the subject, the resulting estimators and corresponding optimal designs are very efficient and easy to implement. This means that they are practically not distinguishable from the weighted least squares estimator and the corresponding optimal designs, which have to be found numerically by nonconvex discrete optimization. The advantages of the new approach are illustrated in several numerical examples.

REFERENCES

AKHIEZER, N. I. and GLAZMAN, I. M. (1993). *Theory of Linear Operators in Hilbert Space*. Dover Publications, New York. MR1255973

BICKEL, P. J. and HERZBERG, A. M. (1979). Robustness of design against autocorrelation in time. I. Asymptotic theory, optimality for location and linear regression. *Ann. Statist.* **7** 77–95. MR0515685

BOLTZE, L. and NÄTHER, W. (1982). On effective observation methods in regression models with correlated errors. *Math. Operationsforsch. Statist. Ser. Statist.* **13** 507–519. MR0682028

CLERC, M. (2006). *Particle Swarm Optimization*. ISTE, London. MR2269598

DETTE, H., KUNERT, J. and PEPELYSHEV, A. (2008). Exact optimal designs for weighted least squares analysis with correlated errors. *Statist. Sinica* **18** 135–154. MR2384982

DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2013). Optimal design for linear models with correlated observations. *Ann. Statist.* **41** 143–176. MR3059413

DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2016). Optimal designs in regression with correlated errors. *Ann. Statist.* **44** 113–152. MR3449764

MSC2010 subject classifications. Primary 62K05; secondary 62M05.

Key words and phrases. Linear regression, correlated observations, optimal design, Gaussian white mouse model, Doob representation, quadrature formulas.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Stat.* **20** 393–403. MR0030732

HARMAN, R. and ŠTULAJTER, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika* **72** 281–294. MR2725102

HARMAN, R. and ŠTULAJTER, F. (2011). Optimality of equidistant sampling designs for the Brownian motion with a quadratic drift. *J. Statist. Plann. Inference* **141** 2750–2758. MR2787742

IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation. Applications of Mathematics* **16**. Springer, New York. MR0620321

KISEL'ÁK, J. and STEHLÍK, M. (2008). Equidistant and D-optimal designs for parameters of Ornstein–Uhlenbeck process. *Statist. Probab. Lett.* **78** 1388–1396. MR2453793

MEHR, C. B. and MCFADDEN, J. A. (1965). Certain properties of Gaussian processes and their first-passage times. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **27** 505–522. MR0199885

MÜLLER, W. G. and PÁZMAN, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika* **90** 423–434. MR1986657

NÄTHER, W. (1985a). *Effective Observation of Random Fields*. BSB B. G. Teubner Verlagsgesellschaft, Leipzig. MR0863287

NÄTHER, W. (1985b). Exact design for regression models with correlated errors. *Statistics* **16** 479–484. MR0803486

PÁZMAN, A. and MÜLLER, W. G. (2001). Optimal design of experiments subject to correlated errors. *Statist. Probab. Lett.* **52** 29–34. MR1820047

PUKELSHEIM, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia, PA. MR2224698

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Stat.* **37** 66–89. MR0192601

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Stat.* **39** 49–69. MR0220424

WONG, W. K., CHEN, R.-B., HUANG, C.-C. and WANG, W. (2015). A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLoS ONE* **10** e0124720.

ZHIGLJAVSKY, A., DETTE, H. R. and PEPELYSHEV, A. (2010). A new approach to optimal design for linear models with correlated observations. *J. Amer. Statist. Assoc.* **105** 1093–1103. MR2752605

The Annals of Statistics
2017, Vol. 45, No. 4, 1609–1637
DOI: 10.1214/16-AOS1502

RARE-EVENT ANALYSIS FOR EXTREMAL EIGENVALUES OF WHITE WISHART MATRICES

BY TIEFENG JIANG*,[1], KEVIN LEDER*,[2] AND GONGJUN XU†,[3]

University of Minnesota and University of Michigan†*

In this paper, we consider the extreme behavior of the extremal eigenvalues of white Wishart matrices, which plays an important role in multivariate analysis. In particular, we focus on the case when the dimension of the feature p is much larger than or comparable to the number of observations n, a common situation in modern data analysis. We provide asymptotic approximations and bounds for the tail probabilities of the extremal eigenvalues. Moreover, we construct efficient Monte Carlo simulation algorithms to compute the tail probabilities. Simulation results show that our method has the best performance among known approximation approaches, and furthermore provides an efficient and accurate way for evaluating the tail probabilities in practice.

REFERENCES

ADLER, R. J., BLANCHET, J. H. and LIU, J. (2012). Efficient Monte Carlo for high excursions of Gaussian random fields. *Ann. Appl. Probab.* **22** 1167–1214. MR2977989

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

ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York. MR2331321

ASMUSSEN, S. and KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.* **38** 545–558. MR2264957

BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory* **57** 2400–2419. MR2809098

BLANCHET, J. and GLYNN, P. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Probab.* **18** 1351–1378. MR2434174

BLANCHET, J., GLYNN, P. and LEDER, K. (2012). On Lyapunov inequalities and subsolutions for efficient importance sampling. *ACM Trans. Model. Comput. Simul.* **22** Art. 13, 27. MR2980826

BLANCHET, J. H. and LIU, J. (2008). State-dependent importance sampling for regularly varying random walks. *Adv. in Appl. Probab.* **40** 1104–1128. MR2488534

BORDENAVE, C. and CAPUTO, P. (2014). A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.* **42** 2454–2496. MR3265172

DUMITRIU, I. (2003). Eigenvalue Statistics for Beta-Ensembles. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

DUMITRIU, I. and EDELMAN, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. MR1936554

MSC2010 subject classifications. 65C05, 60B20.

Key words and phrases. Importance sampling, extremal eigenvalues, random matrix, β-Laguerre ensemble.

DUPUIS, P., LEDER, K. and WANG, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans*. *Model*. *Comput*. *Simul*. **17** 1–14.

EDELMAN, A. and RAO, N. R. (2005). Random matrix theory. *Acta Numer.* **14** 233–297. MR2168344

EL KAROUI, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. Preprint. Available at arXiv:Math/0309355.

HIAI, F. and PETZ, D. (1998). Eigenvalue density of the Wishart matrix and large deviations. *Infin*. *Dimens*. *Anal*. *Quantum Probab*. *Relat*. *Top*. **1** 633–646. MR1665279

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann*. *Math*. *Stat*. **35** 475–501. MR0181057

JIANG, T., LEDER, K. and XU, G. (2017). Supplement to "Rare-event analysis for extremal eigenvalues of white Wishart matrices." DOI:10.1214/16-AOS1502SUPP.

JIANG, T. and LI, D. (2014). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J*. *Theoret*. *Probab*. To appear.

JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm*. *Math*. *Phys*. **209** 437–476. MR1737991

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961

JOHNSTONE, I. M. and MA, Z. (2012). Fast approach to the Tracy–Widom law at the edge of GOE and GUE. *Ann*. *Appl*. *Probab*. **22** 1962–1988. MR3025686

JOHNSTONE, I. M., MA, Z., PERRY, P. O. and SHAHRAM, M. (2010). RMTstat: Distributions, statistics and tests derived from random matrix theory. Unpublished manuscript.

KWAPIEŃ, J., DROŻDŻ, S. and SPETH, J. (2003). Alternation of different fluctuation regimes in the stock market dynamics. *Phys*. *A* **330** 605–621. MR2044843

KWISSA, M., NAKAYA, H. I., ONLAMOON, N., WRAMMERT, J., VILLINGER, F., PERNG, G. C., YOKSAN, S., PATTANAPANYASAT, K., CHOKEPHAIBULKIT, K., AHMED, R. and PULENDRAN, B. (2014). Dengue virus infection induces expansion of a CD14+ CD16+ monocyte population that stimulates plasmablast differentiation. *Cell Host & Microbe* **16** 115–127.

LIU, J. and XU, G. (2014a). On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *Ann*. *Appl*. *Probab*. **24** 1691–1738. MR3211008

LIU, J. and XU, G. (2014b). Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *ACM Trans*. *Model*. *Comput*. *Simul*. **24** Art. 9, 24. MR3195228

MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359. MR2888709

MACDONALD, I. G. (1995). *Symmetric Functions and Hall Polynomials*, 2nd ed. Oxford Univ. Press, New York. MR1354144

MAÏDA, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron*. *J*. *Probab*. **12** 1131–1150 (electronic). MR2336602

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math*. **1** 457–483.

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

PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet*. **2** e190.

RAMÍREZ, J. A., RIDER, B. and VIRÁG, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc*. **24** 919–944. MR2813333

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann*. *Math*. *Stat*. **24** 220–238. MR0057519

SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann*. *Statist*. **4** 673–684. MR0418369

XU, G., LIN, G. and LIU, J. (2014). Rare-event simulation for the stochastic Korteweg–de Vries equation. *SIAM/ASA J. Uncertain*. *Quantificat*. **2** 698–716. MR3280104

The Annals of Statistics
2017, Vol. 45, No. 4, 1638–1663
DOI: 10.1214/16-AOS1503

ROBUST DISCRIMINATION DESIGNS OVER HELLINGER NEIGHBOURHOODS[1]

BY RUI HU AND DOUGLAS P. WIENS

MacEwan University and University of Alberta

To aid in the discrimination between two, possibly nonlinear, regression models, we study the construction of experimental designs. Considering that each of these two models might be only approximately specified, robust "maximin" designs are proposed. The rough idea is as follows. We impose neighbourhood structures on each regression response, to describe the uncertainty in the specifications of the true underlying models. We determine the least favourable—in terms of Kullback–Leibler divergence—members of these neighbourhoods. Optimal designs are those maximizing this minimum divergence. Sequential, adaptive approaches to this maximization are studied. Asymptotic optimality is established.

REFERENCES

ATKINSON, A. C. and FEDOROV, V. V. (1975a). The design of experiments for discriminating between two rival models. *Biometrika* **62** 57–70. MR0370955

ATKINSON, A. C. and FEDOROV, V. V. (1975b). Optimal design: Experiments for discriminating between several models. *Biometrika* **62** 289–303. MR0381163

BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654. MR0108872

CHAUDHURI, P. and MYKLAND, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood. *J. Amer. Statist. Assoc.* **88** 538–546. MR1224379

DETTE, H. (1994). Discrimination designs for polynomial regression on compact intervals. *Ann. Statist.* **22** 890–903. MR1292546

DETTE, H. and HALLER, G. (1998). Optimal designs for the identification of the order of a Fourier regression. *Ann. Statist.* **26** 1496–1521. MR1647689

DETTE, H. and TITOFF, S. (2009). Optimal discrimination designs. *Ann. Statist.* **37** 2056–2082. MR2533479

FEDOROV, V. V. (1975). Optimal experimental designs for discriminating two rival regression models. In *A Survey of Statistical Design and Linear Models* (*Proc. Internat. Sympos.*, *Colorado State Univ.*, *Ft. Collins*, *Colo.*, 1973) 155–164. North-Holland, Amsterdam. MR0375666

FEDOROV, V. V. and PAZMAN, A. (1968). Design of physical experiments. *Fortschr. Phys.* **16** 325–355.

FORD, I., TITTERINGTON, D. M. and KITSOS, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49–60. MR0997670

HILL, P. D. H. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics* **20** 15–21.

MSC2010 subject classifications. Primary 62K99, 62H30; secondary 62F35.

Key words and phrases. Adaptive design, Hellinger distance, Kullback–Leibler divergence, maximin, Michaelis–Menten model, Neyman–Pearson test, nonlinear regression, optimal design, robustness, sequential design.

HU, R. (2016). Robust designs for model discrimination and prediction of a threshold probability. Ph.D. dissertation, Univ. Alberta.

HU, R. and WIENS, D. P. (2017). Supplement to "Robust discrimination designs over Hellinger neighbourhoods." DOI:10.1214/16-AOS1503SUPP.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York. MR0606374

HUNTER, W. G. and REINER, A. M. (1965). Designs for discriminating between two rival models. *Technometrics* **7** 307–323. MR0192615

LÓPEZ-FIDALGO, J., TOMMASI, C. and TRANDAFIR, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 231–242. MR2325274

OOSTERHOFF, J. and VAN ZWET, W. R. (2012). *A Note on Contiguity and Hellinger Distance*. Springer, New York.

SINHA, S. and WIENS, D. P. (2003). Asymptotics for robust sequential designs in misspecified regression models. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **42** 233–247. IMS, Beachwood, OH. MR2138295

WIENS, D. P. (2009a). Robust discrimination designs. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 805–829. MR2750096

WIENS, D. P. (2009b). Asymptotic properties of an Neyman–Pearson test for model discrimination, with an application to experimental design. *J. Stat. Theory Pract.* **3** 419–427. MR2751609

WIENS, D. P. and LI, P. (2014). V-optimal designs for heteroscedastic regression. *J. Statist. Plann. Inference* **145** 125–138. MR3125354

WYNN, H. P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Statist.* **41** 1655–1664. MR0267704

The Annals of Statistics
2017, Vol. 45, No. 4, 1664–1693
DOI: 10.1214/16-AOS1504

NONPARAMETRIC BAYESIAN POSTERIOR CONTRACTION RATES FOR DISCRETELY OBSERVED SCALAR DIFFUSIONS[1]

BY RICHARD NICKL AND JAKOB SÖHL

University of Cambridge

We consider nonparametric Bayesian inference in a reflected diffusion model $dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t$, with discretely sampled observations $X_0, X_\Delta, \dots, X_{n\Delta}$. We analyse the nonlinear inverse problem corresponding to the "low frequency sampling" regime where $\Delta > 0$ is fixed and $n \to \infty$. A general theorem is proved that gives conditions for prior distributions Π on the diffusion coefficient σ and the drift function b that ensure minimax optimal contraction rates of the posterior distribution over Hölder–Sobolev smoothness classes. These conditions are verified for natural examples of nonparametric random wavelet series priors. For the proofs, we derive new concentration inequalities for empirical processes arising from discretely observed diffusions that are of independent interest.

REFERENCES

[1] ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985

[2] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften* **348**. Springer, Cham. MR3155209

[3] BARAUD, Y. (2010). A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli* **16** 1064–1085. MR2759169

[4] BASS, R. F. (1998). *Diffusions and Elliptic Operators*. Springer, New York. MR1483890

[5] BASS, R. F. (2011). *Stochastic Processes*. *Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. MR2856623

[6] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[7] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856

[8] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473

[9] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

[10] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274

[11] GINÉ, E. and NICKL, R. (2011). Rates on contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395

MSC2010 subject classifications. Primary 62G05; secondary 60J60, 62F15, 62G20.
Key words and phrases. Nonlinear inverse problem, Bayesian inference, diffusion model.

[12] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, Cambridge.
[13] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509
[14] GOLIGHTLY, A. and WILKINSON, D. J. (2005). Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61** 781–788. MR2196166
[15] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. MR3212631
[16] KOSKELA, J., SPANO, D. and JENKINS, P. A. (2015). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. Preprint. Available at arXiv:1506.04709.
[17] MEYER, Y. (1992). *Wavelets and Operators*. *Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge. MR1228209
[18] NICKL, R. and SÖHL, J. (2017). Supplement to "Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions." DOI:10.1214/16-AOS1504SUPP.
[19] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G. O. and STUART, A. M. (2012). Nonparametric estimation of diffusions: A differential equations approach. *Biometrika* **99** 511–531. MR2966767
[20] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 1–32. MR3383563
[21] 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
[22] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. MR3117105
[23] ROBERTS, G. O. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88** 603–621. MR1859397
[24] 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.
[25] STROOCK, D. W. and VARADHAN, S. R. S. (1971). Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.* **24** 147–225. MR0277037
[26] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785
[27] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861
[28] TRIEBEL, H. (2010). *Theory of Function Spaces*. Birkhäuser/Springer Basel AG, Basel. MR3024598
[29] 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
[30] VAN ZANTEN, H. (2013). Nonparametric Bayesian methods for one-dimensional diffusion models. *Math. Biosci.* **243** 215–222. MR3065207
[31] VAN DER MEULEN, F., SCHAUER, M. and VAN ZANTEN, H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* **71** 615–632. MR3131993
[32] VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485
[33] 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

[34] 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

The Annals of Statistics
2017, Vol. 45, No. 4, 1694–1727
DOI: 10.1214/16-AOS1506

ASYMPTOTIC AND FINITE-SAMPLE PROPERTIES OF ESTIMATORS BASED ON STOCHASTIC GRADIENTS[1]

BY PANOS TOULIS AND EDOARDO M. AIROLDI

University of Chicago and Harvard University

Stochastic gradient descent procedures have gained popularity for parameter estimation from large data sets. However, their statistical properties are not well understood, in theory. And in practice, avoiding numerical instability requires careful tuning of key parameters. Here, we introduce implicit stochastic gradient descent procedures, which involve parameter updates that are implicitly defined. Intuitively, implicit updates shrink standard stochastic gradient descent updates. The amount of shrinkage depends on the observed Fisher information matrix, which does not need to be explicitly computed; thus, implicit procedures increase stability without increasing the computational burden. Our theoretical analysis provides the first full characterization of the asymptotic behavior of both standard and implicit stochastic gradient descent-based estimators, including finite-sample error bounds. Importantly, analytical expressions for the variances of these stochastic gradient-based estimators reveal their exact loss of efficiency. We also develop new algorithms to compute implicit stochastic gradient descent-based estimators for generalized linear models, Cox proportional hazards, M-estimators, in practice, and perform extensive experiments. Our results suggest that implicit stochastic gradient descent procedures are poised to become a workhorse for approximate inference from large data sets.

REFERENCES

AMARI, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Comput*. **10** 251–276.

AMARI, S.-I., PARK, H. and FUKUMIZU, K. (2000). Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural Comput*. **12** 1399–1409.

BACH, F. and MOULINES, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $o(1/n)$. In *Advances in Neural Information Processing Systems* 773–781.

BATHER, J. A. (1989). Stochastic approximation: A generalisation of the Robbins–Monro procedure. In *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics* (*Prague*, 1988) 13–27. Charles Univ., Prague. MR1051424

BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci*. **2** 183–202. MR2486527

BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, Berlin. MR1082341

BERTSEKAS, D. P. (2011). Incremental proximal methods for large scale convex optimization. *Math. Program*. **129** 163–195. MR2837879

MSC2010 subject classifications. 62L20, 62F10, 62L12, 62F12, 62F35.

Key words and phrases. Stochastic approximation, implicit updates, asymptotic variance, generalized linear models, Cox proportional hazards, M-estimation, maximum likelihood, exponential family, statistical efficiency, numerical stability.

BORDES, A., BOTTOU, L. and GALLINARI, P. (2009). SGD-QN: Careful quasi-Newton stochastic gradient descent. *J. Mach. Learn. Res.* **10** 1737–1754. MR2534877

BORKAR, V. S. (2008). *Stochastic Approximation*. Cambridge Univ. Press, Cambridge. MR2442439

BOTTOU, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'*2010 177–186. Springer, Heidelberg. MR3362066

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

CAPPÉ, O. and MOULINES, E. (2009). On-line expectation-maximization algorithm for latent data models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 593–613. MR2749909

CHEN, H. F., LEI, G. and GAO, A. J. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231. MR0931029

CHUNG, K. L. (1954). On a stochastic approximation method. *Ann. Math. Stat.* **25** 463–483. MR0064365

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

DAVISON, A. C. (2003). *Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **11**. Cambridge Univ. Press, Cambridge. MR1998913

DEAN, J., CORRADO, G., MONGA, R., CHEN, K., DEVIN, M., MAO, M., SENIOR, A., TUCKER, P., YANG, K., LE, Q. V. et al. (2012). Large scale distributed deep networks. In *Advances in Neural Information Processing Systems* 1223–1231.

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

DOMINICI, F., DANIELS, M., ZEGER, S. L. and SAMET, J. M. (2002). Air pollution and mortality: Estimating regional and national dose-response relationships. *J. Amer. Statist. Assoc.* **97** 100–111. MR1963390

DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043

DUCHI, J., HAZAN, E. and SINGER, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **12** 2121–2159. MR2825422

DUCHI, J. and SINGER, Y. (2009). Efficient online and batch learning using forward backward splitting. *J. Mach. Learn. Res.* **10** 2899–2934. MR2579916

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012

FABIAN, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Stat.* **39** 1327–1332. MR0231429

FABIAN, V. (1978). On asymptotically efficient recursive estimation. *Ann. Statist.* **6** 854–866. MR0478506

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **222** 309–368.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

GEORGE, A. P. and POWELL, W. B. (2006). Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming. *Mach. Learn.* **65** 167–198.

GREEN, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. Roy. Statist. Soc. Ser. B* **46** 149–192. MR0781879

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. MR2722294

HENNIG, P. and KIEFEL, M. (2013). Quasi-Newton methods: A new direction. *J. Mach. Learn. Res.* **14** 843–865. MR3049491

HOFFMAN, J. D. and FRANKEL, S. (2001). *Numerical Methods for Engineers and Scientists*. CRC press, Boca Raton, FL.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35** 73–101. MR0161415

JAIN, P., TEWARI, A. and KAR, P. (2014). On iterative hard thresholding methods for high-dimensional m-estimation. In *Advances in Neural Information Processing Systems* 685–693.

KLEIN, J. P. and MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media New York.

KRAKOWSKI, K. A., MAHONY, R. E., WILLIAMSON, R. C. and WARMUTH, M. K. (2007). A geometric view of non-linear on-line stochastic gradient descent. Available at https://users.soe.ucsc.edu/ manfred/pubs/T3.pdf.

LANGE, K. (2010). *Numerical Analysis for Statisticians*, 2nd ed. Springer, New York. MR2655999

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics* **31**. Springer, New York. MR1639875

LIONS, P.-L. and MERCIER, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16** 964–979. MR0551319

LJUNG, L., PFLUG, G. and WALK, H. (1992). *Stochastic Approximation and Optimization of Random Systems*. *DMV Seminar* **17**. Birkhäuser, Basel. MR1162311

MARSHALL, A. W. and OLKIN, I. (1960). Multivariate Chebyshev inequalities. *Ann. Math. Stat.* **31** 1001–1014. MR0119234

MARTIN, R. D. and MASRELIEZ, C. J. (1975). Robust estimation via stochastic approximation. *IEEE Trans. Inform. Theory* **IT-21** 263–271. MR0395111

MESTRE, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory* **54** 5113–5129. MR2589886

MILLER, A. J. (1992). Algorithm as 274: Least squares routines to supplement those of gentleman. *Applied Statistics* 458–478.

MOULINES, E. and BACH, F. R. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems* 451–459.

NAGUMO, J.-I. and NODA, A. (1967). A learning method for system identification. *IEEE Trans. Automat. Control* **12** 282–287.

NATIONAL RESEARCH COUNCIL (2013). *Frontiers in Massive Data Analysis*. The National Academies Press, Washington, DC.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

NEMIROVSKI, A., JUDITSKY, A., LAN, G. and SHAPIRO, A. (2008). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. MR2486041

NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York. MR0702836

NEVELSON, M. B. and KHASMINSKIĬ, R. Z. (1973). *Stochastic Approximation and Recursive Estimation* **47**. Amer. Math. Society, Washington.

PARIKH, N. and BOYD, S. (2013). Proximal algorithms. *Found. Trends Optim.* **1** 123–231.

POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814

POLJAK, B. T. and TSYPKIN, JA. Z. (1980). Robust identification. *Automatica J. IFAC* **16** 53–63. MR0571554

POLYAK, B. T. and TSYPKIN, YA. Z. (1979). Adaptive estimation algorithms (convergence, optimality, stability). *Avtomat. i Telemekh.* **3** 71–84. MR0544876

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668

ROCKAFELLAR, R. T. (1976). Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14** 877–898. MR0410483
ROSASCO, L., VILLA, S. and VŨ, B. C. (2014). Convergence of stochastic proximal gradient algorithm. Preprint. Available at arXiv:1403.5074.
RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process. Technical report, Dept. Operations Research and Industrial Engineering, Cornell Univ., Ithaca, NY.
SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Stat.* **29** 373–405. MR0098427
SAKRISON, D. J. (1965). Efficient recursive estimation; application to estimating the parameters of a covariance function. *Internat. J. Engrg. Sci.* **3** 461–483. MR0182082
SAMET, J. M., ZEGER, S. L., DOMINICI, F., CURRIERO, F., COURSAC, I., DOCKERY, D. W., SCHWARTZ, J. and ZANOBETTI, A. (2000). The national morbidity, mortality, and air pollution study. Part II: Morbidity and mortality from air pollution in the United States. *Res. Rep. Health Eff. Inst.* **94** 5–79.
SCHMIDT, M., LE ROUX, N. and BACH, F. (2013). Minimizing finite sums with the stochastic average gradient. Technical report, HAL 00860051.
SHAMIR, O. and ZHANG, T. (2012). Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. Preprint. Available at arXiv:1212.1824.
SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2011). Regularization paths for Cox's proportional Hazards model via coordinate descent. *J. Stat. Softw.* **39** 1–13.
SINGER, Y. and DUCHI, J. C. (2009). Efficient learning using forward-backward splitting. In *Advances in Neural Information Processing Systems* 495–503.
SLOCK, D. T. M. (1993). On the convergence behavior of the lms and the normalized lms algorithms. *IEEE Trans. Signal Process.* **41** 2811–2825.
TOULIS, P. and AIROLDI, E. M. (2015a). Scalable estimation strategies based on stochastic approximations: Classical results and new insights. *Stat. Comput.* **25** 781–795. MR3360492
TOULIS, P. and AIROLDI, E. M. (2015b). Implicit stochastic approximation. Preprint. Available at arXiv:1510.00967.
TOULIS, P. and AIROLDI, E. M. (2017). Supplement to "Asymptotic and finite-sample properties of estimators based on stochastic gradients." DOI:10.1214/16-AOS1506SUPP.
TOULIS, P., AIROLDI, E. M. and RENNIE, J. (2014). Statistical analysis of stochastic gradient methods for generalized linear models. *J. Mach. Learn. Res. W&CP* **32 (ICML)** 667–675.
TOULIS, P., TRAN, D. and AIROLDI, E. M. (2016). Towards stability and optimality in stochastic gradient descent. *J. Mach. Learn. Res. W&CP* **51 (AISTATS)**.
TRAN, D., TOULIS, P. and AIROLDI, E. M. (2015). Stochastic gradient descent methods for estimation with large data sets. Preprint. Available at arXiv:1509.06459.
WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the* 28*th International Conference on Machine Learning* (*ICML*-11) 681–688.
WIDROW, B. and HOFF, M. E. (1960). Adaptive switching circuits. *Defense Technical Information Center*.
WOOD, S. N., GOUDE, Y. and SHAW, S. (2015). Generalized additive models for large data sets. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 139–155. MR3293922
XIAO, L. and ZHANG, T. (2014). A proximal stochastic gradient method with progressive variance reduction. *SIAM J. Optim.* **24** 2057–2075. MR3285905
XU, W. (2011). Towards optimal one pass large scale learning with averaged stochastic gradient descent. Preprint. Available at arXiv:1107.2490.
ZHANG, T. (2004). Solving large scale linear prediction problems using gradient descent algorithms. In *Proceedings of the Twenty-First International Conference on Machine Learning* 116. ACM, New York.

The Annals of Statistics
2017, Vol. 45, No. 4, 1728–1758
DOI: 10.1214/16-AOS1507

FUNCTIONAL CENTRAL LIMIT THEOREMS FOR SINGLE-STAGE SAMPLING DESIGNS

BY HÉLÈNE BOISTARD*, HENDRIK P. LOPUHAÄ† AND ANNE RUIZ-GAZEN*

Toulouse School of Economics and Delft University of Technology†*

For a joint model-based and design-based inference, we establish functional central limit theorems for the Horvitz–Thompson empirical process and the Hájek empirical process centered by their finite population mean as well as by their super-population mean in a survey sampling framework. The results apply to single-stage unequal probability sampling designs and essentially only require conditions on higher order correlations. We apply our main results to a Hadamard differentiable statistical functional and illustrate its limit behavior by means of a computer simulation.

REFERENCES

[1] BARRETT, G. F. and DONALD, S. G. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare. *J. Bus. Econom. Statist.* **27** 1–17. MR2484980

[2] BERGER, Y. G. (1998). Rate of convergence for asymptotic variance of the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **74** 149–168. MR1665125

[3] BERGER, Y. G. (1998). Rate of convergence to normal distribution for the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693

[4] BERGER, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan J. Statist.* **27** 407–426. MR2919728

[5] BERGER, Y. G. and SKINNER, C. J. (2003). Variance estimation for a low income proportion. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **52** 457–468. MR2012970

[6] BERGER, Y. G. and SKINNER, C. J. (2005). A jackknife variance estimator for unequal probability sampling. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 79–89. MR2136640

[7] BERTAIL, P., CHAUTRU, E. and CLÉMENÇON, S. (2016). Empirical processes in survey sampling. *Scand. J. Stat.*. DOI:10.1111/sjos.12243.

[8] BHATTACHARYA, D. (2007). Inference on inequality from household survey data. *J. Econometrics* **137** 674–707. MR2354960

[9] BHATTACHARYA, D. and MAZUMDER, B. (2011). A nonparametric analysis of black–white differences in intergenerational income mobility in the United States. *Quant. Econ.* **2** 335–379.

[10] BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482. MR0740906

[11] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[12] BINDER, D. and ROBERTS, G. (2009). Design- and model-based inference for model parameters. In *Handbook of Statistics* 29*B*: *Sample Surveys*: *Design*, *Methods and Applications* (D. Pfefferman and C. R. Rao, eds.) 33–54. Elsevier, Amsterdam.

MSC2010 subject classifications. 62D05.
Key words and phrases. Design and model-based inference, Hájek Process, Horvitz–Thompson process, rejective sampling, Poisson sampling, high entropy designs, poverty rate.

[13] BOISTARD, H., LOPUHAÄ, H. P. and RUIZ-GAZEN, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electron. J. Stat.* **6** 1967–1983. MR3020253
[14] BOISTARD, H., LOPUHAÄ, H. P. and RUIZ-GAZEN, A. (2017). Supplement to "Functional central limit theorems for single-stage samplings designs." DOI:10.1214/16-AOS1507SUPP.
[15] BREIDT, F. J. and OPSOMER, J. D. (2000). Local polynomial regresssion estimators in survey sampling. *Ann. Statist.* **28** 1026–1053. MR1810918
[16] BRESLOW, N. E. and WELLNER, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Stat.* **34** 86–102. MR2325244
[17] CARDOT, H., CHAOUCH, M., GOGA, C. and LABRUÈRE, C. (2010). Properties of design-based functional principal components analysis. *J. Statist. Plann. Inference* **140** 75–91. MR2568123
[18] CHAUVET, G. (2015). Coupling methods for multistage sampling. *Ann. Statist.* **43** 2484–2506. MR3405601
[19] CONTI, P. L. (2014). On the estimation of the distribution function of a finite population under high entropy sampling designs, with applications. *Sankhya B* **76** 234–259. MR3302272
[20] CONTI, P. L., MARELLA, D. and MECATTI, F. (2015). Recovering sampling distributions of statistics of finite populations via resampling: A predictive approach. Preprint.
[21] DAVIDSON, R. (2009). Reliable inference for the Gini index. *J. Econometrics* **150** 30–40. MR2525992
[22] DELL, F. and D'HAULTFŒUILLE, X. (2008). Measuring the evolution of complex indicators: Theory and application to the poverty rate in France. *Ann. Écon. Stat.* **90** 259–290.
[23] DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804
[24] DORFMAN, A. H. (2009). Chapter 36: Inference on distribution functions and quantiles. In *Handbook of Statistics* 29*B*: *Sample Surveys*: *Design*, *Methods and Applications* (D. Pfefferman and C. R. Rao, eds.) 371–395. Elsevier, Amsterdam.
[25] DUPAČOVÁ, J. (1979). A note on rejective sampling. In *Contributions to Statistics* 71–78. Reidel, Dordrecht. MR0561260
[26] ESCOBAR, E. L. and BERGER, Y. G. (2013). A jackknife variance estimator for self-weighted two-stage samples. *Statist. Sinica* **23** 595–613. MR3086648
[27] FRANCISCO, C. A. and FULLER, W. A. (1991). Quantile estimation with a complex survey design. *Ann. Statist.* **19** 454–469. MR1091862
[28] FULLER, W. A. (2009). *Sampling Statistics*. Wiley, New York.
[29] GRAF, E. and TILLÉ, Y. (2014). Variance estimation using linearization for poverty and social exclusion indicators. *Surv. Methodol.* **40** 61–79.
[30] HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Čas. Pěst. Mat.* **84** 387–423. MR0121897
[31] HÁJEK, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.* **35** 1491–1523. MR0178555
[32] ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029
[33] KORN, E. L. and GRAUBARD, B. I. (1998). Variance estimation for superpopulation parameters. *Statist. Sinica* **8** 1131–1151. MR1666241
[34] KREWSKI, D. and RAO, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9** 1010–1019. MR0628756
[35] LIN, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika* **87** 37–47. MR1766826

[36] OGUZ-ALPER, M. and BERGER, Y. G. (2015). Variance estimation of change of poverty based upon the Turkish EU-SILC survey. *J. Off. Stat.* **31** 155–175.

[37] PFEFFERMANN, D. and SVERCHKOV, M. (2009). Inference under informative sampling. In *Handbook of Statistics* 29*B*: *Sample Surveys*: *Inference and Analysis* (D. Pfefferman and C. R. Rao, eds.) 455–487. Elsevier, Amsterdam.

[38] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

[39] PRÁSKOVÁ, Z. and SEN, P. K. (2009). Asymptotic in finite population sampling. In *Handbook of Statistics* 29*B*: *Sample Surveys*: *Inference and Analysis* (D. Pfefferman and C. R. Rao, eds.) 489–522. Elsevier, Amsterdam.

[40] 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

[41] RUBIN-BLEUER, S. and SCHIOPU KRATINA, I. (2005). On the two-phase framework for joint model and design-based inference. *Ann. Statist.* **33** 2789–2810. MR2253102

[42] SAEGUSA, T. and WELLNER, J. A. (2013). Weighted likelihood estimation under two-phase sampling. *Ann. Statist.* **41** 269–295. MR3059418

[43] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London. MR0848134

[44] THOMPSON, M. E. (1997). *Theory of Sample Surveys*. *Monographs on Statistics and Applied Probability* **74**. Chapman & Hall, London. MR1462619

[45] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[46] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

[47] VÍŠEK, J. Á. (1979). Asymptotic distribution of simple estimate for rejective, Sampford and successive sampling. In *Contributions to Statistics* 263–275. Reidel, Dordrecht. MR0561274

[48] WANG, J. C. (2012). Sample distribution function based goodness-of-fit test for complex surveys. *Comput. Statist. Data Anal.* **56** 664–679. MR2853763

The Annals of Statistics
2017, Vol. 45, No. 4, 1759–1788
DOI: 10.1214/16-AOS1508

ASYMPTOTIC NORMALITY OF SCRAMBLED GEOMETRIC NET QUADRATURE

BY KINJAL BASU[1] AND RAJARSHI MUKHERJEE

LinkedIn Corporation and Stanford University

In a very recent work, Basu and Owen [*Found. Comput. Math.* **17** (2017) 467–496] propose the use of scrambled geometric nets in numerical integration when the domain is a product of s arbitrary spaces of dimension d having a certain partitioning constraint. It was shown that for a class of smooth functions, the integral estimate has variance $O(n^{-1-2/d}(\log n)^{s-1})$ for scrambled geometric nets compared to $O(n^{-1})$ for ordinary Monte Carlo. The main idea of this paper is to expand on the work by Loh [*Ann. Statist.* **31** (2003) 1282–1324] to show that the scrambled geometric net estimate has an asymptotic normal distribution for certain smooth functions defined on products of suitable subsets of $\mathbb{R}^d$.

REFERENCES

[1] ARVO, J., FAJARDO, M. H. P., JENSEN, H. W., MITCHELL, D., PHARR, M. and SHIRLEY, P. (2001). State of the art in Monte Carlo ray tracing for realistic image synthesis. In *ACM Siggraph* 2001 ACM, New York.

[2] BARBOUR, A. D. and CHEN, L. H. Y. (2005). *An Introduction to Stein's Method* **4**. World Scientific, Singapore.

[3] BASU, K. and MUKHERJEE, R. (2017). Supplement to "Asymptotic normality of scrambled geometric net quadrature." DOI:10.1214/16-AOS1508SUPP.

[4] BASU, K. (2015). Quasi-Monte Carlo tractability of high dimensional integration over products of simplices. *J. Complexity* **31** 817–834. MR3400995

[5] BASU, K. and OWEN, A. B. (2015). Low discrepancy constructions in the triangle. *SIAM J. Numer. Anal.* **53** 743–761. MR3323544

[6] BASU, K. and OWEN, A. B. (2016). Transformations and Hardy-Krause variation. *SIAM J. Numer. Anal.* **54** 1946–1966.

[7] BASU, K. and OWEN, A. B. (2017). Scrambled geometric net integration over general product spaces. *Found. Comput. Math.* **17** 467–496.

[8] BRANDOLINI, L., COLZANI, L., GIGANTE, G. and TRAVAGLINI, G. (2013). A Koksma–Hlawka inequality for simplices. In *Trends in Harmonic Analysis*. *Springer INdAM Ser.* **3** 33–46. Springer, Milan. MR3026347

[9] DICK, J. and PILLICHSHAMMER, F. (2010). *Digital Sequences, Discrepancy and Quasi-Monte Carlo Integration*. Cambridge Univ. Press, Cambridge.

[10] KELLER, A. (2013). Quasi-Monte Carlo image synthesis in a nutshell. In *Monte Carlo and Quasi-Monte Carlo Methods* 2012 (J. Dick, F. Y. Kuo, G. W. Peters and I. H. Sloan, eds.). *Springer Proceedings in Mathematics & Statistics* **65** 213–249. Springer, Berlin.

MSC2010 subject classifications. Primary 62E20; secondary 62D05, 65D30.

Key words and phrases. Asymptotic normality, numerical integration, quasi-Monte Carlo, scrambled geometric net, Stein's method.

[11] L'ECUYER, P. and LEMIEUX, C. (2002). A survey of randomized quasi-Monte Carlo methods. In *Modeling Uncertainty*: *An Examination of Stochastic Theory*, *Methods*, *and Applications* (M. Dror, P. L'Ecuyer and F. Szidarovszki, eds.) 419–474. Kluwer Academic, Norwell.

[12] LOH, W.-L. (2003). On the asymptotic distribution of scrambled net quadrature. *Ann*. *Statist*. **31** 1282–1324. MR2001651

[13] NIEDERREITER, H. (1987). Point sets and sequences with small discrepancy. *Monatsh*. *Math*. **104** 273–337. MR0918037

[14] OWEN, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.-S. Shiue, eds.) 299–317. Springer-Verlag, New York.

[15] OWEN, A. B. (1997). Monte Carlo variance of scrambled net quadrature. *SIAM J*. *Numer. Anal*. **34** 1884–1910. MR1472202

[16] OWEN, A. B. (1997). Scrambled net variance for integrals of smooth functions. *Ann*. *Statist*. **25** 1541–1562. MR1463564

[17] OWEN, A. B. (1998). Scrambling Sobol' and Niederreiter–Xing points. *J*. *Complexity* **14** 466–489. MR1659008

[18] OWEN, A. B. (2003). Variance with alternative scramblings of digital nets. *ACM Trans*. *Model*. *Comput*. *Simul*. **13** 363–378.

[19] OWEN, A. B. (2008). Local antithetic sampling with scrambled nets. *Ann*. *Statist*. **36** 2319–2343. MR2458189

[20] PILLARDS, T. and COOLS, R. (2004). A theoretical view on transforming low-discrepancy sequences from a cube to a simplex. *Monte Carlo Methods Appl*. **10** 511–529. MR2105078

[21] PILLARDS, T. and COOLS, R. (2005). Transforming low-discrepancy sequences from a cube to a simplex. *J*. *Comput*. *Appl*. *Math*. **174** 29–42. MR2102647

[22] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *II*: *Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

[23] STROMBERG, K. R. (1994). *Probability for Analysts*. Chapman & Hall, New York. MR1271144

The Annals of Statistics
2017, Vol. 45, No. 4, 1789–1809
DOI: 10.1214/16-AOS1509

YULE'S "NONSENSE CORRELATION" SOLVED!

BY PHILIP A. ERNST*, LARRY A. SHEPP†,1 AND ABRAHAM J. WYNER†

Rice University and University of Pennsylvania†*

In this paper, we resolve a longstanding open statistical problem. The problem is to mathematically prove Yule's 1926 empirical finding of "nonsense correlation" [*J. Roy. Statist. Soc.* **89** (1926) 1–63], which we do by analytically determining the second moment of the empirical correlation coefficient

$$\theta := \frac{\int_0^1 W_1(t)W_2(t)\,dt - \int_0^1 W_1(t)\,dt \int_0^1 W_2(t)\,dt}{\sqrt{\int_0^1 W_1^2(t)\,dt - (\int_0^1 W_1(t)\,dt)^2}\sqrt{\int_0^1 W_2^2(t)\,dt - (\int_0^1 W_2(t)\,dt)^2}},$$

of two *independent* Wiener processes, W_1, W_2. Using tools from Fredholm integral equation theory, we successfully calculate the second moment of θ to obtain a value for the standard deviation of θ of nearly 0.5. The "nonsense" correlation, which we call "volatile" correlation, is volatile in the sense that its distribution is heavily dispersed and is frequently large in absolute value. It is induced because each Wiener process is "self-correlated" in time. This is because a Wiener process is an integral of pure noise, and thus its values at different time points are correlated. In addition to providing an explicit formula for the second moment of θ, we offer implicit formulas for higher moments of θ.

REFERENCES

[1] ALDRICH, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statist. Sci.* **10** 364–376.
[2] BOAS, R. P. JR. (1954). *Entire Functions*. Academic Press, New York. MR0068627
[3] ERDÖS, P. and KAC, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302. MR0015705
[4] FELLER, W. (1957). *An Introduction to Probability Theory and Its Applications*. Wiley, New York.
[5] GRANGER, C. W. J. and NEWBOLD, D. (1974). Spurious regression in econometrics. *J. Econometrics* **2** 111–120.
[6] HENDRY, D. F. (1986). Economic modelling with cointegrated variables: An overview. *Oxf. Bull. Econ. Stat.* **48** 201–212.
[7] LOGAN, B. F., MALLOWS, C. L., RICE, S. O. and SHEPP, L. A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **1** 788–809. MR0362449
[8] LOGAN, B. F. and SHEPP, L. A. (1968). Real zeros of random polynomials. *Proc. Lond. Math. Soc.* (3) **18** 29–35. MR0234512
[9] MAGNUS, J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Ann. Écon. Stat.* **4** 95–109. MR0880741

MSC2010 subject classifications. Primary 60J65, 60G15; secondary 60G05.
Key words and phrases. Volatile correlation, Wiener processes, integral equations.

[10] MANN, M. E., ZHANG, Z., HUGHES, M. K., BRADLEY, R. S., MILLER, S. K., RUTHERFORD, S. and NI, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millenia. *Proc. Natl. Acad. Sci. USA* **105** 13252–13257.

[11] MCSHANE, B. B. and WYNER, A. J. (2011). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.* **5** 5–44.

[12] PHILLIPS, P. C. B. (1986). Understanding spurious regressions in econometrics. *J. Econometrics* **33** 311–340. MR0867979

[13] PHILLIPS, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* **66** 1299–1325. MR1654311

[14] ROGERS, L. C. G. and SHEPP, L. A. (2006). The correlation of the maxima of correlated Brownian motions. *J. Appl. Probab.* **43** 880–883. MR2274808

[15] SHEPP, L. A. (1979). The joint density of the maximum and its location for a Wiener process with drift. *J. Appl. Probab.* **16** 423–427. MR0531776

[16] YULE, G. U. (1926). Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series. *J. Roy. Statist. Soc.* **89** 1–63.

The Annals of Statistics
2017, Vol. 45, No. 4, 1810–1833
DOI: 10.1214/16-AOS1514

SHARP DETECTION IN PCA UNDER CORRELATIONS: ALL EIGENVALUES MATTER

BY EDGAR DOBRIBAN[1]

Stanford University

Principal component analysis (PCA) is a widely used method for dimension reduction. In high-dimensional data, the "signal" eigenvalues corresponding to weak principal components (PCs) do not necessarily separate from the bulk of the "noise" eigenvalues. Therefore, popular tests based on the largest eigenvalue have little power to detect weak PCs. In the special case of the spiked model, certain tests asymptotically equivalent to linear spectral statistics (LSS)—averaging effects over *all* eigenvalues—were recently shown to achieve some power.

We consider a "local alternatives" model for the spectrum of covariance matrices that allows a general correlation structure. We develop new tests to detect PCs in this model. While the top eigenvalue contains little information, due to the strong correlations between the eigenvalues we can detect weak PCs by averaging over all eigenvalues using LSS. We show that it is possible to find the optimal LSS, by solving a certain integral equation. To solve this equation, we develop efficient algorithms that build on our recent method for computing the limit empirical spectrum [Dobriban (2015)]. The solvability of this equation also presents a new perspective on phase transitions in spiked models.

REFERENCES

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620

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

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

BAI, Z. and SILVERSTEIN, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201

MSC2010 subject classifications. Primary 62H25; secondary 62H15, 45B05.

Key words and phrases. Principal component analysis, linear spectral statistic, random matrix theory, linear integral equation, optimal testing.

BOGACHEV, V. I. (2007). *Measure Theory*, *Vol*. 1. Springer, Berlin.

COUILLET, R. and DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge. MR2884783

DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. ArXiv preprint arXiv:1411.3875.

DOBRIBAN, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices Theory Appl*. **4** 1550019, 36. MR3418848

DOBRIBAN, E. (2017). Supplement to "Sharp detection in PCA under correlations: All eigenvalues matter." DOI:10.1214/16-AOS1514SUPP.

DOBRIBAN, E. and WAGER, S. (2015). High-dimensional asymptotics of prediction: Ridge regression and classification. ArXiv preprint arXiv:1507.03003.

GROETSCH, C. W. (1977). *Generalized Inverses of Linear Operators*: *Representation and Approximation*. *Monographs and Textbooks in Pure and Applied Mathematics* **37**. Dekker, New York. MR0458859

HACHEM, W., HARDY, A. and NAJIM, J. (2015). A survey on the eigenvalues local behavior of large complex correlated Wishart matrices. In *Modélisation Aléatoire et Statistique—Journées MAS* 2014. *ESAIM Proc*. *Surveys* **51** 150–174. EDP Sci., Les Ulis. MR3440796

HUANG, J. (2014). Mesoscopic perturbations of large random matrices. ArXiv preprint arXiv:1412.4193.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961

JOHNSTONE, I. M. and ONATSKI, A. (2015). Testing in high-dimensional spiked models. ArXiv preprint arXiv:1509.07269.

KRESS, R. (2014). *Linear Integral Equations*, 3rd ed. Springer, New York. MR3184286

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

MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat*. *Sb*. **114** 507–536.

ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J*. *Econometrics* **168** 244–258. MR2923766

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann*. *Statist*. **41** 1204–1231. MR3113808

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann*. *Statist*. **42** 225–254. MR3189485

PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann*. *Appl*. *Probab*. **18** 1232–1270. MR2418244

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

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J*. *Statist*. *Plann*. *Inference* **150** 1–29. MR3206718

SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J*. *Multivariate Anal*. **54** 295–309. MR1345541

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge Univ. Press, New York. MR3468554

ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann*. *Statist*. **43** 546–591. MR3316190