

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Memorial issue for Charles Stein

Articles

- Editorial: Memorial issue for Charles Stein
RICHARD J. SAMWORTH AND MING YUAN 1811
- Charles Stein and invariance: Beginning with the Hunt–Stein theorem
MORRIS L. EATON AND EDWARD I. GEORGE 1815
- On Charles Stein’s contributions to (in)admissibility WILLIAM E. STRAWDERMAN 1823
- Stein 1956: Efficient nonparametric testing and estimation
A. W. VAN DER VAART AND J. A. WELLNER 1836
- Stein’s method of normal approximation: Some recollections and reflections
LOUIS H. Y. CHEN 1850
- Second-order Stein: SURE for SURE and other applications in high-dimensional
inference PIERRE C. BELLEC AND CUN-HUI ZHANG 1864
- Consistent nonparametric estimation for heavy-tailed sparse graphs
CHRISTIAN BORGES, JENNIFER T. CHAYES,
HENRY COHN AND SHIRSHENDU GANGULY 1904
- Nonexchangeable random partition models for microclustering
GIUSEPPE DI BENEDETTO, FRANÇOIS CARON AND YEE WHYE TEH 1931
- Peskun–Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario
CHRISTOPHE ANDRIEU AND SAMUEL LIVINGSTONE 1958
- Volatility coupling JEAN JACOD, JIA LI AND ZHIPENG LIAO 1982
- Asymptotic distributions of high-dimensional distance correlation inference
LAN GAO, YINGYING FAN, JINCHI LV AND QI-MAN SHAO 1999
- Confidence intervals for multiple isotonic regression and other monotone models
HANG DENG, QIYANG HAN AND CUN-HUI ZHANG 2021
- On extended admissible procedures and their nonstandard Bayes risk
HAOSUI DUANMU AND DANIEL M. ROY 2053
- Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian
randomization TING YE, JUN SHAO AND HYUNSEUNG KANG 2079
- Boosted nonparametric hazards with time-dependent covariates
DONALD K. K. LEE, NINGYUAN CHEN AND HEMANT ISHWARAN 2101
- Universal Bayes consistency in metric spaces
STEVE HANNEKE, ARYEH KONTOROVICH, SIVAN SABATO AND ROI WEISS 2129
- Minimax optimal conditional independence testing
MATEY NEYKOV, SIVARAMAN BALAKRISHNAN AND LARRY WASSERMAN 2151
- Estimation of the number of components of nonparametric multivariate finite mixture
models CALEB KWON AND ERIC MBAKOP 2178
- Robust k -means clustering for distributions with two moments
YEGOR KLOCHKOV, ALEXEY KROSHNIN AND NIKITA ZHIVOTOVSKIY 2206
- On the rate of convergence of fully connected deep neural network regression estimates
MICHAEL KOHLER AND SOPHIE LANGER 2231
- Infinite-dimensional gradient-based descent for alpha-divergence minimisation
KAMÉLIA DAUDEL, RANDAL DOUC AND FRANÇOIS PORTIER 2250

continued

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

- Monitoring for a change point in a sequence of distributions
LAJOS HORVÁTH, PIOTR KOKOSZKA AND SHIXUAN WANG 2271
- What is resolution? A statistical minimax testing perspective on superresolution
microscopy GYTIS KULAITIS, AXEL MUNK AND FRANK WERNER 2292
- The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter
tuning LÉO MIOLANE AND ANDREA MONTANARI 2313
- Optimal change-point estimation in time series
NGAI HANG CHAN, WAI LEONG NG, CHUN YIP YAU AND HAIHAN YU 2336
- Propriety of the reference posterior distribution in Gaussian process modeling
JOSEPH MURÉ 2356
- Community detection with dependent connectivity YUBAI YUAN AND ANNIE QU 2378

Correction note

- “Statistical inference for the mean outcome under a possibly nonunique optimal treatment
rule” ALEX LUEDTKE, AURÉLIEN BIBAUT AND MARK VAN DER LAAN 2429

THE ANNALS OF STATISTICS

Vol. 49, No. 4, pp. 1811–2430 August 2021

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

President-Elect: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

Past President: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

The Annals of Statistics. *Editors:* Richard J. Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK. Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027, USA

The Annals of Applied Statistics. *Editor-in-Chief:* Karen Kafadar, Department of Statistics, University of Virginia, Heidelberg Institute for Theoretical Studies, Charlottesville, VA 22904-4135, USA

The Annals of Probability. *Editors:* Alice Guionnet, Unité de Mathématiques Pures et Appliquées, ENS de Lyon, Lyon, France. Christophe Garban, Institut Camille Jordan, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

The Annals of Applied Probability. *Editors:* François Delarue, Laboratoire J. A. Dieudonné, Université de Nice Sophia-Antipolis, France-06108 Nice Cedex 2. Peter Friz, Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany and Weierstrass-Institut für Angewandte Analysis und Stochastik, 10117 Berlin, Germany

Statistical Science. *Editor:* Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. *Editor:* Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

The Annals of Statistics [ISSN 0090-5364 (print); ISSN 2168-8966 (online)], Volume 49, Number 4, August 2021. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

EDITORIAL: MEMORIAL ISSUE FOR CHARLES STEIN

BY RICHARD J. SAMWORTH¹ AND MING YUAN²

¹*Statistical Laboratory, University of Cambridge, r.samworth@statslab.cam.ac.uk*

²*Department of Statistics, Columbia University, ming.yuan@columbia.edu*

REFERENCES

- CHEN, L. (2021). Stein’s method of normal approximation: Some recollections and reflections. *Ann. Statist.* **49** 1850–1863.
- EATON, M. L. and GEORGE, E. I. (2021). Charles Stein and invariance: Beginning with the Hunt–Stein theorem. *Ann. Statist.* **49** 1815–1822.
- SAMWORTH, R. (2003). A note on methods of restoring consistency to the bootstrap. *Biometrika* **90** 985–990. [MR2024773 https://doi.org/10.1093/biomet/90.4.985](https://doi.org/10.1093/biomet/90.4.985)
- SAMWORTH, R. (2005). Small confidence sets for the mean of a spherically symmetric distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 343–361. [MR2155342 https://doi.org/10.1111/j.1467-9868.2005.00505.x](https://doi.org/10.1111/j.1467-9868.2005.00505.x)
- STEIN, C. (1956a). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 187–195. Univ. California Press, Berkeley and Los Angeles. [MR0084921](https://doi.org/10.1111/j.1467-9868.2005.00505.x)
- STEIN, C. (1956b). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 197–206. Univ. California Press, Berkeley and Los Angeles. [MR0084922](https://doi.org/10.1111/j.1467-9868.2005.00505.x)
- STRAWDERMAN, W. E. (2021). On Charles Stein’s contributions to (in)admissibility. *Ann. Statist.* **49** 1823–1835.
- VAN DER VAART, A. W. and WELLNER, J. A. (2021). Stein 1956: Efficient nonparametric testing and estimation. *Ann. Statist.* **49** 1836–1849.

REFERENCES

- ANDERSON, T. W., STEIN, C. and ZAMAN, A. (1985). Best invariant estimation of a direction parameter. *Ann. Statist.* **13** 526–533. [MR0790554 https://doi.org/10.1214/aos/1176349536](https://doi.org/10.1214/aos/1176349536)
- BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics* 59–81. Academic Press, Boston, MA. [MR1031278](https://doi.org/10.1214/aos/1176349536)
- DIACONIS, P. and STEIN, C. (1978). Some Tauberian theorems related to coin tossing. *Ann. Probab.* **6** 483–490. [MR0474449 https://doi.org/10.1214/aop/1176995532](https://doi.org/10.1214/aop/1176995532)
- EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. [MR0615434](https://doi.org/10.1214/aos/1176349536)
- GIRI, N., KIEFER, J. and STEIN, C. (1963). Minimax character of Hotelling’s T^2 test in the simplest case. *Ann. Math. Stat.* **34** 1524–1535. [MR0156408 https://doi.org/10.1214/aoms/1177703884](https://doi.org/10.1214/aoms/1177703884)
- JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. [MR0133191](https://doi.org/10.1214/aoms/1177703884)
- LEHMANN, E. L. and STEIN, C. (1948). Most powerful tests of composite hypotheses. I. Normal distributions. *Ann. Math. Stat.* **19** 495–516. [MR0030167 https://doi.org/10.1214/aoms/1177730147](https://doi.org/10.1214/aoms/1177730147)
- LEHMANN, E. L. and STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Stat.* **20** 28–45. [MR0030168](https://doi.org/10.1214/aoms/1177730147)
- LEHMANN, E. L. and STEIN, C. (1950). Completeness in the sequential case. *Ann. Math. Stat.* **21** 376–385. [MR0036969 https://doi.org/10.1214/aoms/1177729796](https://doi.org/10.1214/aoms/1177729796)
- LEHMANN, E. L. and STEIN, C. M. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Stat.* **24** 473–479. [MR0056249 https://doi.org/10.1214/aoms/1177728987](https://doi.org/10.1214/aoms/1177728987)
- PORTNOY, S. and STEIN, C. (1971). Inadmissibility of the best invariant test in three or more dimensions. *Ann. Math. Stat.* **42** 799–801. [MR0286211 https://doi.org/10.1214/aoms/1177693435](https://doi.org/10.1214/aoms/1177693435)
- STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.* **16** 243–258. [MR0013885 https://doi.org/10.1214/aoms/1177731088](https://doi.org/10.1214/aoms/1177731088)
- STEIN, C. (1946). A note on cumulative sums. *Ann. Math. Stat.* **17** 498–499. [MR0018369 https://doi.org/10.1214/aoms/1177730890](https://doi.org/10.1214/aoms/1177730890)
- STEIN, C. (1950). Unbiased estimates with minimum variance. *Ann. Math. Stat.* **21** 406–415. [MR0036968 https://doi.org/10.1214/aoms/1177729799](https://doi.org/10.1214/aoms/1177729799)

- STEIN, C. M. (1951). A property of some tests of composite hypotheses. *Ann. Math. Stat.* **22** 475–476. [MR0042651 https://doi.org/10.1214/aoms/1177729599](https://doi.org/10.1214/aoms/1177729599)
- STEIN, C. M. (1953). *A Two-Sample Test for a Linear Hypothesis Having Power Independent of the Variance*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Columbia University. [MR2938394](https://doi.org/10.1214/aoms/1177728497)
- STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Stat.* **26** 518–522. [MR0070929 https://doi.org/10.1214/aoms/1177728497](https://doi.org/10.1214/aoms/1177728497)
- STEIN, C. (1956a). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 187–195. Univ. California Press, Berkeley and Los Angeles. [MR0084921](https://doi.org/10.1214/aoms/1177728497)
- STEIN, C. (1956b). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 197–206. Univ. California Press, Berkeley and Los Angeles. [MR0084922](https://doi.org/10.1214/aoms/1177728497)
- STEIN, C. (1956c). The admissibility of Hotelling's T^2 -test. *Ann. Math. Stat.* **27** 616–623. [MR0080413 https://doi.org/10.1214/aoms/1177728171](https://doi.org/10.1214/aoms/1177728171)
- STEIN, C. (1959a). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Stat.* **30** 877–880. [MR0125680 https://doi.org/10.1214/aoms/1177706072](https://doi.org/10.1214/aoms/1177706072)
- STEIN, C. (1959b). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Stat.* **30** 970–979. [MR0109392 https://doi.org/10.1214/aoms/1177706080](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1960). Multiple regression. In *Contributions to Probability and Statistics* 424–443. Stanford Univ. Press, Stanford, CA. [MR0120718](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1962a). A remark on the likelihood principle. *J. Roy. Statist. Soc. Ser. A* **125** 565–568. [MR0149580 https://doi.org/10.2307/2982612](https://doi.org/10.2307/2982612)
- STEIN, C. M. (1962b). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296. [MR0148184](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160. [MR0171344 https://doi.org/10.1007/BF02868569](https://doi.org/10.1007/BF02868569)
- STEIN, C. (1965). Approximation of improper prior measures by prior probability measures. In *Proc. Internat. Res. Sem., Statist. Lab., Univ. California, Berkeley, Calif., 1963* 217–240. Springer, New York. [MR0199937](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics (Festschrift J. Neyman)* 351–366. John Wiley, London. [MR0210232](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. M. (1968). Some recent developments in mathematical statistics. In *Proc. Internat. Congr. Math. (Moscow, 1966)* 140–157. Izdat. "Mir", Moscow. [MR0233444](https://doi.org/10.1214/aoms/1177706080)
- 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. [MR0402873](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1974). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague Symposium on Asymptotic Statistics (Charles Univ., Prague, 1973), Vol. II* 345–381. [MR0381062](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1977). Lectures on multivariate estimation theory. (In Russian). *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **74** 4–65.
- STEIN, C. M. (1978). Asymptotic evaluation of the number of Latin rectangles. *J. Combin. Theory Ser. A* **25** 38–49. [MR0499035 https://doi.org/10.1016/0097-3165\(78\)90029-8](https://doi.org/10.1016/0097-3165(78)90029-8)
- STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. M. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics. Banach Center Publ.* **16** 485–514. PWN, Warsaw. [MR0847495](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series 7*. IMS, Hayward, CA. [MR0882007](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. (1992). A way of using auxiliary randomization. In *Probability Theory (Singapore, 1989)* 159–180. de Gruyter, Berlin. [MR1188718](https://doi.org/10.1214/aoms/1177706080)
- STEIN, C. and WALD, A. (1947). Sequential confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Stat.* **18** 427–433. [MR0022061 https://doi.org/10.1214/aoms/1177730389](https://doi.org/10.1214/aoms/1177730389)
- STEIN, C., DIACONIS, P., HOLMES, S. and REINERT, G. (2004). Use of exchangeable pairs in the analysis of simulations. In *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes—Monograph Series 46* 1–26. IMS, Beachwood, OH. [MR2118600 https://doi.org/10.1214/aoms/1177706080](https://doi.org/10.1214/aoms/1177706080)

CHARLES STEIN AND INVARIANCE: BEGINNING WITH THE HUNT-STEIN THEOREM

BY MORRIS L. EATON¹ AND EDWARD I. GEORGE²

¹*School of Statistics, University of Minnesota, eaton@stat.umn.edu*

²*Department of Statistics, University of Pennsylvania, edgeorge@wharton.upenn.edu*

When statistical decision theory was emerging as a promising new paradigm, Charles Stein was to play a major role in the development of minimax theory for invariant statistical problems. In some of his earliest work with Gil Hunt, he set out to prove that, in problems where invariant procedures have constant risk, any best invariant test would be minimax among all tests. Although finding it not quite true in general, this led to the legendary Hunt–Stein theorem, which established the result under restrictive conditions on the underlying group of transformations. In decision problems invariant under such suitable groups, an overall minimax test was guaranteed to reside within the class of invariant procedures where it would typically be much easier to find. But when it did not seem possible to establish this result for invariance under the full linear group, he instead turned to prove its impossibility with counterexamples such as the nonminimaxity of the usual sample covariance estimator where the full linear group was just too big for the Hunt–Stein theorem to apply. Further explorations of invariance such as the sometimes problematic inference under a fiducial distribution, or the characterization of a best invariant procedure as a formal Bayes procedure under a right Haar prior, are further examples of the far reaching influence of Stein’s contributions to invariance theory.

REFERENCES

- BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR0804611](#) <https://doi.org/10.1007/978-1-4757-4286-2>
- BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. Verw. Gebiete* **57** 103–128. [MR0623457](#) <https://doi.org/10.1007/BF00533716>
- BROWN, L. D. (1986). Commentary on paper [19]. In *Collected Papers of Jack Carl Kiefer. Supplementary Volume* (L. D. Brown, I. Olkin, J. Sacks and H. P. Wynn, eds.) 20–27. Springer-Verlag and IMS.
- DEGROOT, M. H. (1986). A conversation with Charles Stein. *Statist. Sci.* **1** 454–462. [MR0874478](#)
- EATON, M. L. (1970). Some problems in covariance estimation. Stanford University Technical Report No. 49. (This report is available at the website <https://statistics.stanford.edu/resources/technical-reports>).
- EATON, M. L. (1989). *Group Invariance Applications in Statistics*. Institute of Mathematical Statistics, Beachwood, OH.
- EATON, M. L. and FREEDMAN, D. A. (2004). Dutch book against some ‘objective’ priors. *Bernoulli* **10** 861–872. [MR2093614](#) <https://doi.org/10.3150/bj/1099579159>
- EATON, M. L., MUIRHEAD, R. J. and PICKERING, E. H. (2006). Assessing a vector of clinical observations. *J. Statist. Plann. Inference* **136** 3383–3414. [MR2256273](#) <https://doi.org/10.1016/j.jspi.2005.02.019>
- EATON, M. L. and SUDDERTH, W. D. (1995). The formal posterior of a standard flat prior in MANOVA is incoherent. *Journal of the Italian Statistical Society* **2** 251–270.
- EATON, M. L. and SUDDERTH, W. D. (1999). Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli* **5** 833–854. [MR1715441](#) <https://doi.org/10.2307/3318446>
- EATON, M. L. and SUDDERTH, W. D. (2001). Best invariant predictive distributions. In *Algebraic Methods in Statistics and Probability (Notre Dame, IN, 2000)* (M. A. G. Viana and D. S. P. Richards, eds.). *Contemp. Math.* **287** 49–62. Amer. Math. Soc., Providence, RI. [MR1873666](#) <https://doi.org/10.1090/conm/287/04776>

- EATON, M. L. and SUDDERTH, W. D. (2002). Group invariant inference and right Haar measure. *J. Statist. Plann. Inference* **103** 87–99. [MR1896985 https://doi.org/10.1016/S0378-3758\(01\)00199-9](#)
- EATON, M. L. and SUDDERTH, W. D. (2004). Properties of right Haar predictive inference. *Sankhyā* **66** 487–512. [MR2108204](#)
- GREENLEAF, F. P. (1969). *Invariant Means on Topological Groups and Their Applications*. Van Nostrand Mathematical Studies, No. 16. Van Nostrand Reinhold Co., New York-Toronto, Ont.-London. [MR0251549](#)
- HORA, R. B. (1964). Fiducial probability theory for distributions with a group structure. Ph.D. thesis and Technical Report No. 33, Dept. Statistics, Univ. Minnesota.
- HORA, R. B. and BUEHLER, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Stat.* **37** 643–656. [MR0199938 https://doi.org/10.1214/aoms/1177699458](#)
- JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. [MR0133191](#)
- KASS, R. E. and WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91** 1343–1370.
- KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Stat.* **28** 573–601. [MR0092325 https://doi.org/10.1214/aoms/1177706874](#)
- KIEFER, J. (1966). Multivariate optimality results. In *Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965)* 255–274. Academic Press, New York. [MR0226785](#)
- KUDO, H. (1955). On minimax invariant estimates of the transformation parameter. *Natur. Sci. Rep. Ochanomizu Univ.* **6** 31–73. [MR0077842](#)
- LEHMANN, E. L. (1950). Some principles of the theory of testing hypotheses. *Ann. Math. Stat.* **21** 1–26. [MR0034001 https://doi.org/10.1214/aoms/1177729884](#)
- LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*, 1st ed. Wiley, New York. [MR0107933](#)
- LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. [MR1639875](#)
- OLKIN, I. and SELLIAH, J. B. (1977). Estimating covariances in a multivariate normal distribution. In *Statistical Decision Theory and Related Topics, II (Proc. Sympos., Purdue Univ., Lafayette, Ind., 1976)* (S. S. Gupta and D. S. Moore, eds.) 313–326. Academic Press, New York. [MR0436407](#)
- PEISAKOFF, M. P. (1950). Transformation parameters. Ph.D. dissertation from Princeton Univ., unpublished.
- PITMAN, E. J. G. (1939a). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.
- PITMAN, E. J. G. (1939b). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215. [MR0000382 https://doi.org/10.1093/biomet/31.1-2.200](#)
- ROBERT, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York.
- SELLIAH, J. B. (1964). Estimation and testing problems in a Wishart distribution. Ph.D. thesis and Technical Report No. 10, Dept. Statistics, Stanford Univ.
- STEIN, C. (1956a). Some problems in multivariate analysis, Part I. Stanford University Technical Report No. 6. (This report is available at the website <https://statistics.stanford.edu/resources/technical-reports>).
- STEIN, C. (1956b). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 197–206. Univ. California Press, Berkeley and Los Angeles. [MR0084922](#)
- STEIN, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Stat.* **30** 877–880. [MR0125680 https://doi.org/10.1214/aoms/1177706072](#)
- STEIN, C. (1965). Approximation of improper prior measures by prior probability measures. In *Proc. Internat. Res. Sem., Statist. Lab., Univ. California, Berkeley, Calif., 1963* (J. Neyman and L. LeCam, eds.) 217–240. Springer, New York. [MR0199937](#)
- VON NEUMAN, J. (1929). Zur allgemeinen theorie des masses. *Fund. Math.* **13** 73–116.
- WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Stat.* **10** 299–326. [MR0000932 https://doi.org/10.1214/aoms/1177732144](#)
- WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482. [MR0012401 https://doi.org/10.2307/1990256](#)
- ZIDEK, J. V. (1969). A representation of Bayes invariant procedures in terms of Haar measure. *Ann. Inst. Statist. Math.* **21** 291–308. [MR0246431 https://doi.org/10.1007/BF02532257](#)

ON CHARLES STEIN'S CONTRIBUTIONS TO (IN)ADMISSIBILITY

BY WILLIAM E. STRAWDERMAN

Department of Statistics, Rutgers University, straw@stat.rutgers.edu

Charles Stein made fundamental contributions to admissibility and inadmissibility in estimation and testing. This paper surveys some of the more important ones. Particular attention will be paid to his monumentally important, and at the time, incredibly surprising discovery of the inadmissibility of the usual estimator of the mean in three and higher dimensions. His result on admissibility of Pitman's estimator of a mean in one and two dimensions, and his results on estimation of a mean matrix and a covariance matrix are also discussed. His work on testing is briefly covered.

REFERENCES

- [1] BLYTH, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Stat.* **22** 22–42. [MR0039966](#) <https://doi.org/10.1214/aoms/1177729690>
- [2] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. [MR0286209](#) <https://doi.org/10.1214/aoms/1177693318>
- [3] EFRON, B. and MORRIS, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59** 335–347. [MR0334386](#) <https://doi.org/10.1093/biomet/59.2.335>
- [4] HAFF, L. R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9** 531–544. [MR0556910](#) [https://doi.org/10.1016/0047-259X\(79\)90056-3](https://doi.org/10.1016/0047-259X(79)90056-3)
- [5] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. i* 361–379. Univ. California Press, Berkeley, CA. [MR0133191](#)
- [6] LEHMANN, E. L. and STEIN, C. (1948). Most powerful tests of composite hypotheses. I. Normal distributions. *Ann. Math. Stat.* **19** 495–516. [MR0030167](#) <https://doi.org/10.1214/aoms/1177730147>
- [7] LEHMANN, E. L. and STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Stat.* **20** 28–45. [MR0030168](#)
- [8] LEHMANN, E. L. and STEIN, C. M. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Stat.* **24** 473–479. [MR0056249](#) <https://doi.org/10.1214/aoms/1177728987>
- [9] PORTNOY, S. and STEIN, C. (1971). Inadmissibility of the best invariant test in three or more dimensions. *Ann. Math. Stat.* **42** 799–801. [MR0286211](#) <https://doi.org/10.1214/aoms/1177693435>
- [10] STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Stat.* **26** 518–522. [MR0070929](#) <https://doi.org/10.1214/aoms/1177728497>
- [11] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. i* 197–206. Univ. California Press, Berkeley and Los Angeles. [MR0084922](#)
- [12] STEIN, C. (1956). The admissibility of Hotelling's T^2 -test. *Ann. Math. Stat.* **27** 616–623. [MR0080413](#) <https://doi.org/10.1214/aoms/1177728171>
- [13] STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Stat.* **30** 970–979. [MR0109392](#) <https://doi.org/10.1214/aoms/1177706080>
- [14] STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160. [MR0171344](#) <https://doi.org/10.1007/BF02868569>
- [15] STEIN, C. (1965). Approximation of improper prior measures by prior probability measures. In *Proc. Internat. Res. Sem., Statist. Lab., Univ. California, Berkeley, Calif., 1963* 217–240. Springer, New York. [MR0199937](#)
- [16] STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics (Festschrift J. Neyman)* 351–366. John Wiley, London. [MR0210232](#)

- [17] STEIN, C. (1974). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague Symposium on Asymptotic Statistics (Charles Univ., Prague, 1973), Vol. II* 345–381. [MR0381062](#)
- [18] STEIN, C. (1975). Estimation of the parameters of a multivariate normal distribution II: Estimation of a covariance matrix. Unpublished Notes. Stanford University.
- [19] STEIN, C. (1986). Lectures on the theory of estimation of many parameters. *J. Math. Sci.* **34** 1373–1403. Also, Stanford University technical report.
- [20] STEIN, C. M. (1951). A property of some tests of composite hypotheses. *Ann. Math. Stat.* **22** 475–476. [MR0042651](#) <https://doi.org/10.1214/aoms/1177729599>
- [21] STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296. [MR0148184](#)
- [22] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [23] STEIN, Č. (1977). Lectures on multivariate estimation theory. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **74** 4–65, 146, 148. Presented at LOMI, 1976. Russian text prepared by the author with the assistance of M. Ermakov, A. Borodin, A. Makšanov and M. S. Nikulin, Studies in statistical estimation theory, I. [MR0501548](#)

STEIN 1956: EFFICIENT NONPARAMETRIC TESTING AND ESTIMATION

BY A. W. VAN DER VAART¹ AND J. A. WELLNER²

¹*Mathematical Institute, Leiden University, avdvaart@math.leidenuniv.nl*

²*Department of Statistics, University of Washington, jaw@stat.washington.edu*

We revisit a paper by Charles Stein, and discuss its follow-up.

REFERENCES

- [1] BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452. [MR0696057](#) <https://doi.org/10.1214/aos/1176346151>
- [2] BERAN, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74. [MR0345295](#)
- [3] BERAN, R. (1977). Robust location estimates. *Ann. Statist.* **5** 431–444. [MR0448699](#)
- [4] BERAN, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463. [MR0448700](#)
- [5] BERAN, R. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313. [MR0518885](#)
- [6] BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. [MR0663424](#)
- [7] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins Univ. Press, Baltimore, MD. [MR1245941](#)
- [8] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. Reprint of the 1993 original. [MR1623559](#)
- [9] BICKEL, P. J. and RITOV, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.* **15** 513–540. [MR0888423](#) <https://doi.org/10.1214/aos/1176350358>
- [10] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. [MR1065550](#)
- [11] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. [MR0341758](#)
- [12] COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39. With a discussion. [MR0893334](#)
- [13] COX, D. R. and REID, N. (1989). On the stability of maximum-likelihood estimators of orthogonal parameters. *Canad. J. Statist.* **17** 229–233. [MR1033105](#) <https://doi.org/10.2307/3314851>
- [14] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. [MR1105839](#) <https://doi.org/10.1214/aos/1176348114>
- [15] DONOHO, D. L. and LOW, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970. [MR1165601](#) <https://doi.org/10.1214/aos/1176348665>
- [16] EFRON, B. (1977). The efficiency of Cox’s likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565. [MR0451514](#)
- [17] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. DMV Seminar **19**. Birkhäuser, Basel. [MR1180321](#) <https://doi.org/10.1007/978-3-0348-8621-5>
- [18] HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Stat.* **33** 1124–1147. [MR0143304](#) <https://doi.org/10.1214/aoms/1177704476>
- [19] HÁJEK, J. (1969/70). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330. [MR0283911](#) <https://doi.org/10.1007/BF00533669>
- [20] HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: *Theory of Statistics* 175–194. [MR0400513](#)
- [21] HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. [MR1963675](#) <https://doi.org/10.3150/bj/1068129013>

- [22] HASMINSKII, R. Z. and IBRAGIMOV, I. A. (1983). On asymptotic efficiency in the presence of an infinite-dimensional nuisance parameter. In *Probability Theory and Mathematical Statistics (Tbilisi, 1982). Lecture Notes in Math.* **1021** 195–229. Springer, Berlin. MR0735986 <https://doi.org/10.1007/BFb0072916>
- [23] IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Applications of Mathematics* **16**. Springer, New York-Berlin. Asymptotic theory, Translated from the Russian by Samuel Kotz. MR0620321
- [24] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 887–906. MR0086464 <https://doi.org/10.1214/aoms/1177728066>
- [25] KLAASSEN, C. A. J. and VAN ZWET, W. R. (1985). On estimating a parameter and its score function. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983). Wadsworth Statist./Probab. Ser.* 827–839. Wadsworth, Belmont, CA. MR0822068
- [26] KOŠEVNIK, J. A. and LEVIT, B. J. (1976). On a nonparametric analogue of the information matrix. *Teor. Veroyatn. Primen.* **21** 759–774. MR0428578
- [27] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York. MR2724368 <https://doi.org/10.1007/978-0-387-74978-5>
- [28] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 <https://doi.org/10.1214/aos/1015957395>
- [29] LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 129–156. Univ. California Press, Berkeley and Los Angeles. MR0084918
- [30] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. Calif. Publ. Stat.* **3** 37–98. MR0126903
- [31] LE CAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Stat.* **41** 802–828. MR0267676 <https://doi.org/10.1214/aoms/1177696960>
- [32] LE CAM, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of Statistics* 245–261. MR0415819
- [33] LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. Calif. Publ. Stat.* **1** 277–329. MR0054913
- [34] LEVIT, B. J. (1975). Conditional estimation of linear functionals. *Problemy Peredachi Informatsii* **11** 39–54. MR0494664
- [35] LEVIT, B. J. (1978). Infinite-dimensional informational bounds. *Teor. Veroyatn. Primen.* **23** 388–394. MR0488446
- [36] LEVIT, B. Y. Asymptotically efficient estimation of nonlinear functionals. MR0533450
- [37] MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252. MR0537670 <https://doi.org/10.1007/BF00537522>
- [38] MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. In *Eleventh Saint Flour Probability Summer School—1981 (Saint Flour, 1981). Lecture Notes in Math.* **976** 75–265. Springer, Berlin. MR0722983 <https://doi.org/10.1007/BFb0067986>
- [39] MILLAR, P. W. (1985). Nonparametric applications of an infinite-dimensional convolution theorem. *Z. Wahrsch. Verw. Gebiete* **68** 545–556. MR0772198 <https://doi.org/10.1007/BF00535344>
- [40] MURPHY, S. A. and VAN DER VAART, A. W. (1996). Likelihood inference in the errors-in-variables model. *J. Multivariate Anal.* **59** 81–108. MR1424904 <https://doi.org/10.1006/jmva.1996.0055>
- [41] MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. With comments and a rejoinder by the authors. MR1803168 <https://doi.org/10.2307/2669386>
- [42] NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32. MR0025113 <https://doi.org/10.2307/1914288>
- [43] PARK, B. U. (1990). Efficient estimation in the two-sample semiparametric location-scale models. *Probab. Theory Related Fields* **86** 21–39. MR1061946 <https://doi.org/10.1007/BF01207511>
- [44] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics* **13**. Springer, New York-Berlin. With the assistance of W. Wefelmeyer. MR0675954
- [45] PFANZAGL, J. (1993). Incidental versus random nuisance parameters. *Ann. Statist.* **21** 1663–1691. MR1245763 <https://doi.org/10.1214/aos/1176349392>
- [46] REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18** 375–389. MR0038054 <https://doi.org/10.2307/1907835>
- [47] ROBINS, J., TCHETGEN TCHETGEN, E., LI, L. and VAN DER VAART, A. (2009). Semiparametric minimax rates. *Electron. J. Stat.* **3** 1305–1321. MR2566189 <https://doi.org/10.1214/09-EJS479>

- [48] ROBINS, J. M., LI, L., MUKHERJEE, R., TCHETGEN, E. T. and VAN DER VAART, A. (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist.* **45** 1951–1987. [MR3718158](#) <https://doi.org/10.1214/16-AOS1515>
- [49] RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. [MR1998720](#) <https://doi.org/10.1017/CBO9780511755453>
- [50] SACKS, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3** 285–298. [MR0359174](#)
- [51] SLUD, E. V. and VONTA, F. (2005). Efficient semiparametric estimators via modified profile likelihood. *J. Statist. Plann. Inference* **129** 339–367. [MR2126854](#) <https://doi.org/10.1016/j.jspi.2004.06.057>
- [52] STEIN, C. (1956). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, Vol. I 187–195. Univ. California Press, Berkeley and Los Angeles. [MR0084921](#)
- [53] STIGLER, S. M. (2007). The epic story of maximum likelihood. *Statist. Sci.* **22** 598–620. [MR2410255](#) <https://doi.org/10.1214/07-STS249>
- [54] STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284. [MR0362669](#)
- [55] STRASSER, H. (1998). Perturbation invariant estimates and incidental nuisance parameters. *Math. Methods Statist.* **7** 1–26. [MR1626568](#)
- [56] TAKEUCHI, K. (1971). A uniformly asymptotically efficient estimator of a location parameter. *J. Amer. Statist. Assoc.* **66** 292–301.
- [57] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. [MR2233926](#)
- [58] VAN DER VAART, A. (1989). On the asymptotic information bound. *Ann. Statist.* **17** 1487–1500. [MR1026295](#) <https://doi.org/10.1214/aos/1176347377>
- [59] VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. [MR1091845](#) <https://doi.org/10.1214/aos/1176347976>
- [60] VAN DER VAART, A. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24** 862–878. [MR1394993](#) <https://doi.org/10.1214/aos/1032894470>
- [61] VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1999). Lecture Notes in Math.* **1781** 331–457. Springer, Berlin. [MR1915446](#)
- [62] VAN DER VAART, A. W. (1988). *Statistical Estimation in Large Parameter Spaces. CWI Tract* **44**. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam. [MR0927725](#)
- [63] VAN DER VAART, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* **16** 1450–1474. [MR0964933](#) <https://doi.org/10.1214/aos/1176351048>
- [64] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#) <https://doi.org/10.1017/CBO9780511802256>
- [65] VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Stat.* **41** 172–181. [MR0263194](#) <https://doi.org/10.1214/aoms/1177697197>
- [66] VAN DER VAART, A. (1991). An asymptotic representation theorem. *Int. Stat. Rev.* **59** 97–121.
- [67] WEISS, L. and WOLFOWITZ, J. (1970). Asymptotically efficient non-parametric estimators of location and scale parameters. *Z. Wahrsch. Verw. Gebiete* **16** 134–150. [MR0323023](#) <https://doi.org/10.1007/BF00535694>
- [68] WELLNER, J. A., KLAASSEN, C. A. J. and RITOV, Y. (2006). Semiparametric models: A review of progress since BKRW (1993). In *Frontiers in Statistics* 25–44. Imp. Coll. Press, London. [MR2325995](#) https://doi.org/10.1142/9781860948886_0002
- [69] WOLFOWITZ, J. (1974). Asymptotically efficient non-parametric estimators of location and scale parameters. II. *Z. Wahrsch. Verw. Gebiete* **30** 117–128. [MR0403054](#) <https://doi.org/10.1007/BF00532264>
- [70] WONG, W. H. and SEVERINI, T. A. (1991). On maximum likelihood estimation in infinite-dimensional parameter spaces. *Ann. Statist.* **19** 603–632. [MR1105838](#) <https://doi.org/10.1214/aos/1176348113>
- [71] YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. [MR1462963](#)

STEIN'S METHOD OF NORMAL APPROXIMATION: SOME RECOLLECTIONS AND REFLECTIONS

BY LOUIS H. Y. CHEN

Department of Mathematics, National University of Singapore, matchyl@nus.edu.sg

This paper is a short exposition of Stein's method of normal approximation from my personal perspective. It focuses mainly on the characterization of the normal distribution and the construction of Stein identities. Through examples, it provides glimpses into the many approaches to constructing Stein identities and the diverse applications of Stein's method to mathematical problems. It also includes anecdotes of historical interest, including how Stein discovered his method and how I found an unpublished proof of his of the Berry–Esseen theorem.

REFERENCES

- [1] BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics* (T. W. Anderson, K. B. Athreya and D. L. Iglehart, eds.) 59–81. Academic Press, Boston, MA. [MR1031278](#)
- [2] BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25**(A) 175–184. [MR0974580](#)
- [3] BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. [MR1035659](#) <https://doi.org/10.1007/BF01197887>
- [4] BARBOUR, A. D. and CHEN, L. H. Y. (2014). Stein's (magic) method. Available at [arXiv:1411.1179](#).
- [5] BHATTACHARJEE, C. and GOLDSTEIN, L. (2019). Dickman approximation in simulation, summations and perpetuities. *Bernoulli* **25** 2758–2792. [MR4003564](#) <https://doi.org/10.3150/18-BEJ1070>
- [6] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. [MR2449121](#) <https://doi.org/10.1007/s00440-007-0118-6>
- [7] CHATTERJEE, S. (2014). A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV* 1–24. Kyung Moon Sa, Seoul. [MR3727600](#)
- [8] CHATTERJEE, S. (2019). Central limit theorem for the free energy of the random field Ising model. *J. Stat. Phys.* **175** 185–202. [MR3927223](#) <https://doi.org/10.1007/s10955-019-02249-9>
- [9] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. [MR0428387](#) <https://doi.org/10.1214/aop/1176996359>
- [10] CHEN, L. HY. (1986). The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. *IMA Preprint Series* **243**.
- [11] CHEN, L. H. Y. (1998). Stein's method: Some perspectives with applications. In *Probability Towards 2000* (New York, 1995). (L. Accardi and C. C. Heyde, eds.) *Lect. Notes Stat.* **128** 97–122. Springer, New York. [MR1632651](#) https://doi.org/10.1007/978-1-4612-2224-8_6
- [12] CHEN, L. H. Y. and FANG, X. (2015). On the error bound in a combinatorial central limit theorem. *Bernoulli* **21** 335–359. [MR3322321](#) <https://doi.org/10.3150/13-BEJ569>
- [13] CHEN, L. HY. and FANG, X. (2015). Multivariate normal approximation by Stein's method: The concentration inequality approach. Available at [arXiv:1111.4073](#).
- [14] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (New York). Springer, Heidelberg. [MR2732624](#) <https://doi.org/10.1007/978-3-642-15007-4>
- [15] CHEN, L. H. Y. and RÖLLIN, A. (2010). Stein couplings for normal approximation. Available at [arXiv:1003.6039v2](#).
- [16] CHEN, L. H. Y., RÖLLIN, A. and XIA, A. (2021). Palm theory, random measures and Stein couplings. *Ann. Appl. Probab.* To appear. Available at [arXiv:2004.05026](#).

MSC2020 subject classifications. Primary 60F05, 62E17; secondary 60B10.

Key words and phrases. Stein's method, normal approximation, Stein equation, Stein identity, concentration inequality, exchangeable pair.

- [17] CHEN, L. H. Y. and SHAO, Q.-M. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Related Fields* **120** 236–254. [MR1841329](#) <https://doi.org/10.1007/PL00008782>
- [18] CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. [MR2073183](#) <https://doi.org/10.1214/009117904000000450>
- [19] CHEN, L. H. Y. and SHAO, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. [MR2331265](#) <https://doi.org/10.3150/07-BEJ5164>
- [20] CHWIAKOWSKI, K., STRATHMANN, H. and GRETTON, A. (2016). A kernel test of goodness of fit. In *International Conference on Machine Learning* 2606–2615. PMLR.
- [21] DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5** 72–81. [MR0422186](#) <https://doi.org/10.1214/aop/1176995891>
- [22] DIACONIS, P. and HOLMES, S. eds. (2004). *Stein's Method: Expository Lectures and Applications*. IMS Lecture Notes-Monogr. Ser. **46**. IMS, Beachwood, OH.
- [23] FANG, X., PENG, S., SHAO, Q.-M. and SONG, Y. (2019). Limit theorems with rate of convergence under sublinear expectations. *Bernoulli* **25** 2564–2596. [MR4003558](#) <https://doi.org/10.3150/18-BEJ1063>
- [24] FATHI, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. [MR3980918](#) <https://doi.org/10.1214/18-AOP1305>
- [25] FATHI, M., GOLDSTEIN, L., REINERT, G. and SAUMARD, A. (2020). Relaxing the Gaussian assumption in shrinkage and SURE in high dimension. Available at [arXiv:2004.01378](#).
- [26] GOLDSTEIN, L. (2010). Bounds on the constant in the mean central limit theorem. *Ann. Probab.* **38** 1672–1689. [MR2663641](#) <https://doi.org/10.1214/10-AOP527>
- [27] GOLDSTEIN, L. and REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. [MR1484792](#) <https://doi.org/10.1214/aoap/1043862419>
- [28] GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. [MR1371949](#) <https://doi.org/10.1017/s0021900200103675>
- [29] GORHAM, J. and MACKEY, L. (2015). Measuring sample quality with Stein's method. In *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, eds.) **28**. Curran Associates, Red Hook, NY.
- [30] GORHAM, J. (2017). Measuring sample quality with kernels. In *International Conference on Machine Learning* 1292–1301. PMLR.
- [31] HERRY, R. (2018). Contributions to functional inequalities and limit theorems on the configuration space Ph.D. thesis, Univ. Luxembourg.
- [32] HO, S. T. and CHEN, L. H. Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6** 231–249. [MR0478291](#) <https://doi.org/10.1214/aop/1176995570>
- [33] Hoeffding, W. (1951). A combinatorial central limit theorem. *Ann. Math. Stat.* **22** 558–566. [MR0044058](#) <https://doi.org/10.1214/aoms/1177729545>
- [34] KALLENBERG, O. (1983). *Random Measures*, 3rd ed. Academic Press, London. [MR0818219](#)
- [35] LACHÈZE-REY, R., SCHULTE, M. and YUKICH, J. E. (2019). Normal approximation for stabilizing functionals. *Ann. Appl. Probab.* **29** 931–993. [MR3910021](#) <https://doi.org/10.1214/18-AAP1405>
- [36] LEDOUX, M., NOURDIN, I. and PECCATI, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** 256–306. [MR3320893](#) <https://doi.org/10.1007/s00039-015-0312-0>
- [37] LIU, Q., LEE, J. and JORDAN, M. (2016). A kernelized Stein discrepancy for good-of-fit tests. In *International Conference on Machine Learning* 276–284. PMLR.
- [38] LIU, Q. and WANG, D. (2016). Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems* 2378–2386.
- [39] LUK, H. M. (1994). *Stein's Method for the Gamma Distribution and Related Statistical Applications*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.), Univ. Southern California. [MR2693204](#)
- [40] NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. [MR2520122](#) <https://doi.org/10.1007/s00440-008-0162-x>
- [41] REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. [MR2573554](#) <https://doi.org/10.1214/09-AOP467>
- [42] RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U -statistics. *Ann. Appl. Probab.* **7** 1080–1105. [MR1484798](#) <https://doi.org/10.1214/aoap/1043862425>
- [43] RÖLLIN, A. (2008). A note on the exchangeability condition in Stein's method. *Statist. Probab. Lett.* **78** 1800–1806. [MR2453918](#) <https://doi.org/10.1016/j.spl.2008.01.043>
- [44] RUDIN, W. (1986). *Real and Complex Analysis*. McGraw-Hill, New York.

- [45] 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. [MR0402873](#)
- [46] STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* 7. IMS, Hayward, CA. [MR0882007](#)
- [47] STEIN, C. (1992). A way of using auxiliary randomization. In *Probability Theory* (L. H. Y. Chen, K. P. Choi, K. Hu and J.-H. Lou, eds.) 159–180. de Gruyter, Berlin. [MR1188718](#)
- [48] WALD, A. and WOLFOWITZ, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Stat.* **15** 358–372. [MR0011424](#) <https://doi.org/10.1214/aoms/1177731207>

SECOND-ORDER STEIN: SURE FOR SURE AND OTHER APPLICATIONS IN HIGH-DIMENSIONAL INFERENCE

BY PIERRE C. BELLEC^{*} AND CUN-HUI ZHANG[†]

Department of Statistics, Rutgers University, ^{*}pierre.bellec@rutgers.edu; [†]czhang@stat.rutgers.edu

Stein's formula states that a random variable of the form $\mathbf{z}^\top f(\mathbf{z}) - \operatorname{div} f(\mathbf{z})$ is mean-zero for all functions f with integrable gradient. Here, $\operatorname{div} f$ is the divergence of the function f and $\mathbf{z}$ is a standard normal vector. This paper aims to propose a second-order Stein formula to characterize the variance of such random variables for all functions $f(\mathbf{z})$ with square integrable gradient, and to demonstrate the usefulness of this second-order Stein formula in various applications.

In the Gaussian sequence model, a remarkable consequence of Stein's formula is Stein's Unbiased Risk Estimate (SURE), an unbiased estimate of the mean squared risk for almost any given estimator $\hat{\boldsymbol{\mu}}$ of the unknown mean vector. A first application of the second-order Stein formula is an Unbiased Risk Estimate for SURE itself (SURE for SURE): an unbiased estimate providing information about the squared distance between SURE and the squared estimation error of $\hat{\boldsymbol{\mu}}$. SURE for SURE has a simple form as a function of the data and is applicable to all $\hat{\boldsymbol{\mu}}$ with square integrable gradient, for example, the Lasso and the Elastic Net.

In addition to SURE for SURE, the following statistical applications are developed: (1) upper bounds on the risk of SURE when the estimation target is the mean squared error; (2) confidence regions based on SURE and using the second-order Stein formula; (3) oracle inequalities satisfied by SURE-tuned estimates under a mild Lipschitz assumption; (4) an upper bound on the variance of the size of the model selected by the Lasso, and more generally an upper bound on the variance of the empirical degrees-of-freedom of convex penalized estimators; (5) explicit expressions of SURE for SURE for the Lasso and the Elastic Net; (6) in the linear model, a general semiparametric scheme to de-bias a differentiable initial estimator for the statistical inference of a low-dimensional projection of the unknown regression coefficient vector, with a characterization of the variance after debiasing; and (7) an accuracy analysis of a Gaussian Monte Carlo scheme to approximate the divergence of functions $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$.

REFERENCES

- [1] ARLOT, S. and BACH, F. R. (2009). Data-driven calibration of linear estimators with minimal penalties. *Adv. Neural Inf. Process. Syst.* 46–54.
- [2] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Springer, Cham. MR3155209 <https://doi.org/10.1007/978-3-319-00227-9>
- [3] BELLEC, P. and TSYBAKOV, A. (2017). Bounds on the prediction error of penalized least squares estimators with convex penalty. In *Modern Problems of Stochastic Analysis and Statistics. Springer Proc. Math. Stat.* **208** 315–333. Springer, Cham. MR3747672 https://doi.org/10.1007/978-3-319-65313-6_13
- [4] BELLEC, P. C. (2018). The noise barrier and the large signal bias of the lasso and other convex estimators. Available at [arXiv:1804.01230](https://arxiv.org/abs/1804.01230).
- [5] BELLEC, P. C. (2018). Optimal bounds for aggregation of affine estimators. *Ann. Statist.* **46** 30–59. MR3766945 <https://doi.org/10.1214/17-AOS1540>

MSC2020 subject classifications. Primary 62H10, 62H12, 62J07; secondary 62G15, 62F35.

Key words and phrases. Stein's formula, variance estimate, risk estimate, SURE, SURE for SURE, Lasso, Elastic Net, model selection, variance of model size, debiased estimation, regression.

- [6] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. [MR3852663](#) <https://doi.org/10.1214/17-AOS1670>
- [7] BELLEC, P. C. and YANG, D. (2019). The cost-free nature of optimally tuning Tikhonov regularizers and other ordered smoothers. Preprint. Available at [arXiv:1905.12517](#).
- [8] BELLEC, P. C. and ZHANG, C.-H. (2018). De-biasing the lasso with degrees-of-freedom adjustment. Preprint.
- [9] BELLEC, P. C. and ZHANG, C.-H. (2019). Second order Poincaré inequalities and de-biasing arbitrary convex regularizers when $p/n \rightarrow \gamma$. Preprint. Available at [arXiv:1912.11943](#).
- [10] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. [MR3207983](#) <https://doi.org/10.1093/restud/rdt044>
- [11] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. [MR3210986](#) <https://doi.org/10.1214/14-AOS1204>
- [12] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#) <https://doi.org/10.1214/08-AOS620>
- [13] BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. [MR1642391](#) <https://doi.org/10.1090/surv/062>
- [14] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. [MR3185193](#) <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [15] BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. [MR3102549](#) <https://doi.org/10.3150/12-BEJSP11>
- [16] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. [MR2807761](#) <https://doi.org/10.1007/978-3-642-20192-9>
- [17] CANDÈS, E. J., SING-LONG, C. A. and TRZASKO, J. D. (2013). Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Trans. Signal Process.* **61** 4643–4657. [MR3105401](#) <https://doi.org/10.1109/TSP.2013.2270464>
- [18] CHATTERJEE, S. (2014). A short survey of Stein’s method. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV* 1–24. Kyung Moon Sa, Seoul. [MR3727600](#)
- [19] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein’s Method. Probability and Its Applications (New York)*. Springer, Heidelberg. [MR2732624](#) <https://doi.org/10.1007/978-3-642-15007-4>
- [20] DAI, D., RIGOLLET, P., XIA, L. and ZHANG, T. (2014). Aggregation of affine estimators. *Electron. J. Stat.* **8** 302–327. [MR3192554](#) <https://doi.org/10.1214/14-EJS886>
- [21] DAI, D., RIGOLLET, P. and ZHANG, T. (2012). Deviation optimal learning using greedy Q -aggregation. *Ann. Statist.* **40** 1878–1905. [MR3015047](#) <https://doi.org/10.1214/12-AOS1025>
- [22] DALALYAN, A. S. and SALMON, J. (2012). Sharp oracle inequalities for aggregation of affine estimators. *Ann. Statist.* **40** 2327–2355. [MR3059085](#) <https://doi.org/10.1214/12-AOS1038>
- [23] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. [MR1379464](#)
- [24] DOSSAL, C., KACHOUR, M., FADILI, M. J., PEYRÉ, G. and CHESNEAU, C. (2013). The degrees of freedom of the Lasso for general design matrix. *Statist. Sinica* **23** 809–828. [MR3086657](#)
- [25] ERDOGDU, M. A. (2016). Newton–Stein method: An optimization method for GLMs via Stein’s lemma. *J. Mach. Learn. Res.* **17** Paper No. 216, 52. [MR3595150](#)
- [26] EVANS, L. C. (1999). Partial differential equations and Monge–Kantorovich mass transfer. In *Current Developments in Mathematics, 1997 (Cambridge, MA)* 65–126. Int. Press, Boston, MA. [MR1698853](#)
- [27] GIRAUD, C., HUET, S. and VERZELEN, N. (2012). High-dimensional regression with unknown variance. *Statist. Sci.* **27** 500–518. [MR3025131](#) <https://doi.org/10.1214/12-STS398>
- [28] HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations. Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. [MR3616141](#)
- [29] JANZAMIN, M., SEDGHI, H. and ANANDKUMAR, A. (2014). Score function features for discriminative learning: Matrix and tensor framework. Preprint. Available at [arXiv:1412.2863](#).
- [30] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. [MR3277152](#)
- [31] JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. [MR3851749](#) <https://doi.org/10.1214/17-AOS1630>
- [32] JOHNSTONE, I. (1988). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)* 361–379. Springer, New York. [MR0927112](#)

- [33] KNEIP, A. (1994). Ordered linear smoothers. *Ann. Statist.* **22** 835–866. [MR1292543](#) <https://doi.org/10.1214/aos/1176325498>
- [34] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. [MR1805785](#) <https://doi.org/10.1214/aos/1015957395>
- [35] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann. Statist.* **46** 611–641. [MR3782379](#) <https://doi.org/10.1214/17-AOS1562>
- [36] LEUNG, G. and BARRON, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inf. Theory* **52** 3396–3410. [MR2242356](#) <https://doi.org/10.1109/TIT.2006.878172>
- [37] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. [MR1015135](#) <https://doi.org/10.1214/aos/1176347253>
- [38] MALLOWS, C. L. (1973). Some comments on Cp. *Technometrics* **15** 661–675.
- [39] MATOUSEK, J. and VONDRAK, J. The Probabilistic Method. Lecture notes.
- [40] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#) <https://doi.org/10.1214/009053606000000281>
- [41] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. [MR3161450](#) <https://doi.org/10.1214/13-AOS1170>
- [42] NICULESCU, C. P. and PERSSON, L.-E. (2006). *Convex Functions and Their Applications: A Contemporary Approach*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC **23**. Springer, New York. [MR2178902](#) <https://doi.org/10.1007/0-387-31077-0>
- [43] RIGOLLET, P. (2012). Kullback–Leibler aggregation and misspecified generalized linear models. *Ann. Statist.* **40** 639–665. [MR2933661](#) <https://doi.org/10.1214/11-AOS961>
- [44] RIGOLLET, P. and TSYBAKOV, A. B. (2012). Sparse estimation by exponential weighting. *Statist. Sci.* **27** 558–575. [MR3025134](#) <https://doi.org/10.1214/12-STS393>
- [45] 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. [MR0402873](#)
- [46] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [47] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. [MR2999166](#) <https://doi.org/10.1093/biomet/ass043>
- [48] SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled lasso. *J. Mach. Learn. Res.* **14** 3385–3418. [MR3144466](#)
- [49] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume I: Basic Examples*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **54**. Springer, Berlin. [MR2731561](#) <https://doi.org/10.1007/978-3-642-15202-3>
- [50] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. [MR3066375](#) <https://doi.org/10.1214/13-EJS815>
- [51] TIBSHIRANI, R. J. and ROSSET, S. (2019). Excess optimism: How biased is the apparent error of an estimator tuned by SURE? *J. Amer. Statist. Assoc.* **114** 697–712. [MR3963173](#) <https://doi.org/10.1080/01621459.2018.1429276>
- [52] TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. [MR2985948](#) <https://doi.org/10.1214/12-AOS1003>
- [53] TROPP, J. A. (2006). Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory* **52** 1030–1051. [MR2238069](#) <https://doi.org/10.1109/TIT.2005.864420>
- [54] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. [MR3224285](#) <https://doi.org/10.1214/14-AOS1221>
- [55] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York. https://doi.org/10.1007/978-1-4757-2545-2_3
- [56] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. [MR2729873](#) <https://doi.org/10.1109/TIT.2009.2016018>
- [57] YANG, Z., BALASUBRAMANIAN, K., WANG, Z. and LIU, H. Learning non-Gaussian multi-index model via second-order Stein’s method.
- [58] YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. [MR2756192](#)

- [59] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#) <https://doi.org/10.1214/09-AOS729>
- [60] ZHANG, C.-H. (2011). Statistical inference for high-dimensional data. *Mathematisches Forschungsinstitut Oberwolfach: Very High Dimensional Semiparametric Models* **48** 28–31. Report.
- [61] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#) <https://doi.org/10.1214/07-AOS520>
- [62] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. [MR3153940](#) <https://doi.org/10.1111/rssb.12026>
- [63] ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. [MR3025135](#) <https://doi.org/10.1214/12-STS399>
- [64] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)

CONSISTENT NONPARAMETRIC ESTIMATION FOR HEAVY-TAILED SPARSE GRAPHS

BY CHRISTIAN BORGES^{1,*}, JENNIFER T. CHAYES^{1,†}, HENRY COHN² AND SHIRSHENDU GANGULY^{1,‡}

¹University of California, Berkeley, *borges@berkeley.edu; †jchayes@berkeley.edu; ‡sganguly@berkeley.edu

²Microsoft Research, cohn@microsoft.com

We study graphons as a nonparametric generalization of stochastic block models, and show how to obtain compactly represented estimators for sparse networks in this framework. In contrast to previous work, we relax the usual boundedness assumption for the generating graphon and instead assume only integrability, so that we can handle networks that have long tails in their degree distributions. We also relax the usual assumption that the graphon is defined on the unit interval, to allow latent position graphs based on more general spaces.

We analyze three algorithms. The first is a least squares algorithm, which gives a consistent estimator for all square-integrable graphons, with errors expressed in terms of the best possible stochastic block model approximation. Next, we analyze an algorithm based on the cut norm, which works for all integrable graphons. Finally, we show that clustering based on degrees works whenever the underlying degree distribution is atomless.

REFERENCES

- [1] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. [MR3447993 https://doi.org/10.1109/TIT.2015.2490670](https://doi.org/10.1109/TIT.2015.2490670)
- [2] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 670–688. IEEE Computer Soc., Los Alamitos, CA. [MR3473334 https://doi.org/10.1109/FOCS.2015.47](https://doi.org/10.1109/FOCS.2015.47)
- [3] ABBE, E. and SANDON, C. (2015). Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems* 28 676–684.
- [4] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- [5] 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* 26 692–700. Extended version with proofs available at [arXiv:1311.1731](https://arxiv.org/abs/1311.1731).
- [6] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937 https://doi.org/10.1016/0047-259X\(81\)90099-3](https://doi.org/10.1016/0047-259X(81)90099-3)
- [7] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. [MR3127859 https://doi.org/10.1214/13-AOS1138](https://doi.org/10.1214/13-AOS1138)
- [8] ANANDKUMAR, A., GE, R., HSU, D. and KAKADE, S. M. (2014). A tensor approach to learning mixed membership community models. *J. Mach. Learn. Res.* **15** 2239–2312. [MR3231594](https://doi.org/10.1214/14-AOS904)
- [9] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
- [10] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. [MR2906868 https://doi.org/10.1214/11-AOS904](https://doi.org/10.1214/11-AOS904)
- [11] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. [MR2337396 https://doi.org/10.1002/rsa.20168](https://doi.org/10.1002/rsa.20168)

- [12] BOLLOBÁS, B. and RIORDAN, O. (2009). Metrics for sparse graphs. In *Surveys in Combinatorics 2009. London Mathematical Society Lecture Note Series* **365** 211–287. Cambridge Univ. Press, Cambridge. [MR2588543](#)
- [13] BOPPANA, R. B. (1987). Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science* 280–285.
- [14] BORGS, C., CHAYES, J. and LOVÁSZ, L. (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.* **19** 1597–1619. [MR2594615](#) <https://doi.org/10.1007/s00039-010-0044-0>
- [15] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics. Algorithms Combin.* **26** 315–371. Springer, Berlin. [MR2249277](#) https://doi.org/10.1007/3-540-33700-8_18
- [16] BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2021). Supplement to “Consistent nonparametric estimation for heavy-tailed sparse graphs.” <https://doi.org/10.1214/20-AOS1985SUPP>
- [17] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. [MR3758733](#) <https://doi.org/10.1214/17-AOP1187>
- [18] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062. [MR3988601](#) <https://doi.org/10.1090/tran/7543>
- [19] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. [MR2455626](#) <https://doi.org/10.1016/j.aim.2008.07.008>
- [20] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)* **176** 151–219. [MR2925382](#) <https://doi.org/10.4007/annals.2012.176.1.2>
- [21] BORGS, C., CHAYES, J. T. and SMITH, A. (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems* 28 1369–1377. Extended version with proofs available at [arXiv:1506.06162](#).
- [22] CAI, D., ACKERMAN, N. and FREER, C. (2014). An iterative step-function estimator for graphons. Preprint. Available at [arXiv:1412.2129](#).
- [23] CHAN, S. H. and AIROLDI, E. M. (2014). A consistent histogram estimator for exchangeable graph models. In *Proceedings of the 31st International Conference on Machine Learning (JMLR Workshop and Conference Proceedings Volume 32)* 208–216.
- [24] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. [MR3285604](#) <https://doi.org/10.1214/14-AOS1272>
- [25] CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. [MR2857452](#) <https://doi.org/10.1214/10-AAP728>
- [26] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Proceedings of the 25th Annual Conference on Learning Theory (JMLR Workshop and Conference Proceedings Volume 23)* 35.1–35.23.
- [27] CHEN, Y., SANGHAVI, S. and XU, H. (2012). Clustering sparse graphs. In *Advances in Neural Information Processing Systems* 25 2204–2212.
- [28] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Proceedings of the 28th Conference on Learning Theory (JMLR Workshop and Conference Proceedings Volume 40)* 391–423.
- [29] CHOI, D. and WOLFE, P. J. (2014). Co-clustering separately exchangeable network data. *Ann. Statist.* **42** 29–63. [MR3161460](#) <https://doi.org/10.1214/13-AOS1173>
- [30] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. [MR2931253](#) <https://doi.org/10.1093/biomet/asr053>
- [31] COJA-OGHLAN, A. (2010). Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** 227–284. [MR2593622](#) <https://doi.org/10.1017/S0963548309990514>
- [32] DASGUPTA, A., HOPCROFT, J., KANNAN, R. and MITRA, P. (2006). Spectral clustering by recursive partitioning. In *Algorithms—ESA 2006. Lecture Notes in Computer Science* **4168** 256–267. Springer, Berlin. [MR2347148](#) https://doi.org/10.1007/11841036_25
- [33] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** 33–61. [MR2463439](#)
- [34] DYER, M. E. and FRIEZE, A. M. (1989). The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489. [MR1022107](#) [https://doi.org/10.1016/0196-6774\(89\)90001-1](https://doi.org/10.1016/0196-6774(89)90001-1)

- [35] FISHKIND, D. E., SUSSMAN, D. L., TANG, M., VOGELSTEIN, J. T. and PRIEBE, C. E. (2013). Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM J. Matrix Anal. Appl.* **34** 23–39. MR3032990 <https://doi.org/10.1137/120875600>
- [36] FRIEZE, A. and KANNAN, R. (1999). Quick approximation to matrices and applications. *Combinatorica* **19** 175–220. MR1723039 <https://doi.org/10.1007/s004930050052>
- [37] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 <https://doi.org/10.1214/15-AOS1354>
- [38] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inf. Theory* **62** 2788–2797. MR3493879 <https://doi.org/10.1109/TIT.2016.2546280>
- [39] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inf. Theory* **62** 5918–5937. MR3552431 <https://doi.org/10.1109/TIT.2016.2594812>
- [40] HOEFFDING, W. (1961). The strong law of large numbers for U -statistics. *North Carolina State University, Institute of Statistics Mimeograph Series* **302**. <https://repository.lib.ncsu.edu/dr/handle/1840.4/2128>.
- [41] 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 <https://doi.org/10.1198/016214502388618906>
- [42] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [43] HOOVER, D. (1979). Relations on probability spaces and arrays of random variables. Institute for Advanced Study, Princeton, NJ. Preprint.
- [44] JANSON, S. (2013). *Graphons, cut norm and distance, couplings and rearrangements*. NYJM Monographs **4**. State University of New York, University at Albany, Albany, NY. 76 pp. MR3043217
- [45] JERRUM, M. and SORKIN, G. B. (1998). The Metropolis algorithm for graph bisection. *Discrete Appl. Math.* **82** 155–175. MR1609981 [https://doi.org/10.1016/S0166-218X\(97\)00133-9](https://doi.org/10.1016/S0166-218X(97)00133-9)
- [46] KALLENBERG, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *J. Theoret. Probab.* **12** 859–883. MR1702867 <https://doi.org/10.1023/A:1021692202530>
- [47] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 <https://doi.org/10.1214/16-AOS1454>
- [48] LATOUCHE, P. and ROBIN, S. (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W -graph models. *Stat. Comput.* **26** 1173–1185. MR3538631 <https://doi.org/10.1007/s11222-015-9607-0>
- [49] LAURITZEN, S. L. (2003). Rasch models with exchangeable rows and columns. In *Bayesian Statistics, 7 (Tenerife, 2002)* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 215–232. Oxford Univ. Press, New York. With a discussion by Michael Goldstein and a reply by the author. MR2003175
- [50] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 <https://doi.org/10.1214/14-AOS1274>
- [51] LEI, J. and ZHU, L. (2017). Generic sample splitting for refined community recovery in degree corrected stochastic block models. *Statist. Sinica* **27** 1639–1659. MR3701502
- [52] LLOYD, J. R., ORBANZ, P., GHAHRAMANI, Z. and ROY, D. M. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems* 25 **25** 1007–1015.
- [53] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 <https://doi.org/10.1016/j.jctb.2006.05.002>
- [54] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997
- [55] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742
- [56] OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.
- [57] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic block-model. In *Advances in Neural Information Processing Systems* 26 3120–3128.
- [58] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 <https://doi.org/10.1214/11-AOS887>
- [59] RYFF, J. V. (1970). Measure preserving transformations and rearrangements. *J. Math. Anal. Appl.* **31** 449–458. MR0419734 [https://doi.org/10.1016/0022-247X\(70\)90038-7](https://doi.org/10.1016/0022-247X(70)90038-7)

- [60] SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. [MR1449742](#) <https://doi.org/10.1007/s003579900004>
- [61] TANG, M., SUSSMAN, D. L. and PRIEBE, C. E. (2013). Universally consistent vertex classification for latent positions graphs. *Ann. Statist.* **41** 1406–1430. [MR3113816](#) <https://doi.org/10.1214/13-AOS1112>
- [62] VU, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combin. Probab. Comput.* **27** 124–140. [MR3734334](#) <https://doi.org/10.1017/S0963548317000463>
- [63] WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. [MR0883333](#)
- [64] FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple socio-metric relations. *J. Amer. Statist. Assoc.* **80** 51–67.
- [65] WHITE, H. C., BOORMAN, S. A. and BREIGER, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *Am. J. Sociol.* **80** 730–780.
- [66] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at [arXiv:1309.5936](#).
- [67] YANG, J. J., HAN, Q. and AIROLDI, E. M. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics (JMLR Workshop and Conference Proceedings Volume 33)* 1060–1067.
- [68] YUN, S.-Y. and PROUTIERE, A. (2014). Accurate community detection in the stochastic block model via spectral algorithms. Preprint. Available at [arXiv:1412.7335](#).
- [69] YUN, S.-Y. and PROUTIERE, A. (2014). Community detection via random and adaptive sampling. In *Proceedings of the 27th Conference on Learning Theory (JMLR Workshop and Conference Proceedings Volume 35)* 138–175.

NONEXCHANGEABLE RANDOM PARTITION MODELS FOR MICROCLUSTERING

BY GIUSEPPE DI BENEDETTO^{*}, FRANÇOIS CARON[†] AND YEE WHYE TEH[‡]

Department of Statistics, University of Oxford, ^{*}dibenegiu@gmail.com; [†]caron@stats.ox.ac.uk; [‡]y.w.teh@stats.ox.ac.uk

Many popular random partition models, such as the Chinese restaurant process and its two-parameter extension, fall in the class of exchangeable random partitions, and have found wide applicability in various fields. While the exchangeability assumption is sensible in many cases, it implies that the size of the clusters necessarily grows linearly with the sample size, and such feature may be undesirable for some applications. We present here a flexible class of nonexchangeable random partition models, which are able to generate partitions whose cluster sizes grow sublinearly with the sample size, and where the growth rate is controlled by one parameter. Along with this result, we provide the asymptotic behaviour of the number of clusters of a given size, and show that the model can exhibit a power-law behaviour, controlled by another parameter. The construction is based on completely random measures and a Poisson embedding of the random partition, and inference is performed using a Sequential Monte Carlo algorithm. Experiments on real data sets emphasise the usefulness of the approach compared to a two-parameter Chinese restaurant process.

REFERENCES

- [1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. [MR0883646](#) <https://doi.org/10.1007/BFb0099421>
- [2] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. [MR2758115](#) <https://doi.org/10.1111/j.1467-9868.2009.00736.x>
- [3] BACALLADO, S., FAVARO, S. and TRIPPA, L. (2015). Looking-backward probabilities for Gibbs-type exchangeable random partitions. *Bernoulli* **21** 1–37. [MR3322311](#) <https://doi.org/10.3150/13-BEJ559>
- [4] BETZ, V., UELTSCHI, D. and VELENIK, Y. (2011). Random permutations with cycle weights. *Ann. Appl. Probab.* **21** 312–331. [MR2759204](#) <https://doi.org/10.1214/10-AAP697>
- [5] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. [MR0898871](#) <https://doi.org/10.1017/CBO9780511721434>
- [6] BLEI, D. M. and FRAZIER, P. I. (2011). Distance dependent Chinese restaurant processes. *J. Mach. Learn. Res.* **12** 2461–2488. [MR2834504](#)
- [7] BLOEM-REDDY, B. (2017). Preferential attachment and vertex arrival times. Available at [arXiv:1710.02159](#).
- [8] BRIX, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.* **31** 929–953. [MR1747450](#) <https://doi.org/10.1239/aap/1029955251>
- [9] BRODERICK, T., JORDAN, M. I. and PITMAN, J. (2012). Beta processes, stick-breaking and power laws. *Bayesian Anal.* **7** 439–475. [MR2934958](#) <https://doi.org/10.1214/12-BA715>
- [10] CAI, D., CAMPBELL, T. and BRODERICK, T. (2016). Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 4249–4257. Curran Associates, Red Hook.
- [11] CAMERLENGHI, F., DUNSON, D. B., LIJOI, A., PRÜNSTER, I. and RODRÍGUEZ, A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Anal.* **14** 1303–1356. [MR4044854](#) <https://doi.org/10.1214/19-BA1169>

MSC2020 subject classifications. 62F15, 60G55, 60G57.

Key words and phrases. Power-law, random partitions, completely random measure, stochastic process, sparse random graph.

- [12] CAMERLENGHI, F., LIJOI, A., ORBANZ, P. and PRÜNSTER, I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.* **47** 67–92. [MR3909927](#) <https://doi.org/10.1214/17-AOS1678>
- [13] CARON, F., DAVY, M. and DOUCET, A. (2007). Generalized Pólya urn for time-varying Dirichlet process mixtures. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence, UAI 2007* 33–40.
- [14] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. [MR3731666](#) <https://doi.org/10.1111/rssb.12233>
- [15] CARON, F., NEISWANGER, W., WOOD, F., DOUCET, A. and DAVY, M. (2017). Generalized Pólya urn for time-varying Pitman-Yor processes. *J. Mach. Learn. Res.* **18** Paper No. 27, 32. [MR3634894](#)
- [16] CARON, F. and ROUSSEAU, J. (2017). On sparsity and power-law properties of graphs based on exchangeable point processes. Preprint. Available at [arXiv:1708.03120](#).
- [17] CHEN, C., RAO, V., BUNTIME, W. and TEH, Y. W. (2013). Dependent normalized random measures. In *International Conference on Machine Learning*.
- [18] CRANE, H. and DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113** 1311–1326. [MR3862359](#) <https://doi.org/10.1080/01621459.2017.1341413>
- [19] DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. [MR2371524](#) <https://doi.org/10.1007/978-0-387-49835-5>
- [20] DEL MORAL, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications (New York)*. Springer, New York. [MR2044973](#) <https://doi.org/10.1007/978-1-4684-9393-1>
- [21] DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. [MR2278333](#) <https://doi.org/10.1111/j.1467-9868.2006.00553.x>
- [22] DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R. H., PRÜNSTER, I. and RUGGIERO, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 212–229.
- [23] DI BENEDETTO, G., CARON, F. and TEH, Y. W. (2021). Supplement to “Nonexchangeable random partition models for microclustering.” <https://doi.org/10.1214/20-AOS2003SUPP>
- [24] DOUCET, A., DE FREITAS, N. and GORDON, N., eds. (2001) *Sequential Monte Carlo Methods in Practice. Statistics for Engineering and Information Science*. Springer, New York. [MR1847783](#) <https://doi.org/10.1007/978-1-4757-3437-9>
- [25] DURRETT, R. and SCHWEINSBERG, J. (2005). Power laws for family sizes in a duplication model. *Ann. Probab.* **33** 2094–2126. [MR2184092](#) <https://doi.org/10.1214/009117905000000369>
- [26] FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *Ann. Appl. Probab.* **23** 1721–1754. [MR3114915](#) <https://doi.org/10.1214/12-AAP843>
- [27] FEINERER, I. and HORNIK, K. (2018). tm: Text Mining Package. R package version 0.7-6.
- [28] FEINERER, I., HORNIK, K. and MEYER, D. (2008). Text mining infrastructure in R. *J. Stat. Softw.* **25** 1–54.
- [29] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. [MR0270403](#)
- [30] FOTI, N. J. and WILLIAMSON, S. A. (2015). A survey of non-exchangeable priors for Bayesian nonparametric models. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 359–371.
- [31] GNEDIN, A., HANSEN, B. and PITMAN, J. (2007). Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Probab. Surv.* **4** 146–171. [MR2318403](#) <https://doi.org/10.1214/07-PS092>
- [32] GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325** 83–102, 244–245. [MR2160320](#) <https://doi.org/10.1007/s10958-006-0335-z>
- [33] GRIFFIN, J. E. and STEEL, M. F. J. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 179–194. [MR2268037](#) <https://doi.org/10.1198/016214505000000727>
- [34] JAMES, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Preprint. Available at [arXiv:math/0205093](#).
- [35] JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799. [MR2166562](#) <https://doi.org/10.1214/009053605000000336>
- [36] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. [MR2508332](#) <https://doi.org/10.1111/j.1467-9469.2008.00609.x>
- [37] JANSON, S. (2018). On edge exchangeable random graphs. *J. Stat. Phys.* **173** 448–484. [MR3876897](#) <https://doi.org/10.1007/s10955-017-1832-9>
- [38] KARLIN, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17** 373–401. [MR0216548](#) <https://doi.org/10.1512/iumj.1968.17.17020>

- [39] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. [MR0210185](#)
- [40] KINGMAN, J. F. C. (1978). Random partitions in population genetics. *Proc. R. Soc. Lond. Ser. A* **361** 1–20. [MR0526801](#) <https://doi.org/10.1098/rspa.1978.0089>
- [41] KINGMAN, J. F. C. (1993). *Poisson Processes. Oxford Studies in Probability* **3**. Oxford University Press, New York. [MR1207584](#)
- [42] KINGMAN, J. F. C. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** 1–22. [MR0368264](#)
- [43] KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1** 705–711. [MR0350950](#) <https://doi.org/10.1214/aop/1176996898>
- [44] LAU, J. W. and GREEN, P. J. (2007). Bayesian model-based clustering procedures. *J. Comput. Graph. Statist.* **16** 526–558. [MR2351079](#) <https://doi.org/10.1198/106186007X238855>
- [45] LESKOVEC, J. and KREVL, A. (2014). SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>.
- [46] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Stat. Inference Stoch. Process.* **8** 283–309. [MR2177315](#) <https://doi.org/10.1007/s11203-005-6071-z>
- [47] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291. [MR2236441](#) <https://doi.org/10.1198/016214505000000132>
- [48] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. [MR2416792](#) <https://doi.org/10.1093/biomet/asm061>
- [49] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 715–740. [MR2370077](#) <https://doi.org/10.1111/j.1467-9868.2007.00609.x>
- [50] LIJOI, A., NIPOTI, B. and PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20** 1260–1291. [MR3217444](#) <https://doi.org/10.3150/13-BEJ521>
- [51] LIJOI, A. and PRÜNSTER, I. (2003). On a normalized random measure with independent increments relevant to Bayesian nonparametric inference. In *Proceedings of the 13th European Young Statisticians Meeting* 123–134. Bernoulli Society.
- [52] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics. Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge Univ. Press, Cambridge. [MR2730661](#)
- [53] LIN, D., GRIMSON, E. and FISHER, J. W. (2010). Construction of dependent Dirichlet processes based on Poisson processes. In *Advances in Neural Information Processing Systems* 1396–1404.
- [54] MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science* 50–55. Amer. Statist. Assoc., Alexandria, VA.
- [55] MCAULEY, J., TARGETT, C., SHI, Q. and VAN DEN HENGEL, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* 43–52. ACM, New York.
- [56] MILLER, J., BETANCOURT, B., ZAIDI, A., WALLACH, H. and STEORTS, R. (2015). Microclustering: When the cluster sizes grow sublinearly with the size of the data set. [arXiv:1512.00792v1](#).
- [57] MÜLLER, P. and RODRIGUEZ, A. (2013). *Nonparametric Bayesian Inference. NSF-CBMS Regional Conference Series in Probability and Statistics* **9**. IMS, Beachwood, OH; Amer. Statist. Assoc., Alexandria, VA. [MR3113683](#)
- [58] NIETO-BARAJAS, L. E., PRÜNSTER, I. and WALKER, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.* **32** 2343–2360. [MR2153987](#) <https://doi.org/10.1214/009053604000000625>
- [59] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. [MR1156448](#) <https://doi.org/10.1007/BF01205234>
- [60] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. [MR1337249](#) <https://doi.org/10.1007/BF01213386>
- [61] PITMAN, J. (2003). Poisson-Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. [MR2004330](#) <https://doi.org/10.1214/lnms/1215091133>
- [62] PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin. [MR2245368](#)
- [63] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. [MR1434129](#) <https://doi.org/10.1214/aop/1024404422>
- [64] RAO, V. and TEH, Y. W. (2009). Spatial normalized gamma processes. In *Advances in Neural Information Processing Systems* 1554–1562.

- [65] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments *Ann. Statist.* **31** 560–585. [MR1983542](#) <https://doi.org/10.1214/aos/1051027881>
- [66] RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York. [MR2271424](#)
- [67] SUDDERTH, E. B. and JORDAN, M. I. (2009). Shared segmentation of natural scenes using dependent Pitman-Yor processes. In *Advances in Neural Information Processing Systems* 1585–1592.
- [68] TEH, Y. W. (2006). A hierarchical Bayesian language model based on Pitman-Yor processes. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics* 985–992. Association for Computational Linguistics.
- [69] ZANELLA, G., BETANCOURT, B., WALLACH, H., MILLER, J. W., ZAIDI, A. and STEORTS, R. C. (2016). Flexible models for microclustering with application to entity resolution. In *Advances in Neural Information Processing Systems* 1417–1425.
- [70] ZHOU, M., FAVARO, S. and WALKER, S. G. (2017). Frequency of frequencies distributions and size-dependent exchangeable random partitions. *J. Amer. Statist. Assoc.* **112** 1623–1635. [MR3750886](#) <https://doi.org/10.1080/01621459.2016.1222290>

PESKUN–TIERNEY ORDERING FOR MARKOVIAN MONTE CARLO: BEYOND THE REVERSIBLE SCENARIO

BY CHRISTOPHE ANDRIEU¹ AND SAMUEL LIVINGSTONE²

¹*School of Mathematics, University of Bristol, c.andrieu@bristol.ac.uk*

²*Department of Statistical Science, University College London, samuel.livingstone@ucl.ac.uk*

Historically time-reversibility of the transitions or processes underpinning Markov chain Monte Carlo methods (MCMC) has played a key role in their development, while the self-adjointness of associated operators together with the use of classical functional analysis techniques on Hilbert spaces have led to powerful and practically successful tools to characterise and compare their performance. Similar results for algorithms relying on nonreversible Markov processes are scarce. We show that for a type of nonreversible Monte Carlo Markov chains and processes, of current or renewed interest in the physics and statistical literatures, it is possible to develop comparison results which closely mirror those available in the reversible scenario. We show that these results shed light on earlier literature, proving some conjectures and strengthening some earlier results.

REFERENCES

- ANDRIEU, C. (2016). On random- and systematic-scan samplers. *Biometrika* **103** 719–726. [MR3551794](#) <https://doi.org/10.1093/biomet/asw019>
- ANDRIEU, C., FORT, G. and VIHOLA, M. (2015). Quantitative convergence rates for subgeometric Markov chains. *J. Appl. Probab.* **52** 391–404. [MR3372082](#) <https://doi.org/10.1239/jap/1437658605>
- ANDRIEU, C. and LIVINGSTONE, S. (2021). Supplement to “Peskun–Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario.” <https://doi.org/10.1214/20-AOS2008SUPP>
- ANDRIEU, C. and VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. [MR3313762](#) <https://doi.org/10.1214/14-AAP1022>
- ANDRIEU, C. and VIHOLA, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Probab.* **26** 2661–2696. [MR3563190](#) <https://doi.org/10.1214/15-AAP1158>
- ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2018). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* to appear. [arXiv:1808.08592](https://arxiv.org/abs/1808.08592).
- BACALLADO, S. (2010). A version of Peskun’s theorem for higher-order Markov chains. Master’s thesis, Stanford.
- BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825. [MR3694318](#) <https://doi.org/10.1017/apr.2017.22>
- BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. [MR3911113](#) <https://doi.org/10.1214/18-AOS1715>
- BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. [MR3655855](#) <https://doi.org/10.1214/16-AAP1217>
- BORNN, L., PILLAI, N. S., SMITH, A. and WOODARD, D. (2017). The use of a single pseudo-sample in approximate Bayesian computation. *Stat. Comput.* **27** 583–590. [MR3613586](#) <https://doi.org/10.1007/s11222-016-9640-7>
- BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. [MR3693523](#) <https://doi.org/10.1214/16-AAP1255>
- BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. [MR3832232](#) <https://doi.org/10.1080/01621459.2017.1294075>
- CAMPOS, C. M. and SANZ-SERNA, J. M. (2015). Extra chance generalized hybrid Monte Carlo. *J. Comput. Phys.* **281** 365–374. [MR3281978](#) <https://doi.org/10.1016/j.jcp.2014.09.037>

MSC2020 subject classifications. Primary 65C40, 65C05; secondary 62J10.

Key words and phrases. Markov chain Monte Carlo, Peskun ordering, piecewise deterministic Markov processes.

- CARACCILO, S., PELISSETTO, A. and SOKAL, A. D. (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.* **60** 1–53. [MR1063212](#) <https://doi.org/10.1007/BF01013668>
- DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. CRC Press, London. [MR1283589](#) <https://doi.org/10.1007/978-1-4899-4483-2>
- DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. [MR1789978](#) <https://doi.org/10.1214/aoap/1019487508>
- DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. [MR3371005](#) <https://doi.org/10.1093/biomet/asu075>
- DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. [MR3960671](#) [https://doi.org/10.1016/0370-2693\(87\)91197-x](https://doi.org/10.1016/0370-2693(87)91197-x)
- DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2018). Piecewise deterministic Markov processes and their invariant measure. arXiv e-prints [arXiv:1807.05421](#).
- ETHIER, S. N. and KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence* **282**. John Wiley & Sons. [MR0838085](#) <https://doi.org/10.1002/9780470316658>
- FAGGIONATO, A., GABRIELLI, D. and RIBEZZI CRIVELLARI, M. (2009). Non-equilibrium thermodynamics of piecewise deterministic Markov processes. *J. Stat. Phys.* **137** 259–304. [MR2559431](#) <https://doi.org/10.1007/s10955-009-9850-x>
- FANG, Y., SANZ-SERNA, J.-M. and SKEEL, R. D. (2014). Compressible generalized hybrid Monte Carlo. *J. Chem. Phys.* **140** 174108.
- FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. [MR3843382](#) <https://doi.org/10.1214/18-STS648>
- GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. [MR1404536](#) <https://doi.org/10.1214/aop/1039639370>
- GUSTAFSON, P. (1998). A guided walk Metropolis algorithm. *Stat. Comput.* **8** 357–364.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. [MR3363437](#) <https://doi.org/10.1093/biomet/57.1.97>
- HOBERT, J. P. and MARCHEV, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *Ann. Statist.* **36** 532–554. [MR2396806](#) <https://doi.org/10.1214/009053607000000569>
- HOROWITZ, A. M. (1991). A generalized guided Monte Carlo algorithm. *Phys. Lett. B* **268** 247–252.
- HUKUSHIMA, K. and SAKAI, Y. (2013). An irreversible Markov-chain Monte Carlo method with skew detailed balance conditions. *J. Phys., Conf. Ser.* **473** 012012.
- KOMOROWSKI, T., LANDIM, C. and OLLA, S. (2012). *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **345**. Springer, Heidelberg. [MR2952852](#) <https://doi.org/10.1007/978-3-642-29880-6>
- LAMB, J. S. and ROBERTS, J. A. (1998). Time-reversal symmetry in dynamical systems: A survey. *Phys. D, Nonlinear Phenom.* **112** 1–39. [MR1605826](#) [https://doi.org/10.1016/S0167-2789\(97\)00199-1](https://doi.org/10.1016/S0167-2789(97)00199-1)
- LEISEN, F. and MIRA, A. (2008). An extension of Peskun and Tierney orderings to continuous time Markov chains. *Statist. Sinica* **18** 1641–1651. [MR2469328](#)
- LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2010). *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, London. [MR2681239](#) <https://doi.org/10.1142/9781848162488>
- LIU, J. S. (1996). Peskun’s theorem and a modified discrete-state Gibbs sampler. *Biometrika* **83** 681–682. [MR1423883](#) <https://doi.org/10.1093/biomet/83.3.681>
- MAIGRET, N. (1978). Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. *Ann. Inst. Henri Poincaré B, Calc. Probab. Stat.* **14** 425–440 (1979). [MR0523221](#)
- MAIRE, F., DOUC, R. and OLSSON, J. (2014). Comparison of asymptotic variances of inhomogeneous Markov chains with application to Markov chain Monte Carlo methods. *Ann. Statist.* **42** 1483–1510. [MR3262458](#) <https://doi.org/10.1214/14-AOS1209>
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- MIRA, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16** 340–350. [MR1888449](#) <https://doi.org/10.1214/ss/1015346319>
- NEAL, R. M. (2004). Improving asymptotic variance of MCMC estimators: Non-reversible chains are better. arXiv preprint [math/0407281](#).
- O’DONNELL, R. (2014). *Analysis of Boolean Functions*. Cambridge Univ. Press, New York. [MR3443800](#) <https://doi.org/10.1017/CBO9781139814782>
- OTTOBRE, M. (2016). Markov chain Monte Carlo and irreversibility. *Rep. Math. Phys.* **77** 267–292. [MR3517569](#) [https://doi.org/10.1016/S0034-4877\(16\)30031-3](https://doi.org/10.1016/S0034-4877(16)30031-3)

- OTTOBRE, M., PILLAI, N. S., PINSKI, F. J. and STUART, A. M. (2016). A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli* **22** 60–106. [MR3449777](#) <https://doi.org/10.3150/14-BEJ621>
- PEDERSEN, G. K. (2012). *Analysis Now. Graduate Texts in Mathematics* **118**. Springer, New York. [MR0971256](#) <https://doi.org/10.1007/978-1-4612-1007-8>
- PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. [MR0362823](#) <https://doi.org/10.1093/biomet/60.3.607>
- PETERS, E. A. J. F. and DE WIT, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* (3) **85** 026703.
- PONCET, R. (2017). Generalized and hybrid MCMC overdamped Langevin algorithms. arXiv Preprint [arXiv:1701.05833](#).
- REY-BELLET, L. and SPILIOPOULOS, K. (2016). Improving the convergence of reversible samplers. *J. Stat. Phys.* **164** 472–494. [MR3519206](#) <https://doi.org/10.1007/s10955-016-1565-1>
- ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Minimising MCMC variance via diffusion limits, with an application to simulated tempering. *Ann. Appl. Probab.* **24** 131–149. [MR3161644](#) <https://doi.org/10.1214/12-AAP918>
- ROBERTS, G. O. and ROSENTHAL, J. S. (2015). Surprising convergence properties of some simple Gibbs samplers under various scans. *International Journal of Statistics and Probability* **5** 51.
- ROSENTHAL, J. S. and ROSENTHAL, P. (2015). Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.* **20** no. 91, 10. [MR3434208](#) <https://doi.org/10.1214/ECP.v20-4528>
- SAKAI, Y. and HUKUSHIMA, K. (2016). Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions. *Phys. Rev. E* (3) **93** 043318. <https://doi.org/10.1103/PhysRevE.93.043318>
- SHERLOCK, C., THIERY, A. H. and LEE, A. (2017). Pseudo-marginal Metropolis–Hastings sampling using averages of unbiased estimators. *Biometrika* **104** 727–734. [MR3694593](#) <https://doi.org/10.1093/biomet/asx031>
- SOHL-DICKSTEIN, J., MUDIGONDA, M. and DEWEESE, M. R. (2014). Hamiltonian Monte Carlo without detailed balance. arXiv Preprint [arXiv:1409.5191](#).
- TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. [MR1620401](#) <https://doi.org/10.1214/aoap/1027961031>
- TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D, Nonlinear Phenom.* **240** 410–414.
- VOIGT, J. (1977). On the perturbation theory for strongly continuous semigroups. *Math. Ann.* **229** 163–171. [MR0445335](#) <https://doi.org/10.1007/BF01351602>
- VUCELJA, M. (2016). Lifting: A nonreversible Markov chain Monte Carlo algorithm. *Am. J. Phys.* **84** 958–968.
- YAGLOM, A. M. (1949). On the statistical reversibility of Brownian motion. *Mat. Sb. (N.S.)* **24(66)** 457–492. [MR0030718](#)

VOLATILITY COUPLING

BY JEAN JACOD¹, JIA LI² AND ZHIPENG LIAO³

¹*Institut de Mathématiques de Jussieu, Sorbonne Université, jean.jacod@gmail.com*

²*Department of Economics, Duke University, jl410@duke.edu*

³*Department of Economics, UCLA, zhipeng.liao@econ.ucla.edu*

This paper provides a strong approximation, or coupling, theory for spot volatility estimators formed using high-frequency data. We show that the t-statistic process associated with the nonparametric spot volatility estimator can be strongly approximated by a growing-dimensional vector of independent variables defined as functions of Brownian increments. We use this coupling theory to study the uniform inference for the volatility process in an infill asymptotic setting. Specifically, we propose uniform confidence bands for spot volatility, beta, idiosyncratic variance processes, and their nonlinear transforms. The theory is also applied to address an open question concerning the inference of monotone nonsmooth integrated volatility functionals such as the occupation time and its quantiles.

REFERENCES

- AIT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton.
- ANDERSEN, T. G. and BOLLERSLEV, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Internat. Econom. Rev.* **39** 885–905.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625. [MR1958138](#) <https://doi.org/10.1111/1468-0262.00418>
- BARNDOFF-NIELSEN, O. E. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925. [MR2051439](#) <https://doi.org/10.1111/j.1468-0262.2004.00515.x>
- BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. [MR3343791](#) <https://doi.org/10.1016/j.jeconom.2015.02.014>
- BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. [MR0348906](#)
- BOLLERSLEV, T. and TODOROV, V. (2011). Estimation of jump tails. *Econometrica* **79** 1727–1783. [MR2895885](#) <https://doi.org/10.3982/ECTA9240>
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#) <https://doi.org/10.1214/13-AOS1161>
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. [MR3262461](#) <https://doi.org/10.1214/14-AOS1230>
- CHERNOZHUKOV, V., LEE, S. and ROSEN, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica* **81** 667–737. [MR3043345](#) <https://doi.org/10.3982/ECTA8718>
- COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8** 291–323. [MR1645101](#) <https://doi.org/10.1111/1467-9965.00057>
- FAN, J. and WANG, Y. (2008). Spot volatility estimation for high-frequency data. *Stat. Interface* **1** 279–288. [MR2476744](#) <https://doi.org/10.4310/SII.2008.v1.n2.a5>
- FOSTER, D. P. and NELSON, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64** 139–174. [MR1366144](#) <https://doi.org/10.2307/2171927>
- GEMAN, D. and HOROWITZ, J. (1980). Occupation densities. *Ann. Probab.* **8** 1–67. [MR0556414](#)

MSC2020 subject classifications. 60F15, 60G44, 62G20.

Key words and phrases. Coupling, high-frequency data, occupation measure, quantiles, semimartingale, uniform inference.

- JACOD, J. and PROTTER, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096 <https://doi.org/10.1007/978-3-642-24127-7>
- JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. MR3113818 <https://doi.org/10.1214/13-AOS1115>
- JOHNSTON, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12** 402–414. MR0666014 [https://doi.org/10.1016/0047-259X\(82\)90074-4](https://doi.org/10.1016/0047-259X(82)90074-4)
- KRISTENSEN, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* **26** 60–93. MR2587103 <https://doi.org/10.1017/S0266466609090616>
- LI, J. and LIAO, Z. (2020). Uniform nonparametric inference for time series. *J. Econometrics* **219** 38–51. MR4152784 <https://doi.org/10.1016/j.jeconom.2019.09.011>
- LI, J., TODOROV, V. and TAUCHEN, G. (2013). Volatility occupation times. *Ann. Statist.* **41** 1865–1891. MR3127851 <https://doi.org/10.1214/13-AOS1135>
- LI, J., TODOROV, V. and TAUCHEN, G. (2017a). Adaptive estimation of continuous-time regression models using high-frequency data. *J. Econometrics* **200** 36–47. MR3683660 <https://doi.org/10.1016/j.jeconom.2017.01.010>
- LI, J., TODOROV, V. and TAUCHEN, G. (2017b). Jump regressions. *Econometrica* **85** 173–195. MR3611769 <https://doi.org/10.3982/ECTA12962>
- LI, J. and XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* **84** 1613–1633. MR3537165 <https://doi.org/10.3982/ECTA12306>
- MANCINI, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *G. Ist. Ital. Attuari* **LXIV** 19–47.
- MARCUS, M. B. and ROSEN, J. (2006). *Markov Processes, Gaussian Processes, and Local Times. Cambridge Studies in Advanced Mathematics* **100**. Cambridge Univ. Press, Cambridge. MR2250510 <https://doi.org/10.1017/CBO9780511617997>
- MYKLAND, P. A. and ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77** 1403–1445. MR2561071 <https://doi.org/10.3982/ECTA7417>
- RENAULT, E., SARISOY, C. and WERKER, B. J. M. (2017). Efficient estimation of integrated volatility and related processes. *Econometric Theory* **33** 439–478. MR3600049 <https://doi.org/10.1017/S0266466616000013>
- TODOROV, V. and TAUCHEN, G. (2012). The realized Laplace transform of volatility. *Econometrica* **80** 1105–1127. MR2963883 <https://doi.org/10.3982/ECTA9133>
- ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 <https://doi.org/10.1214/16-AOS1512>

ASYMPTOTIC DISTRIBUTIONS OF HIGH-DIMENSIONAL DISTANCE CORRELATION INFERENCE

BY LAN GAO^{1,*}, YINGYING FAN^{1,†}, JINCHI LV^{1,‡} AND QI-MAN SHAO^{2,3}

¹*Data Sciences and Operations Department, Marshall School of Business, University of Southern California,*

**gaolan@marshall.usc.edu; †fanyingy@marshall.usc.edu; ‡jinchilv@marshall.usc.edu*

²*Department of Statistics and Data Science, Southern University of Science and Technology, shaohqm@sustech.edu.cn*

³*Department of Statistics, The Chinese University of Hong Kong*

Distance correlation has become an increasingly popular tool for detecting the nonlinear dependence between a pair of potentially high-dimensional random vectors. Most existing works have explored its asymptotic distributions under the null hypothesis of independence between the two random vectors when only the sample size or the dimensionality diverges. Yet its asymptotic null distribution for the more realistic setting when both sample size and dimensionality diverge in the full range remains largely underdeveloped. In this paper, we fill such a gap and develop central limit theorems and associated rates of convergence for a rescaled test statistic based on the bias-corrected distance correlation in high dimensions under some mild regularity conditions and the null hypothesis. Our new theoretical results reveal an interesting phenomenon of blessing of dimensionality for high-dimensional distance correlation inference in the sense that the accuracy of normal approximation can increase with dimensionality. Moreover, we provide a general theory on the power analysis under the alternative hypothesis of dependence, and further justify the capability of the rescaled distance correlation in capturing the pure nonlinear dependency under moderately high dimensionality for a certain type of alternative hypothesis. The theoretical results and finite-sample performance of the rescaled statistic are illustrated with several simulation examples and a blockchain application.

REFERENCES

- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall’s tau. *Bernoulli* **20** 1006–1028. [MR3178526](#) <https://doi.org/10.3150/13-BEJ514>
- BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. [MR4060981](#)
- BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. [MR0125690](#) <https://doi.org/10.1214/aoms/1177705055>
- CHAKRABORTY, S. and ZHANG, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *J. Amer. Statist. Assoc.* **114** 1638–1650. [MR4047289](#) <https://doi.org/10.1080/01621459.2018.1513364>
- DAVIS, R. A., MATSUI, M., MIKOSCH, T. and WAN, P. (2018). Applications of distance correlation to time series. *Bernoulli* **24** 3087–3116. [MR3779711](#) <https://doi.org/10.3150/17-BEJ955>
- ESCOUFIER, Y. (1973). Le traitement des variables vectorielles. *Biometrics* **29** 751–760. [MR0334416](#) <https://doi.org/10.2307/2529140>
- FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433.
- GAO, L., FAN, Y., LV, J. and SHAO, Q.-M. (2021). Supplement to “Asymptotic distributions of high-dimensional distance correlation inference.” <https://doi.org/10.1214/20-AOS2024SUPP>

MSC2020 subject classifications. Primary 62E20, 62H20; secondary 62G10, 62G20.

Key words and phrases. Nonparametric inference, high dimensionality, distance correlation, test of independence, nonlinear dependence detection, central limit theorem, rate of convergence, power, blockchain.

- GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. [MR2249882](#)
- HAEUSLER, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *Ann. Probab.* **16** 275–299. [MR0920271](#)
- HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. [MR0029139](#) <https://doi.org/10.1214/aoms/1177730150>
- HUANG, C. and HUO, X. (2017). A statistically and numerically efficient independence test based on random projections and distance covariance. arXiv preprint. Available at [arXiv:1701.06054](#).
- HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. [MR3556612](#) <https://doi.org/10.1080/00401706.2015.1054435>
- JIN, Z. and MATTESON, D. S. (2018). Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. *J. Multivariate Anal.* **168** 304–322. [MR3858367](#) <https://doi.org/10.1016/j.jmva.2018.08.006>
- KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.
- KONG, J., WANG, S. and WAHBA, G. (2015). Using distance covariance for improved variable selection with application to learning genetic risk models. *Stat. Med.* **34** 1708–1720. [MR3334686](#) <https://doi.org/10.1002/sim.6441>
- KONG, Y., LI, D., FAN, Y. and LV, J. (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *Ann. Statist.* **45** 897–922. [MR3650404](#) <https://doi.org/10.1214/16-AOS1474>
- LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. [MR3010900](#) <https://doi.org/10.1080/01621459.2012.695654>
- LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. [MR3127883](#) <https://doi.org/10.1214/12-AOP803>
- MATSUI, M., MIKOSCH, T. and SAMORODNITSKY, G. (2017). Distance covariance for stochastic processes. *Probab. Math. Statist.* **37** 355–372. [MR3745391](#) <https://doi.org/10.19195/0208-4147.37.2.9>
- MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* **112** 623–637. [MR3671757](#) <https://doi.org/10.1080/01621459.2016.1150851>
- PEARSON, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* **58** 240–242.
- ROBERT, P. and ESCOUFIER, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **25** 257–265. [MR0440801](#) <https://doi.org/10.2307/2347233>
- ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14. [MR0428579](#)
- SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. [MR4124333](#) <https://doi.org/10.1214/19-AOS1857>
- SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. [MR3265698](#) <https://doi.org/10.1080/01621459.2014.887012>
- SPEARMAN, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* **15** 72–101.
- STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. [MR1924302](#) <https://doi.org/10.1111/1467-9868.00346>
- SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. [MR2752127](#) <https://doi.org/10.1214/09-AOAS312>
- SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t -test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. [MR3053543](#) <https://doi.org/10.1016/j.jmva.2013.02.012>
- SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. [MR3269983](#) <https://doi.org/10.1214/14-AOS1255>
- SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. [MR2382665](#) <https://doi.org/10.1214/009053607000000505>
- VEPAKOMMA, P., TONDE, C. and ELGAMMAL, A. (2018). Supervised dimensionality reduction via distance correlation maximization. *Electron. J. Stat.* **12** 960–984. [MR3772810](#) <https://doi.org/10.1214/18-EJS1403>
- WANG, X., PAN, W., HU, W., TIAN, Y. and ZHANG, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734. [MR3449068](#) <https://doi.org/10.1080/01621459.2014.993081>
- WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. [MR3842884](#) <https://doi.org/10.1093/biomet/as021>
- YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. [MR3798874](#) <https://doi.org/10.1111/rssb.12259>

- ZHANG, X., YAO, S. and SHAO, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Ann. Statist.* **46** 219–246. [MR3766951](#) <https://doi.org/10.1214/17-AOS1548>
- ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. [MR2915095](#) <https://doi.org/10.1111/j.1467-9892.2011.00780.x>
- ZHU, C., YAO, S., ZHANG, X. and SHAO, X. (2020). Distance-based and RKHS-based dependence metrics in high dimension. *Ann. Statist.* To appear.

CONFIDENCE INTERVALS FOR MULTIPLE ISOTONIC REGRESSION AND OTHER MONOTONE MODELS

BY HANG DENG^{*}, QIYANG HAN[†] AND CUN-HUI ZHANG[‡]

Department of Statistics, Rutgers University, ^{*}hdeng@stat.rutgers.edu; [†]qh85@stat.rutgers.edu; [‡]czhang@stat.rutgers.edu

We consider the problem of constructing pointwise confidence intervals in the multiple isotonic regression model. Recently, Han and Zhang (2020) obtained a pointwise limit distribution theory for the so-called block max–min and min–max estimators (Fokianos, Leucht and Neumann (2020); Deng and Zhang (2020)) in this model, but inference remains a difficult problem due to the nuisance parameter in the limit distribution that involves multiple unknown partial derivatives of the true regression function.

In this paper, we show that this difficult nuisance parameter can be effectively eliminated by taking advantage of information beyond point estimates in the block max–min and min–max estimators. Formally, let $\hat{u}(x_0)$ (resp. $\hat{v}(x_0)$) be the maximizing lower-left (resp. minimizing upper-right) vertex in the block max–min (resp. min–max) estimator, and $\bar{f}_n$ be the average of the block max–min and min–max estimators. If all (first-order) partial derivatives of f_0 are nonvanishing at x_0 , then the following pivotal limit distribution theory holds:

$$\sqrt{n_{\hat{u}, \hat{v}}(x_0)}(\hat{f}_n(x_0) - f_0(x_0)) \rightsquigarrow \sigma \cdot \mathbb{L}_{\mathbf{1}_d}.$$

Here $n_{\hat{u}, \hat{v}}(x_0)$ is the number of design points in the block $[\hat{u}(x_0), \hat{v}(x_0)]$, σ is the standard deviation of the errors, and $\mathbb{L}_{\mathbf{1}_d}$ is a universal limit distribution free of nuisance parameters. This immediately yields confidence intervals for $f_0(x_0)$ with asymptotically exact confidence level and oracle length. Notably, the construction of the confidence intervals, even new in the univariate setting, requires no more efforts than performing an isotonic regression once using the block max–min and min–max estimators, and can be easily adapted to other common monotone models including, for example, (i) monotone density estimation, (ii) interval censoring model with current status data, (iii) counting process model with panel count data, and (iv) generalized linear models. Extensive simulations are carried out to support our theory.

REFERENCES

- [1] BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. [MR2509075](#) <https://doi.org/10.1214/08-AOS609>
- [2] BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k -monotone density: Limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. [MR2382657](#) <https://doi.org/10.1214/009053607000000262>
- [3] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. [MR2341693](#) <https://doi.org/10.1214/009053606000001578>
- [4] BANERJEE, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statist. Sinica* **18** 467–492. [MR2411614](#)
- [5] BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. [MR2336859](#) <https://doi.org/10.1214/009053606000001415>
- [6] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. [MR1891743](#) <https://doi.org/10.1214/aos/1015345959>

MSC2020 subject classifications. Primary 60E15; secondary 62G05.

Key words and phrases. Limit distribution theory, confidence interval, multiple isotonic regression, Gaussian process, shape constraints.

- [7] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, London-Sydney. Wiley Series in Probability and Mathematical Statistics. [MR0326887](#)
- [8] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. [MR3782383](#) <https://doi.org/10.1214/17-AOS1566>
- [9] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* 177–197. Cambridge Univ. Press, London. [MR0277070](#)
- [10] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. [MR3357878](#) <https://doi.org/10.1214/15-AOS1324>
- [11] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. [MR3706788](#) <https://doi.org/10.3150/16-BEJ865>
- [12] CHATTERJEE, S. and LAFFERTY, J. (2019). Adaptive risk bounds in unimodal regression. *Bernoulli* **25** 1–25. [MR3892309](#) <https://doi.org/10.3150/16-bej922>
- [13] CHEN, Y. and SAMWORTH, R. J. (2016). Generalized additive and index models with shape constraints. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 729–754. [MR3534348](#) <https://doi.org/10.1111/rssb.12137>
- [14] DENG, H., HAN, Q. and ZHANG, C.-H. (2021). Supplement to “Confidence intervals for multiple isotonic regression and other monotone models.” <https://doi.org/10.1214/20-AOS2025SUPP>
- [15] DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann. Statist.* **48** 3672–3698. [MR4185824](#) <https://doi.org/10.1214/20-AOS1947>
- [16] DOSS, C. R. (2019). Concave regression: Value-constrained estimation and likelihood ratio-based inference. *Math. Program.* **174** 5–39. [MR3935071](#) <https://doi.org/10.1007/s10107-018-1338-5>
- [17] DOSS, C. R. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s -concave densities. *Ann. Statist.* **44** 954–981. [MR3485950](#) <https://doi.org/10.1214/15-AOS1394>
- [18] DOSS, C. R. and WELLNER, J. A. (2019). Inference for the mode of a log-concave density. *Ann. Statist.* **47** 2950–2976. [MR3988778](#) <https://doi.org/10.1214/18-AOS1770>
- [19] DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. [MR2546798](#) <https://doi.org/10.3150/08-BEJ141>
- [20] DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. [MR2816336](#) <https://doi.org/10.1214/10-AOS853>
- [21] FANG, B., GUNTUBOYINA, A. and SEN, B. (2019). Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy–Krause variation. Preprint. Available at [arXiv:1903.01395](https://arxiv.org/abs/1903.01395).
- [22] FENG, O. Y., GUNTUBOYINA, A., KIM, A. K. H. and SAMWORTH, R. J. (2018). Adaptation in multivariate log-concave density estimation. *Ann. Statist.* To appear. Available at [arXiv:1812.11634](https://arxiv.org/abs/1812.11634).
- [23] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2020). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Trans. Inf. Theory* **66** 6389–6402. [MR4173546](#)
- [24] GHOSAL, P. and SEN, B. (2017). On univariate convex regression. *Sankhya A* **79** 215–253. [MR3707421](#) <https://doi.org/10.1007/s13171-017-0104-8>
- [25] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. [MR0093415](#) <https://doi.org/10.1080/03461238.1956.10414944>
- [26] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*. Wadsworth Statist./Probab. Ser. 539–555. Wadsworth, Belmont, CA. [MR0822052](#)
- [27] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. [MR0981568](#) <https://doi.org/10.1007/BF00343738>
- [28] GROENEBOOM, P. (2015). Rcpp scripts. Available at https://github.com/pietg/book/tree/master/Rcpp_scripts/.
- [29] GROENEBOOM, P. and JONGBLOED, G. (1995). Isotonic estimation and rates of convergence in Wicksell’s problem. *Ann. Statist.* **23** 1518–1542. [MR1370294](#) <https://doi.org/10.1214/aos/1176324310>
- [30] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. [MR3445293](#) <https://doi.org/10.1017/CBO9781139020893>
- [31] GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. [MR3375875](#) <https://doi.org/10.1214/15-AOS1335>

- [32] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). A canonical process for estimation of convex functions: The “invelope” of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652. [MR1891741](#) <https://doi.org/10.1214/aos/1015345957>
- [33] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. [MR1891742](#) <https://doi.org/10.1214/aos/1015345958>
- [34] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. [MR1180321](#) <https://doi.org/10.1007/978-3-0348-8621-5>
- [35] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. [MR3405621](#) <https://doi.org/10.1007/s00440-014-0595-3>
- [36] HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1991). On estimation of noise variance in two-dimensional signal processing. *Adv. in Appl. Probab.* **23** 476–495.
- [37] HAN, Q. (2019). Global empirical risk minimizers with “shape constraints” are rate optimal in general dimensions. Preprint. Available at [arXiv:1905.12823](#).
- [38] HAN, Q. and KATO, K. (2019). Berry–Esseen bounds for Chernoff-type non-standard asymptotics in isotonic regression. Preprint. Available at [arXiv:1910.09662](#).
- [39] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. [MR3988762](#) <https://doi.org/10.1214/18-AOS1753>
- [40] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s -concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359. [MR3485962](#) <https://doi.org/10.1214/15-AOS1408>
- [41] HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Preprint. Available at [arXiv:1601.06844](#).
- [42] HAN, Q. and ZHANG, C.-H. (2020). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist.* **48** 3251–3282. [MR4185808](#) <https://doi.org/10.1214/19-AOS1928>
- [43] HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050. [MR0426273](#)
- [44] HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. [MR0065093](#)
- [45] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli* **15** 1010–1035. [MR2597581](#) <https://doi.org/10.3150/09-BEJ202>
- [46] KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.* **46** 2279–2306. [MR3845018](#) <https://doi.org/10.1214/17-AOS1619>
- [47] KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. [MR3576560](#) <https://doi.org/10.1214/16-AOS1480>
- [48] KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. [MR2722462](#) <https://doi.org/10.1214/10-AOS814>
- [49] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat. (IMS) Collect.* **1** 282–292. IMS, Beachwood, OH. [MR2462212](#) <https://doi.org/10.1214/193940307000000202>
- [50] KUOSMANEN, T. (2008). Representation theorem for convex nonparametric least squares. *Econom. J.* **11** 308–325.
- [51] LIM, E. and GLYNN, P. W. (2012). Consistency of multidimensional convex regression. *Oper. Res.* **60** 196–208. [MR2911667](#) <https://doi.org/10.1287/opre.1110.1007>
- [52] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. [MR1105842](#) <https://doi.org/10.1214/aos/1176348118>
- [53] MAMMEN, E., MARRON, J. S., TURLACH, B. A. and WAND, M. P. (2001). A general projection framework for constrained smoothing. *Statist. Sci.* **16** 232–248. [MR1874153](#) <https://doi.org/10.1214/ss/1009213727>
- [54] MUNK, A., BISSANTZ, N., WAGNER, T. and FREITAG, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 19–41. [MR2136637](#) <https://doi.org/10.1111/j.1467-9868.2005.00486.x>
- [55] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. [MR0267677](#)
- [56] PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Stat.* **41** 507–519. [MR0260133](#) <https://doi.org/10.1214/aoms/1177697091>
- [57] RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230. [MR0760684](#) <https://doi.org/10.1214/aos/1176346788>
- [58] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, Chichester. [MR0961262](#)

- [59] SEIJO, E. and SEN, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39** 1580–1607. [MR2850213](#) <https://doi.org/10.1214/11-AOS874>
- [60] SEIJO, E. and SEN, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* **39** 1633–1657. [MR2850215](#) <https://doi.org/10.1214/10-AOS852>
- [61] SEN, B. and BANERJEE, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika* **94** 71–86. [MR2307901](#) <https://doi.org/10.1093/biomet/asm011>
- [62] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880](#) <https://doi.org/10.1214/09-AOS777>
- [63] SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. [MR2766867](#) <https://doi.org/10.1214/10-AOS840>
- [64] SUN, J. and KALBFLEISCH, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5** 279–289. [MR1329298](#)
- [65] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. [MR1385671](#) <https://doi.org/10.1007/978-1-4757-2545-2>
- [66] WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. [MR1792787](#) <https://doi.org/10.1214/aos/1015951998>
- [67] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. [MR0606630](#)
- [68] XU, M. and SAMWORTH, R. J. (2021). High-dimensional nonparametric density estimation via symmetry and shape constraints. *Ann. Statist.* To appear. Available at [arXiv:1903.06092](#).
- [69] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. [MR1902898](#) <https://doi.org/10.1214/aos/1021379864>

ON EXTENDED ADMISSIBLE PROCEDURES AND THEIR NONSTANDARD BAYES RISK

BY HAOSUI DUANMU AND DANIEL M. ROY

Department of Statistical Sciences, University of Toronto

For finite parameter spaces, among decision procedures with finite risk functions, a decision procedure is extended admissible if and only if it is Bayes. Various relaxations of this classical equivalence have been established for infinite parameter spaces, but these extensions are each subject to technical conditions that limit their applicability, especially to modern (semi and nonparametric) statistical problems. Using results in mathematical logic and nonstandard analysis, we extend this equivalence to *arbitrary* statistical decision problems: informally, we show that, among decision procedures with finite risk functions, a decision procedure is extended admissible if and only if it has infinitesimal excess Bayes risk. In contrast to existing results, our equivalence holds in complete generality, that is, *without* regularity conditions or restrictions on the model or loss function. We also derive a nonstandard analogue of Blyth's method that yields sufficient conditions for admissibility, and apply the nonstandard theory to derive a purely standard theorem: when risk functions are continuous on a compact Hausdorff parameter space, a procedure is extended admissible if and only if it is Bayes.

REFERENCES

- [1] ANDERSON, R. M. (1982). Star-finite representations of measure spaces. *Trans. Amer. Math. Soc.* **271** 667–687. [MR0654856](#) <https://doi.org/10.2307/1998904>
- [2] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR0804611](#) <https://doi.org/10.1007/978-1-4757-4286-2>
- [3] BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801. [MR0478426](#)
- [4] BHASKARA RAO, K. P. S. and BHASKARA RAO, M. (1983). *Theory of Charges: A Study of Finitely Additive Measures. Pure and Applied Mathematics* **109**. Academic Press [Harcourt Brace Jovanovich, Publishers], New York. [MR0751777](#)
- [5] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York. [MR0070134](#)
- [6] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. [MR0286209](#) <https://doi.org/10.1214/aoms/1177693318>
- [7] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. [MR0882001](#)
- [8] COHEN, M. P. and KUO, L. (1985). The admissibility of the empirical distribution function. *Ann. Statist.* **13** 262–271. [MR0773166](#) <https://doi.org/10.1214/aos/1176346591>
- [9] CUTLAND, N. J. (1997). Nonstandard real analysis. In *Nonstandard Analysis* (L. O. Arkeryd, N. J. Cutland and C. W. Henson, eds.). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **493** 51–76. Kluwer Academic, Dordrecht. [MR1603229](#) https://doi.org/10.1007/978-94-011-5544-1_2
- [10] CUTLAND, N. J., NEVES, V., OLIVEIRA, F. and SOUSA-PINTO, J. (1995). Loeb measure theory. In *Developments in Nonstandard Mathematics* (N. J. Cutland, V. Neves, F. Oliveira and J. Sousa-Pinto, eds.). *Pitman Research Notes in Mathematics Series* **336** 151–176. Longman, Harlow. [MR1394215](#)
- [11] DUANMU, H. (2017). Applications of nonstandard analysis to Markov processes and statistical decision theory. Ph.D. thesis, Univ. Toronto. Available at <http://hdl.handle.net/1807/89711>. [MR3858218](#)
- [12] DUANMU, H. and ROY, D. M. (2021). Supplement to “On extended admissible procedures and their non-standard Bayes risk.” <https://doi.org/10.1214/20-AOS2026SUPP>

- [13] DUANMU, H. and WEISS, W. (2018). Finitely-additive, countably-additive and internal probability measures. *Comment. Math. Univ. Carolin.* **59** 467–485. MR3914713 <https://doi.org/10.14712/1213-7243.2015.270>
- [14] FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach. Probability and Mathematical Statistics 1*. Academic Press, New York. MR0215390
- [15] FREMLIN, D. H. (2009). Real-valued-measurable cardinals. Version 19.9.09. Available at <https://www1.essex.ac.uk/maths/people/fremlin/rvmc.pdf>. Accessed 2021-07-01.
- [16] HALPERN, J. Y. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games Econom. Behav.* **68** 155–179. MR2577384 <https://doi.org/10.1016/j.geb.2009.03.013>
- [17] HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345. MR0464450
- [18] HSUAN, F. C. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860–868. MR0532249
- [19] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. 1* 361–379. Univ. California Press, Berkeley, CA. MR0133191
- [20] KADANE, J. B., SCHERVISH, M. J. and SEIDENFELD, T. (1999). Statistical implications of finitely additive probability. In *Rethinking the Foundations of Statistics* 211.
- [21] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Stat.* **28** 573–601. MR0092325 <https://doi.org/10.1214/aoms/1177706874>
- [22] KUDŌ, H. (1955). On minimax invariant estimates of the transformation parameter. *Natur. Sci. Rep. Ochanomizu Univ.* **6** 31–73. MR0077842
- [23] LECAM, L. (1955). An extension of Wald’s theory of statistical decision functions. *Ann. Math. Stat.* **26** 69–81. MR0067443 <https://doi.org/10.1214/aoms/1177728594>
- [24] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875
- [25] RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory. Studies in Managerial Economics*. Division of Research, Graduate School of Business Administration, Harvard Univ., Boston, MA. MR0117844
- [26] SACKS, J. (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Stat.* **34** 751–768. MR0150908 <https://doi.org/10.1214/aoms/1177704001>
- [27] STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Stat.* **26** 518–522. MR0070929 <https://doi.org/10.1214/aoms/1177728497>
- [28] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, Vol. I 197–206. Univ. California Press, Berkeley and Los Angeles. MR0084922
- [29] STONE, M. (1967). Generalized Bayes decision functions, admissibility and the exponential family. *Ann. Math. Stat.* **38** 818–822. MR0226780 <https://doi.org/10.1214/aoms/1177698876>
- [30] WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Stat.* **10** 299–326. MR0000932 <https://doi.org/10.1214/aoms/1177732144>
- [31] WALD, A. (1947). An essentially complete class of admissible decision functions. *Ann. Math. Stat.* **18** 549–555. MR0023499 <https://doi.org/10.1214/aoms/1177730345>
- [32] WALD, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica* **15** 279–313. MR0024113 <https://doi.org/10.2307/1905331>
- [33] WALD, A. (1949). Statistical decision functions. *Ann. Math. Stat.* **20** 165–205. MR0044802 <https://doi.org/10.1214/aoms/1177730030>
- [34] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York. MR0036976

DEBIASED INVERSE-VARIANCE WEIGHTED ESTIMATOR IN TWO-SAMPLE SUMMARY-DATA MENDELIAN RANDOMIZATION

BY TING YE¹, JUN SHAO^{2,3,*} AND HYUNSEUNG KANG^{3,†}

¹Department of Statistics, University of Pennsylvania, tingye@wharton.upenn.edu

²School of Statistics, East China Normal University, *shao@stat.wisc.edu

³Department of Statistics, University of Wisconsin-Madison, †hyunseung@stat.wisc.edu

Mendelian randomization (MR) has become a popular approach to study the effect of a modifiable exposure on an outcome by using genetic variants as instrumental variables. A challenge in MR is that each genetic variant explains a relatively small proportion of variance in the exposure and there are many such variants, a setting known as many weak instruments. To this end, we provide a theoretical characterization of the statistical properties of two popular estimators in MR: the inverse-variance weighted (IVW) estimator and the IVW estimator with screened instruments using an independent selection dataset, under many weak instruments. We then propose a debiased IVW estimator, a simple modification of the IVW estimator, that is robust to many weak instruments and does not require screening. Additionally, we present two instrument selection methods to improve the efficiency of the new estimator when a selection dataset is available. An extension of the debiased IVW estimator to handle balanced horizontal pleiotropy is also discussed. We conclude by demonstrating our results in simulated and real datasets.

REFERENCES

- [1] ABBOTT, L., BRYANT, S., CHURCHHOUSE, C. et al. (2018). Round 2 GWAS results of thousands of phenotypes in the UK BioBank. Available at <http://www.nealelab.is/uk-biobank> (14 November 2018, date last accessed).
- [2] AKIYAMA, M., OKADA, Y., KANAI, M., TAKAHASHI, A., MOMOZAWA, Y., IKEDA, M., IWATA, N., IKEGAWA, S., HIRATA, M. et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49** 1458–1467.
- [3] ANDREWS, D. W. K. and STOCK, J. H. (2005). Inference with weak instruments. Working Paper No. 313, National Bureau of Economic Research.
- [4] ANGRIST, J. D. and KRUEGER, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. Econ. Perspect.* **15** 69–85.
- [5] BAIocchi, M., CHENG, J. and SMALL, D. S. (2014). Instrumental variable methods for causal inference. *Stat. Med.* **33** 2297–2340. [MR3257582 https://doi.org/10.1002/sim.6128](https://doi.org/10.1002/sim.6128)
- [6] BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44** 512–525.
- [7] BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** 304–314.
- [8] BOWDEN, J., DEL GRECO M, F., MINELLI, C., DAVEY SMITH, G., SHEEHAN, N. and THOMPSON, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36** 1783–1802. [MR3648622 https://doi.org/10.1002/sim.7221](https://doi.org/10.1002/sim.7221)
- [9] BOWDEN, J., DEL GRECO M, F., MINELLI, C., ZHAO, Q., LAWLOR, D. A., SHEEHAN, N. A., THOMPSON, J. and DAVEY SMITH, G. (2019). Improving the accuracy of two-sample summary-data Mendelian randomization: Moving beyond the NOME assumption. *Int. J. Epidemiol.* **48** 728–742.
- [10] BURGESS, S., BUTTERWORTH, A. and THOMPSON, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** 658–665.

MSC2020 subject classifications. Primary 62E30, 60K35; secondary 46N60, 62P10.

Key words and phrases. Causal inference, inverse variance weighted estimator, many weak instruments, Mendelian randomization, summary data.

- [11] BURGESS, S., SMALL, D. S. and THOMPSON, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26** 2333–2355. MR3712236 <https://doi.org/10.1177/0962280215597579>
- [12] BURGESS, S. and THOMPSON, S. G. (2011). Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat. Med.* **30** 1312–1323. MR2828901 <https://doi.org/10.1002/sim.4197>
- [13] BURGESS, S. and THOMPSON, S. G. (2012). Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Stat. Med.* **31** 1582–1600. MR2947529 <https://doi.org/10.1002/sim.4498>
- [14] BURGESS, S., THOMPSON, S. G. and COLLABORATION, C. C. G. (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40** 755–764.
- [15] CHAO, J. C. and SWANSON, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica* **73** 1673–1692. MR2156676 <https://doi.org/10.1111/j.1468-0262.2005.00632.x>
- [16] CIRULLI, E. T. and GOLDSTEIN, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11** 415–425.
- [17] CORBIN, L. J., RICHMOND, R. C., WADE, K. H., BURGESS, S., BOWDEN, J., SMITH, G. D. and TIMPSON, N. J. (2016). BMI as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomization. *Diabetes* **65** 3002–3007. <https://doi.org/10.2337/db16-0418>
- [18] DAVEY SMITH, G. and EBRAHIM, S. (2003). ‘Mendelian randomization’: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32** 1–22.
- [19] DIDELEZ, V. and SHEEHAN, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16** 309–330. MR2395652 <https://doi.org/10.1177/0962280206077743>
- [20] GIBSON, G. (2012). Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13** 135–145.
- [21] HARTWIG, F. P., SMITH, G. D. and BOWDEN, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46** 1985–1998. <https://doi.org/10.1093/ije/dyx102>
- [22] HEMANI, G., BOWDEN, J. and SMITH, G. D. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27** R195–R208. <https://doi.org/10.1093/hmg/ddy163>
- [23] HEMANI, G., ZHENG, J., ELSWORTH, B., WADE, K. H., HABERLAND, V., BAIRD, D., LAURIN, C., BURGESS, S., BOWDEN, J. et al. (2018). The MR-base platform supports systematic causal inference across the human phenome. *eLife* **7** e34408.
- [24] HERNAN, M. A. and ROBINS, J. M. (2020). *Causal Inference: What If*. CRC Press/CRC, Boca Raton.
- [25] KAMSTRUP, P. R., TYBJÆRG-HANSEN, A., STEFFENSEN, R. and NORDESTGAARD, B. G. (2009). Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* **301** 2331–2339.
- [26] LAWLOR, D. A., HARBORD, R. M., STERNE, J. A. C., TIMPSON, N. and SMITH, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27** 1133–1163. MR2420151 <https://doi.org/10.1002/sim.3034>
- [27] PIERCE, B. L. and BURGESS, S. (2013). Efficient design for Mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178** 1177–1184.
- [28] PINGAULT, J.-B., O’REILLY, P. F., SCHOELER, T., PLOUBIDIS, G. B., RIJSDIJK, F. and DUDBRIDGE, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19** 566–580.
- [29] QI, G. and CHATTERJEE, N. (2019). Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat. Commun.* **10** 1941. <https://doi.org/10.1038/s41467-019-09432-2>
- [30] SAWA, T. (1969). The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *J. Amer. Statist. Assoc.* **64** 923–937. MR0245118
- [31] SMITH, G. D. and EBRAHIM, S. (2004). Mendelian randomization: Prospects, potentials, and limitations. *Int. J. Epidemiol.* **33** 30–42. <https://doi.org/10.1093/ije/dyh132>
- [32] SOLOVIEFF, N., COTSAPAS, C., LEE, P. H., PURCELL, S. M. and SMOLLER, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nat. Rev. Genet.* **14** 483–495.
- [33] STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65** 557–586. MR1445622 <https://doi.org/10.2307/2171753>
- [34] STOCK, J. H., WRIGHT, J. H. and YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *J. Bus. Econom. Statist.* **20** 518–529. MR1973801 <https://doi.org/10.1198/073500102288618658>

- [35] THE CARDIOGRAMPLUSC4D CONSORTIUM, NIKPAY, M., GOEL, A., WON, H.-H., HALL, L. M., WILLENBORG, C. et al. (2015). A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47** 1121–1130.
- [36] VERBANCK, M., CHEN, C.-Y., NEALE, B. and DO, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50** 693–698.
- [37] VISSCHER, P. M., HILL, W. G. and WRAY, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9** 255–266.
- [38] WANG, S. and KANG, H. (2019). Weak-instrument robust tests in two-sample summary-data Mendelian randomization. Available at [arXiv:1909.06950](https://arxiv.org/abs/1909.06950).
- [39] YAVORSKA, O. O. and BURGESS, S. (2017). MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46** 1734–1739.
- [40] YE, T., SHAO, J. and KANG, H. (2021). Supplement to “Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization.” <https://doi.org/10.1214/20-AOS2027SUPP>
- [41] ZHAO, Q., CHEN, Y., WANG, J. and SMALL, D. S. (2019). Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization. *Int. J. Epidemiol.*
- [42] ZHAO, Q., WANG, J., HEMANI, G., BOWDEN, J. and SMALL, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Statist.* **48** 1742–1769. [MR4124342 https://doi.org/10.1214/19-AOS1866](https://doi.org/10.1214/19-AOS1866)
- [43] ZHENG, J., BAIRD, D., BORGES, M.-C., BOWDEN, J., HEMANI, G., HAYCOCK, P., EVANS, D. M. and SMITH, G. D. (2017). Recent developments in Mendelian randomization studies. *Current Epidemiology Reports* **4** 330–345.

BOOSTED NONPARAMETRIC HAZARDS WITH TIME-DEPENDENT COVARIATES

BY DONALD K. K. LEE¹, NINGYUAN CHEN² AND HEMANT ISHWARAN³

¹Goizueta Business School and Department of Biostatistics and Bioinformatics, Emory University, donald.lee@emory.edu

²Rotman School of Management, University of Toronto, ningyuan.chen@utoronto.ca

³Division of Biostatistics, University of Miami, hemant.ishwaran@gmail.com

Given functional data from a survival process with time-dependent covariates, we derive a smooth convex representation for its nonparametric log-likelihood functional and obtain its functional gradient. From this, we devise a generic gradient boosting procedure for estimating the hazard function nonparametrically. An illustrative implementation of the procedure using regression trees is described to show how to recover the unknown hazard. The generic estimator is consistent if the model is correctly specified; alternatively, an oracle inequality can be demonstrated for tree-based models. To avoid overfitting, boosting employs several regularization devices. One of them is stepsize restriction, but the rationale for this is somewhat mysterious from the viewpoint of consistency. Our work brings some clarity to this issue by revealing that stepsize restriction is a mechanism for preventing the curvature of the risk from derailing convergence.

REFERENCES

- [1] AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726. [MR0491547](#)
- [2] ADELSON, K., LEE, D. K. K., VELJI, S., MA, J., LIPKA, S. K., RIMAR, J., LONGLEY, P., VEGA, T., PEREZ-IRIZARRY, J. et al. (2018). Development of Imminent Mortality Predictor for Advanced Cancer (IMPAC), a tool to predict short-term mortality in hospitalized patients with advanced cancer. *Journal of Oncology Practice* **14** e168–e175.
- [3] BIAU, G. and CADRE, B. (2017). Optimization by gradient boosting. arXiv preprint [arXiv:1707.05023](#).
- [4] BINDER, H. and SCHUMACHER, M. (2008). Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC Bioinform.* **9** 14. <https://doi.org/10.1186/1471-2105-9-14>
- [5] BLANCHARD, G., LUGOSI, G. and VAYATIS, N. (2004). On the rate of convergence of regularized boosting classifiers. *J. Mach. Learn. Res.* **4** 861–894. [MR2076000](#) <https://doi.org/10.1162/1532443041424319>
- [6] BREIMAN, L. (1997). Arcing the edge. U.C. Berkeley Dept. of Statistics Technical Report 486.
- [7] BREIMAN, L. (1999). Prediction games and arcing algorithms. *Neural Comput.* **11** 1493–1517.
- [8] BREIMAN, L. (2004). Population theory for boosting ensembles. *Ann. Statist.* **32** 1–11. [MR2050998](#) <https://doi.org/10.1214/aos/1079120126>
- [9] BÜHLMANN, P. (2002). Consistency for L_2 boosting and matching pursuit with trees and tree-type basis functions. Research Report 109. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH).
- [10] BÜHLMANN, P. and HOTHORN, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statist. Sci.* **22** 477–505. [MR2420454](#) <https://doi.org/10.1214/07-STS242>
- [11] BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339. [MR1995709](#) <https://doi.org/10.1198/016214503000125>
- [12] EHRLINGER, J. and ISHWARAN, H. (2012). Characterizing L_2 Boosting. *Ann. Statist.* **40** 1074–1101. [MR2985944](#) <https://doi.org/10.1214/12-AOS997>
- [13] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. [MR1873328](#) <https://doi.org/10.1214/aos/1013203451>

MSC2020 subject classifications. Primary 62N02; secondary 62G05, 90B22.

Key words and phrases. Survival analysis, gradient boosting, functional data, stepsize shrinkage, regression trees, likelihood functional.

- [14] HOORFAR, A. and HASSANI, M. (2008). Inequalities on the Lambert W function and hyperpower function. *JIPAM. J. Inequal. Pure Appl. Math.* **9** Article 51, 5. MR2417333 <https://doi.org/10.1145/1342991.1342997>
- [15] HOTHORN, T. (2020). Transformation boosting machines. *Stat. Comput.* **30** 141–152. MR4057476 <https://doi.org/10.1007/s11222-019-09870-4>
- [16] HOTHORN, T., BÜHLMANN, P., DUDOIT, S., MOLINARO, A. and VAN DER LAAN, M. J. (2006). Survival ensembles. *Biostatistics* **7** 355–373.
- [17] HOTHORN, T. and ZEILEIS, A. (2017). Transformation forests. arXiv preprint [arXiv:1701.02110](https://arxiv.org/abs/1701.02110).
- [18] HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. MR2247210 <https://doi.org/10.1111/j.1541-0420.2006.00562.x>
- [19] HUANG, J. Z. and STONE, C. J. (1998). The L_2 rate of convergence for event history regression with time-dependent covariates. *Scand. J. Stat.* **25** 603–620. MR1666788 <https://doi.org/10.1111/1467-9469.00124>
- [20] LOWSKY, D. J., DING, Y., LEE, D. K. K., MCCULLOCH, C. E., ROSS, L. F., THISTLETHWAITE, J. R. and ZENIOS, S. A. (2013). A K -nearest neighbors survival probability prediction method. *Stat. Med.* **32** 2062–2069. MR3067356 <https://doi.org/10.1002/sim.5673>
- [21] MASON, L., BAXTER, J., BARTLETT, P. L. and FREAN, M. R. (1999). Functional gradient techniques for combining hypotheses. In *Advances in Neural Information Processing Systems* 221–246.
- [22] MASON, L., BAXTER, J., BARTLETT, P. L. and FREAN, M. R. (2000). Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems* 512–518.
- [23] MEYER, D. M., ROGERS, J. G., EDWARDS, L. B., CALLAHAN, E. R., WEBBER, S. A., JOHNSON, M. R., VEGA, J. D., ZUCKER, M. J. and CLEVELAND JR., J. C. (2015). The future direction of the adult heart allocation system in the United States. *Am. J. Transplant.* **15** 44–54.
- [24] MÜLLER, H.-G. and YAO, F. (2010). Additive modelling of functional gradients. *Biometrika* **97** 791–805. MR2746152 <https://doi.org/10.1093/biomet/asq056>
- [25] NESTEROV, Y. (2004). *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization **87**. Kluwer Academic, Boston, MA. MR2142598 <https://doi.org/10.1007/978-1-4419-8853-9>
- [26] RIDGEWAY, G. (1999). The state of boosting. *Computing Science and Statistics* **31** 172–181.
- [27] SCHMID, M. and HOTHORN, T. (2008). Flexible boosting of accelerated failure time models. *BMC Bioinform.* **9** 269. <https://doi.org/10.1186/1471-2105-9-269>
- [28] TELGARSKY, M. (2013). Margins, shrinkage, and boosting. In *Proceedings of the 30th International Conference on International Conference on Machine Learning* 28 II–307.
- [29] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York. MR1385671 <https://doi.org/10.1007/978-1-4757-2545-2>
- [30] VUONG, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica* **57** 307–333. MR0996939 <https://doi.org/10.2307/1912557>
- [31] ZHANG, T. and YU, B. (2005). Boosting with early stopping: Convergence and consistency. *Ann. Statist.* **33** 1538–1579. MR2166555 <https://doi.org/10.1214/009053605000000255>
- [32] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the “degrees of freedom” of the lasso. *Ann. Statist.* **35** 2173–2192. MR2363967 <https://doi.org/10.1214/009053607000000127>

UNIVERSAL BAYES CONSISTENCY IN METRIC SPACES

BY STEVE HANNEKE¹, ARYEH KONTOROVICH^{2,*}, SIVAN SABATO^{2,†} AND ROI WEISS³

¹*Toyota Technological Institute at Chicago, steve.hanneke@gmail.com*

²*Department of Computer Science, Ben-Gurion University of the Negev, *karyeh@cs.bgu.ac.il; †sabatos@cs.bgu.ac.il*

³*Department of Computer Science, Ariel University, roiw@ariel.ac.il*

We extend a recently proposed 1-nearest-neighbor based multiclass learning algorithm and prove that our modification is universally strongly Bayes consistent in all metric spaces admitting *any* such learner, making it an “optimistically universal” Bayes-consistent learner. This is the first learning algorithm known to enjoy this property; by comparison, the k -NN classifier and its variants are not generally universally Bayes consistent, except under additional structural assumptions, such as an inner product, a norm, finite dimension or a Besicovitch-type property.

The metric spaces in which universal Bayes consistency is possible are the “essentially separable” ones—a notion that we define, which is more general than standard separability. The existence of metric spaces that are not essentially separable is widely believed to be independent of the ZFC axioms of set theory. We prove that essential separability exactly characterizes the existence of a universal Bayes-consistent learner for the given metric space. In particular, this yields the first impossibility result for universal Bayes consistency.

Taken together, our results completely characterize strong and weak universal Bayes consistency in metric spaces.

REFERENCES

- [1] ABRAHAM, C., BIAU, G. and CADRE, B. (2006). On the kernel rule for function classification. *Ann. Inst. Statist. Math.* **58** 619–633. [MR2327897](#) <https://doi.org/10.1007/s10463-006-0032-1>
- [2] BEN-DAVID, S., HRUBES, P., MORAN, S., SHPILKA, A. and YEHUDAYOFF, A. (2019). Learnability can be undecidable. *Nat. Mach. Intell.* **1** 44–48. <https://doi.org/10.1038/s42256-018-0002-3>
- [3] BIAU, G., BUNEA, F. and WEGKAMP, M. H. (2005). Functional classification in Hilbert spaces. *IEEE Trans. Inf. Theory* **51** 2163–2172. [MR2235289](#) <https://doi.org/10.1109/TIT.2005.847705>
- [4] BIAU, G., CÉROU, F. and GUYADER, A. (2010). Rates of convergence of the functional k -nearest neighbor estimate. *IEEE Trans. Inf. Theory* **56** 2034–2040. [MR2654492](#) <https://doi.org/10.1109/TIT.2010.2040857>
- [5] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. [MR1700749](#) <https://doi.org/10.1002/9780470316962>
- [6] CÉROU, F. and GUYADER, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat.* **10** 340–355. [MR2247925](#) <https://doi.org/10.1051/ps:2006014>
- [7] CHAUDHURI, K. and DASGUPTA, S. (2014). Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems* 3437–3445.
- [8] CHRISTMANN, A. and STEINWART, I. (2010). Universal kernels on non-standard input spaces. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems* 2010 406–414.
- [9] COVER, T. M. and HART, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13** 21–27.
- [10] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. Springer, New York. [MR1383093](#) <https://doi.org/10.1007/978-1-4612-0711-5>

MSC2020 subject classifications. Primary 54E70, 97K80, 62C12; secondary 03E17, 03E55.

Key words and phrases. Metric space, nearest neighbor, classification, Bayes consistency.

- [11] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. [MR1720712](#) <https://doi.org/10.1017/CBO9780511665622>
- [12] FIX, E. and HODGES J. L., J. (1989). Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **57** 238–247.
- [13] FORZANI, L., FRAIMAN, R. and LLOP, P. (2012). Consistent nonparametric regression for functional data under the Stone–Besicovitch conditions. *IEEE Trans. Inf. Theory* **58** 6697–6708. [MR2991802](#) <https://doi.org/10.1109/TIT.2012.2209628>
- [14] FREMLIN, D. H. (1993). Real-valued-measurable cardinals. In *Set Theory of the Reals (Ramat Gan, 1991)*. *Israel Math. Conf. Proc.* **6** 151–304. Bar-Ilan Univ., Ramat Gan. [MR1234282](#)
- [15] FREMLIN, D. H. (2000). *Measure Theory* **1–5**. Torres Fremlin, Colchester.
- [16] GITIK, M. and SHELAH, S. (1989). Forcings with ideals and simple forcing notions. *Israel J. Math.* **68** 129–160. [MR1035887](#) <https://doi.org/10.1007/BF02772658>
- [17] GOTTLIEB, L.-A., KONTOROVICH, A. and KRAUTHGAMER, R. (2014). Efficient classification for metric data. *IEEE Trans. Inf. Theory* **60** 5750–5759. [MR3252418](#) <https://doi.org/10.1109/TIT.2014.2339840>
- [18] GOTTLIEB, L.-A., KONTOROVICH, A. and NISNEVITCH, P. (2017). Nearly optimal classification for semi-metrics. *J. Mach. Learn. Res.* **18** Paper No. 37, 22. [MR3646632](#)
- [19] GOTTLIEB, L.-A., KONTOROVICH, A. and NISNEVITCH, P. (2018). Near-optimal sample compression for nearest neighbors. *IEEE Trans. Inf. Theory* **64** 4120–4128. [MR3809730](#) <https://doi.org/10.1109/TIT.2018.2822267>
- [20] GRAEPEL, T., HERBRICH, R. and SHAWE-TAYLOR, J. (2005). PAC-Bayesian compression bounds on the prediction error of learning algorithms for classification. *Mach. Learn.* **59** 55–76.
- [21] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. Springer, New York. [MR1920390](#) <https://doi.org/10.1007/b97848>
- [22] HANNEKE, S. (2017). Learning whenever learning is possible: Universal learning under general stochastic processes. Preprint. Available at [arXiv:1706.01418](#).
- [23] HANNEKE, S. and KONTOROVICH, A. (2019). A sharp lower bound for agnostic learning with sample compression schemes. In *Algorithmic Learning Theory 2019. Proc. Mach. Learn. Res. (PMLR)* **98** 489–505. [MR3932856](#)
- [24] HANNEKE, S., KONTOROVICH, A., SABATO, S. and WEISS, R. (2021). Supplement to “Universal Bayes consistency in metric spaces.” <https://doi.org/10.1214/20-AOS2029SUPP>
- [25] HRBACEK, K. and JECH, T. (1999). *Introduction to Set Theory*, 3rd ed. *Monographs and Textbooks in Pure and Applied Mathematics* **220**. Dekker, New York. [MR1697766](#)
- [26] JECH, T. (2003). *Set Theory*. *Springer Monographs in Mathematics*. Springer, Berlin. The third millennium edition, revised and expanded. [MR1940513](#)
- [27] KONTOROVICH, A., SABATO, S. and URNER, R. (2017). Active nearest-neighbor learning in metric spaces. *J. Mach. Learn. Res.* **18** Paper No. 195, 38. [MR3827083](#)
- [28] KONTOROVICH, A., SABATO, S. and WEISS, R. (2017). Nearest-neighbor sample compression: Efficiency, consistency, infinite dimensions. In *Advances in Neural Information Processing Systems* 1573–1583.
- [29] KONTOROVICH, A., SABATO, S. and WEISS, R. (2017). Nearest-neighbor sample compression: Efficiency, consistency, infinite dimensions. Preprint. Available at [arXiv:1705.08184](#).
- [30] KONTOROVICH, A. and WEISS, R. (2014). Maximum margin multiclass nearest neighbors. In *International Conference on Machine Learning (ICML 2014)*.
- [31] KONTOROVICH, A. and WEISS, R. (2014). A Bayes consistent 1-NN classifier. In *Artificial Intelligence and Statistics (AISTATS 2015)*.
- [32] KPOTUFE, S. and VERMA, N. (2017). Time-accuracy tradeoffs in kernel prediction: Controlling prediction quality. *J. Mach. Learn. Res.* **18** Paper No. 44, 29. [MR3655309](#)
- [33] KULKARNI, S. R. and POSNER, S. E. (1995). Rates of convergence of nearest neighbor estimation under arbitrary sampling. *IEEE Trans. Inf. Theory* **41** 1028–1039. [MR1366756](#) <https://doi.org/10.1109/18.391248>
- [34] LITTLESTONE, N. and WARMUTH, M. K. (1986). Relating data compression and learnability. Unpublished.
- [35] PELILLO, M. (2014). Alhazen and the nearest neighbor rule. *Pattern Recogn. Lett.* **38** 34–37. <https://doi.org/10.1016/j.patrec.2013.10.022>
- [36] PREISS, D. (1979). Invalid Vitali theorems. In *Abstracta. 7th Winter School on Abstract Analysis* 58–60.
- [37] PREISS, D. (1981). Gaussian measures and the density theorem. *Comment. Math. Univ. Carolin.* **22** 181–193. [MR0609946](#)
- [38] SHALEV-SHWARTZ, S. and BEN-DAVID, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge Univ. Press, Cambridge.
- [39] STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. [MR0443204](#)

- [40] TIŠER, J. (2003). Vitali covering theorem in Hilbert space. *Trans. Amer. Math. Soc.* **355** 3277–3289. [MR1974687](#) <https://doi.org/10.1090/S0002-9947-03-03296-3>
- [41] ULAM, S. M. (1930). *Zur Masstheorie in der Allgemeinen Mengenlehre*. Uniwersytet, seminarjum matematyczne.
- [42] ZHAO, L. C. (1987). Exponential bounds of mean error for the nearest neighbor estimates of regression functions. *J. Multivariate Anal.* **21** 168–178. [MR0877849](#) [https://doi.org/10.1016/0047-259X\(87\)90105-9](https://doi.org/10.1016/0047-259X(87)90105-9)

MINIMAX OPTIMAL CONDITIONAL INDEPENDENCE TESTING

BY MATEY NEYKOV^{*}, SIVARAMAN BALAKRISHNAN[†] AND LARRY WASSERMAN[‡]

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

We consider the problem of conditional independence testing of X and Y given Z where X , Y and Z are three real random variables and Z is continuous. We focus on two main cases—when X and Y are both discrete, and when X and Y are both continuous. In view of recent results on conditional independence testing [*Ann. Statist.* **48** (2020) 1514–1538], one cannot hope to design nontrivial tests, which control the type I error for all absolutely continuous conditionally independent distributions, while still ensuring power against interesting alternatives. Consequently, we identify various, natural smoothness assumptions on the conditional distributions of X , $Y|Z = z$ as z varies in the support of Z , and study the hardness of conditional independence testing under these smoothness assumptions. We derive matching lower and upper bounds on the critical radius of separation between the null and alternative hypotheses in the total variation metric. The tests we consider are easily implementable and rely on binning the support of the continuous variable Z . To complement these results, we provide a new proof of the hardness result of Shah and Peters [*Ann. Statist.* **48** (2020) 1514–1538].

REFERENCES

- [1] AGRESTI, A. (1992). A survey of exact inference for contingency tables. *Statist. Sci.* **7** 131–177. [MR1173420](#)
- [2] ARIAS-CASTRO, E., PELLETIER, B. and SALIGRAMA, V. (2018). Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *J. Nonparametr. Stat.* **30** 448–471. [MR3794401](#) <https://doi.org/10.1080/10485252.2018.1435875>
- [3] AZADKIA, M. and CHATTERJEE, S. (2019). A simple measure of conditional dependence. Preprint. Available at [arXiv:1910.12327](#).
- [4] BALAKRISHNAN, S. and WASSERMAN, L. (2018). Hypothesis testing for high-dimensional multinomials: A selective review. *Ann. Appl. Stat.* **12** 727–749. [MR3834283](#) <https://doi.org/10.1214/18-AOAS1155SF>
- [5] BALAKRISHNAN, S. and WASSERMAN, L. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. [MR3953439](#) <https://doi.org/10.1214/18-AOS1729>
- [6] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. [MR1935648](#)
- [7] BERGSMA, W. (2010). Nonparametric testing of conditional independence by means of the partial copula. Available at SSRN 1702981.
- [8] BERGSMA, W. P. (2004). Testing conditional independence for continuous random variables. *Eurandom*.
- [9] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2018). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. [MR4060981](#) <https://doi.org/10.1111/rssb.12340>
- [10] BLAIS, E., CANONNE, C. L. and GUR, T. (2019). Distribution testing lower bounds via reductions from communication complexity. *ACM Trans. Comput. Theory* **11** 1–37. [MR3940784](#) <https://doi.org/10.1145/3305270>
- [11] CANONNE, C. L. (2015). A survey on distribution testing: Your data is big. But is it blue? In *Electronic Colloquium on Computational Complexity (ECCC)* **22** 1–1.

- [12] CANONNE, C. L., DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2018). Testing conditional independence of discrete distributions. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 735–748. ACM, New York. [MR3826290](#) <https://doi.org/10.1145/3188745.3188756>
- [13] DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31. [MR0535541](#)
- [14] DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* 685–694. IEEE Computer Soc., Los Alamitos, CA. [MR3631031](#)
- [15] DORAN, G., MUANDET, K., ZHANG, K. and SCHÖLKOPF, B. (2014). A permutation-based kernel conditional independence test. In *UAI* 132–141.
- [16] FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* 489–496.
- [17] GRETTON, A. and GYÖRFI, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.* **11** 1391–1423. [MR2645456](#)
- [18] GYÖRFI, L. and WALK, H. (2012). Strongly consistent nonparametric tests of conditional independence. *Statist. Probab. Lett.* **82** 1145–1150. [MR2915081](#) <https://doi.org/10.1016/j.spl.2012.02.023>
- [19] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#) <https://doi.org/10.1214/aoms/1177730196>
- [20] HUANG, T.-M. (2010). Testing conditional independence using maximal nonlinear conditional correlation. *Ann. Statist.* **38** 2047–2091. [MR2676883](#) <https://doi.org/10.1214/09-AOS770>
- [21] INGSTER, Y. I. Minimax nonparametric detection of signals in white Gaussian noise. *Problemy Peredachi Informatsii* **18** 61–73. [MR0689340](#)
- [22] 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](#) <https://doi.org/10.1007/978-0-387-21580-8>
- [23] KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models: Principles and Techniques. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2778120](#)
- [24] MARGARITIS, D. (2005). Distribution-free learning of Bayesian network structure in continuous domains. In *AAAI* **5** 825–830.
- [25] NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2021). Supplement to “Minimax optimal conditional independence testing.” <https://doi.org/10.1214/20-AOS2030SUPP>.
- [26] PATRA, R. K., SEN, B. and SZÉKELY, G. J. (2016). On a nonparametric notion of residual and its applications. *Statist. Probab. Lett.* **109** 208–213. [MR3434980](#) <https://doi.org/10.1016/j.spl.2015.10.011>
- [27] PEARL, J. (2014). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Elsevier.
- [28] ROSENBAUM, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika* **49** 425–435. [MR0760206](#) <https://doi.org/10.1007/BF02306030>
- [29] SEN, R., THEERTHA SURESH, A., SHANMUGAM, K., DIMAKIS, A. G. and SHAKKOTAI, S. (2017). Model-powered conditional independence test. In *Advances in Neural Information Processing Systems* 2951–2961.
- [30] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- [31] SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. [MR4124333](#) <https://doi.org/10.1214/19-AOS1857>
- [32] SONG, K. (2009). Testing conditional independence via Rosenblatt transforms. *Ann. Statist.* **37** 4011–4045. [MR2572451](#) <https://doi.org/10.1214/09-AOS704>
- [33] SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR1815675](#)
- [34] SU, L. and WHITE, H. (2007). A consistent characteristic function-based test for conditional independence. *J. Econometrics* **141** 807–834. [MR2413488](#) <https://doi.org/10.1016/j.jeconom.2006.11.006>
- [35] SU, L. and WHITE, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* **24** 829–864. [MR2428851](#) <https://doi.org/10.1017/S0266466608080341>
- [36] SU, L. and WHITE, H. (2014). Testing conditional independence via empirical likelihood. *J. Econometrics* **182** 27–44. [MR3212759](#) <https://doi.org/10.1016/j.jeconom.2014.04.006>
- [37] VALIANT, G. and VALIANT, P. (2017). An automatic inequality prover and instance optimal identity testing. *SIAM J. Comput.* **46** 429–455. [MR3614697](#) <https://doi.org/10.1137/151002526>
- [38] VALIANT, G. and VALIANT, P. (2017). Estimating the unseen: Improved estimators for entropy and other properties. *J. ACM* **64** Art. 37, 41. [MR3713795](#) <https://doi.org/10.1145/3125643>

- [39] WANG, X. and HONG, Y. (2018). Characteristic function based testing for conditional independence: A non-parametric regression approach. *Econometric Theory* **34** 815–849. [MR3815468](#) <https://doi.org/10.1017/S026646661700010X>
- [40] YAO, Q. and TRITCHLER, D. (1993). An exact analysis of conditional independence in several 2×2 contingency tables. *Biometrics* **49** 233–236. [MR1221407](#) <https://doi.org/10.2307/2532617>
- [41] ZHANG, K., PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2012). Kernel-based conditional independence test and application in causal discovery. Preprint. Available at [arXiv:1202.3775](#).

ESTIMATION OF THE NUMBER OF COMPONENTS OF NONPARAMETRIC MULTIVARIATE FINITE MIXTURE MODELS

BY CALEB KWON¹ AND ERIC MBAKOP²

¹Harvard Business School, Harvard University, calebkwon@g.harvard.edu

²Department of Economics, University of Calgary, eric.mbakop@ucalgary.edu

We propose a novel estimator for the number of mixture components (denoted by M) in a nonparametric finite mixture model. The setting that we consider is one where the analyst has repeated observations of $K \geq 2$ variables that are conditionally independent given a finitely supported latent variable with M support points. Under a mild assumption on the joint distribution of the observed and latent variables, we show that an integral operator T that is identified from the data has rank equal to M . We use this observation, in conjunction with the fact that singular values of operators are stable under perturbations, to propose an estimator of M , which essentially consists of a thresholding rule that counts the number of singular values of a consistent estimator of T that are greater than a data-driven threshold. We prove that our estimator of M is consistent, and establish nonasymptotic results, which provide finite sample performance guarantees for our estimator. We present a Monte Carlo study, which shows that our estimator performs well for samples of moderate size.

REFERENCES

- [1] AGUIRREGABIRIA, V. and MIRA, P. (2019). Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. *Quant. Econ.* **10** 1659–1701. [MR4048251](#) <https://doi.org/10.3982/qe666>
- [2] ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. [MR2549554](#) <https://doi.org/10.1214/09-AOS689>
- [3] AN, Y., HU, Y. and SHUM, M. (2010). Estimating first-price auctions with an unknown number of bidders: A misclassification approach. *J. Econometrics* **157** 328–341. [MR2661605](#) <https://doi.org/10.1016/j.jeconom.2010.02.002>
- [4] BENAGLIA, T., CHAUVEAU, D. and HUNTER, D. R. (2009). An EM-like algorithm for semi- and non-parametric estimation in multivariate mixtures. *J. Comput. Graph. Statist.* **18** 505–526. [MR2749842](#) <https://doi.org/10.1198/jcgs.2009.07175>
- [5] BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.
- [6] BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2016). Non-parametric estimation of finite mixtures from repeated measurements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 211–229. [MR3453653](#) <https://doi.org/10.1111/rssb.12110>
- [7] BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2016). Estimating multivariate latent-structure models. *Ann. Statist.* **44** 540–563. [MR3476609](#) <https://doi.org/10.1214/15-AOS1376>
- [8] CLOGG, C. (1981). Latent structure models of mobility. *Am. J. Sociol.* **86** 836–868.
- [9] COMPIANI, G. and KITAMURA, Y. (2016). Using mixtures in econometric models: A brief review and some new results. *Econom. J.* **19** C95–C127. [MR3597287](#) <https://doi.org/10.1111/ectj.12068>
- [10] HALL, P., NEEMAN, A., PAKYARI, R. and ELMORE, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika* **92** 667–678. [MR2202653](#) <https://doi.org/10.1093/biomet/92.3.667>
- [11] HALL, P. and ZHOU, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31** 201–224. [MR1962504](#) <https://doi.org/10.1214/aos/1046294462>

MSC2020 subject classifications. Primary 62G05; secondary 62G15, 62H30, 47A55, 47G10, 47N30.

Key words and phrases. Finite mixture model, latent model, nonparametric mixture, conditional independence, multivariate data.

- [12] HETTMANSPERGER, T. P. and THOMAS, H. (2000). Almost nonparametric inference for repeated measures in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 811–825. [MR1796294](#) <https://doi.org/10.1111/1467-9868.00266>
- [13] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [14] HORN, R. A. and JOHNSON, C. R. (1991). *Topics in Matrix Analysis*. Cambridge Univ. Press, Cambridge. [MR1091716](#) <https://doi.org/10.1017/CBO9780511840371>
- [15] HU, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *J. Econometrics* **144** 27–61. [MR2439921](#) <https://doi.org/10.1016/j.jeconom.2007.12.001>
- [16] HU, Y., MCADAMS, D. and SHUM, M. (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *J. Econometrics* **174** 186–193. [MR3045027](#) <https://doi.org/10.1016/j.jeconom.2013.02.005>
- [17] HU, Y. and SHUM, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *J. Econometrics* **171** 32–44. [MR2970334](#) <https://doi.org/10.1016/j.jeconom.2012.05.023>
- [18] KASAHARA, H. and SHIMOTSU, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* **77** 135–175. [MR2477846](#) <https://doi.org/10.3982/ECTA6763>
- [19] KASAHARA, H. and SHIMOTSU, K. (2014). Non-parametric identification and estimation of the number of components in multivariate mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 97–111. [MR3153935](#) <https://doi.org/10.1111/rssb.12022>
- [20] KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167. [MR1781185](#) <https://doi.org/10.2307/3318636>
- [21] KRESS, R. (1999). *Linear Integral Equations*, 2nd ed. *Applied Mathematical Sciences* **82**. Springer, New York. [MR1723850](#) <https://doi.org/10.1007/978-1-4612-0559-3>
- [22] KWON, C. and MBAKOP, E. (2021). Supplement to “Estimation of the number of components of nonparametric multivariate finite mixture models.” <https://doi.org/10.1214/20-AOS2032SUPP>
- [23] LEVINE, M., HUNTER, D. R. and CHAUVEAU, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika* **98** 403–416. [MR2806437](#) <https://doi.org/10.1093/biomet/asq079>
- [24] MBAKOP, E. (2017). Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity. Working Papers, Univ. Calgary.
- [25] MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics: Applied Probability and Statistics*. Wiley Interscience, New York. [MR1789474](#) <https://doi.org/10.1002/0471721182>
- [26] MISLEVY, R. J. (1984). Estimating latent distribution. *Psychometrika* **49** 359–381.
- [27] PINELIS, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. [MR1331198](#)
- [28] ROSASCO, L., BELKIN, M. and DE VITO, E. (2010). On learning with integral operators. *J. Mach. Learn. Res.* **11** 905–934. [MR2600634](#)
- [29] SMALE, S. and ZHOU, D.-X. (2007). Learning theory estimates via integral operators and their approximations. *Constr. Approx.* **26** 153–172. [MR2327597](#) <https://doi.org/10.1007/s00365-006-0659-y>
- [30] VAN DER HEIJDEN, P. G. M., VAN DER ARK, L. A. and MOOIJART, A. (2002). Some examples of latent budget analysis and its extensions. In *Applied Latent Class Analysis* 107–136. Cambridge Univ. Press, Cambridge. [MR1927667](#) <https://doi.org/10.1017/CBO9780511499531.005>
- [31] XIAO, R. (2018). Identification and estimation of incomplete information games with multiple equilibria. *J. Econometrics* **203** 328–343. [MR3770830](#) <https://doi.org/10.1016/j.jeconom.2017.12.005>
- [32] ZWALD, L. and BLANCHARD, G. (2005). On the convergence of eigenspaces in kernel principal component analysis. In *NIPS*.

ROBUST k -MEANS CLUSTERING FOR DISTRIBUTIONS WITH TWO MOMENTS

BY YEGOR KLOCHKOV¹, ALEXEY KROSHNIN² AND NIKITA ZHIVOTOVSKIY³

¹Cambridge-INET, Faculty of Economics, University of Cambridge, yk376@cam.ac.uk

²HSE University and Institute for Information Transmission Problems, RAS, akroshnin@hse.ru

³Google Research, Brain Team, Zürich, zhivotovskiy@google.com

We consider the robust algorithms for the k -means clustering problem where a quantizer is constructed based on N independent observations. Our main results are median of means based nonasymptotic excess distortion bounds that hold under the two bounded moments assumption in a general separable Hilbert space. In particular, our results extend the renowned asymptotic result of (Ann. Statist. **9** (1981) 135–140) who showed that the existence of two moments is sufficient for strong consistency of an empirically optimal quantizer in $\mathbb{R}^d$. In a special case of clustering in $\mathbb{R}^d$, under two bounded moments, we prove matching (up to constant factors) nonasymptotic upper and lower bounds on the excess distortion, which depend on the probability mass of the lightest cluster of an optimal quantizer. Our bounds have the sub-Gaussian form, and the proofs are based on the versions of uniform bounds for robust mean estimators.

REFERENCES

- ALON, N., MATIAS, Y. and SZEGEDY, M. (1999). The space complexity of approximating the frequency moments. *J. Comput. System Sci.* **58** 137–147. [MR1688610](#) <https://doi.org/10.1006/jcss.1997.1545>
- ANTOS, A. (2005). Improved minimax bounds on the test and training distortion of empirically designed vector quantizers. *IEEE Trans. Inf. Theory* **51** 4022–4032. [MR2239018](#) <https://doi.org/10.1109/TIT.2005.856980>
- BACHEM, O., LUCIC, M., HASSANI, S. H. and KRAUSE, A. (2017). Uniform deviation bounds for k -means clustering. In *International Conference on Machine Learning* 283–291.
- BARTLETT, P. L., LINDER, T. and LUGOSI, G. (1998). The minimax distortion redundancy in empirical quantizer design. *IEEE Trans. Inf. Theory* **44** 1802–1813. [MR1664098](#) <https://doi.org/10.1109/18.705560>
- BIAU, G., DEVROYE, L. and LUGOSI, G. (2008). On the performance of clustering in Hilbert spaces. *IEEE Trans. Inf. Theory* **54** 781–790. [MR2444554](#) <https://doi.org/10.1109/TIT.2007.913516>
- BOUCHERON, S., BOUSQUET, O. and LUGOSI, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM Probab. Stat.* **9** 323–375. [MR2182250](#) <https://doi.org/10.1051/ps:2005018>
- BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. [MR3185193](#) <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- BRÉCHETEAU, C., FISCHER, A. and LEVRARD, C. (2018). Robust Bregman clustering. Preprint. Available at [arXiv:1812.04356](https://arxiv.org/abs/1812.04356).
- BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. [MR3405602](#) <https://doi.org/10.1214/15-AOS1350>
- CADRE, B. and PARIS, Q. (2012). On Hölder fields clustering. *TEST* **21** 301–316. [MR2935361](#) <https://doi.org/10.1007/s11749-011-0244-4>
- CUTURI, M. and DOUCET, A. (2014). Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning* 685–693.
- FEFFERMAN, C., MITTER, S. and NARAYANAN, H. (2016). Testing the manifold hypothesis. *J. Amer. Math. Soc.* **29** 983–1049. [MR3522608](#) <https://doi.org/10.1090/jams/852>
- FISCHER, A. (2010). Quantization and clustering with Bregman divergences. *J. Multivariate Anal.* **101** 2207–2221. [MR2671211](#) <https://doi.org/10.1016/j.jmva.2010.05.008>
- FOSTER, D. J. and RAKHLIN, A. (2019). ℓ_∞ -Vector contraction for Rademacher complexity. Preprint. Available at [arXiv:1911.06468](https://arxiv.org/abs/1911.06468).

- GENEVAY, A., DULAC-ARNOLD, G. and VERT, J.-P. (2019). Differentiable deep clustering with cluster size constraints. Preprint. Available at [arXiv:1910.09036](https://arxiv.org/abs/1910.09036).
- GRAF, S. and LUSCHGY, H. (2007). *Foundations of Quantization for Probability Distributions. Lecture Notes in Math.* **1730**. Springer, Berlin. [MR1764176 https://doi.org/10.1007/BFb0103945](https://doi.org/10.1007/BFb0103945)
- JOHNSON, W. B. and LINDENSTRAUSS, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)*. *Contemp. Math.* **26** 189–206. Amer. Math. Soc., Providence, RI. [MR0737400 https://doi.org/10.1090/conm/026/737400](https://doi.org/10.1090/conm/026/737400)
- KLOCHKOV, Y., KROSHNIN, A. and ZHIVOTOVSKIY, N. (2021). Supplement to “Robust k -means clustering for distributions with two moments.” <https://doi.org/10.1214/20-AOS2033SUPP>
- LECUÉ, G. and LERASLE, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. [MR4102681 https://doi.org/10.1214/19-AOS1828](https://doi.org/10.1214/19-AOS1828)
- LECUÉ, G., LERASLE, M. and MATHIEU, T. (2020). Robust classification via MOM minimization. *Mach. Learn.* **109** 1635–1665. [MR4137195 https://doi.org/10.1007/s10994-019-05863-6](https://doi.org/10.1007/s10994-019-05863-6)
- LEVRARD, C. (2013). Fast rates for empirical vector quantization. *Electron. J. Stat.* **7** 1716–1746. [MR3080408 https://doi.org/10.1214/13-EJS822](https://doi.org/10.1214/13-EJS822)
- LEVRARD, C. (2015). Nonasymptotic bounds for vector quantization in Hilbert spaces. *Ann. Statist.* **43** 592–619. [MR3316191 https://doi.org/10.1214/14-AOS1293](https://doi.org/10.1214/14-AOS1293)
- LINDER, T. (2002). Learning-theoretic methods in vector quantization. In *Principles of Nonparametric Learning (Udine, 2001)*. *CISM Courses and Lect.* **434** 163–210. Springer, Vienna. [MR1987659](https://doi.org/10.1007/978-3-319-46379-7_1)
- LUGOSI, G. and MENDELSON, S. (2019a). Near-optimal mean estimators with respect to general norms. *Probab. Theory Related Fields* **175** 957–973. [MR4026610 https://doi.org/10.1007/s00440-019-00906-4](https://doi.org/10.1007/s00440-019-00906-4)
- LUGOSI, G. and MENDELSON, S. (2019b). Regularization, sparse recovery, and median-of-means tournaments. *Bernoulli* **25** 2075–2106. [MR3961241 https://doi.org/10.3150/18-BEJ1046](https://doi.org/10.3150/18-BEJ1046)
- LUGOSI, G. and MENDELSON, S. (2019c). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. [MR3909950 https://doi.org/10.1214/17-AOS1639](https://doi.org/10.1214/17-AOS1639)
- MAURER, A. (2016). A vector-contraction inequality for Rademacher complexities. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **9925** 3–17. Springer, Cham. [MR3590979 https://doi.org/10.1007/978-3-319-46379-7_1](https://doi.org/10.1007/978-3-319-46379-7_1)
- MAURER, A. and PONTIL, M. (2010). k -dimensional coding schemes in Hilbert spaces. *IEEE Trans. Inf. Theory* **56** 5839–5846. [MR2808936 https://doi.org/10.1109/TIT.2010.2069250](https://doi.org/10.1109/TIT.2010.2069250)
- MENDELSON, S. and ZHIVOTOVSKIY, N. (2020). Robust covariance estimation under L_4 - L_2 norm equivalence. *Ann. Statist.* **48** 1648–1664. [MR4124338 https://doi.org/10.1214/19-AOS1862](https://doi.org/10.1214/19-AOS1862)
- MINSKER, S. (2018). Uniform bounds for robust mean estimators. Preprint. Available at [arXiv:1812.03523](https://arxiv.org/abs/1812.03523).
- MINSKER, S. and MATHIEU, T. (2019). Excess risk bounds in robust empirical risk minimization. Preprint. Available at [arXiv:1910.07485](https://arxiv.org/abs/1910.07485).
- NARAYANAN, H. and MITTER, S. (2010). Sample complexity of testing the manifold hypothesis. In *Advances in Neural Information Processing Systems* 1786–1794.
- NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. A Wiley-Interscience Publication*. Wiley, New York. [MR0702836](https://doi.org/10.1002/97804702836)
- NG, M. K. (2000). A note on constrained k -means algorithms. *Pattern Recognit.* **33** 515–519.
- POLLARD, D. (1981). Strong consistency of k -means clustering. *Ann. Statist.* **9** 135–140. [MR0600539](https://doi.org/10.1214/aos/1176344953)
- SREBRO, N., SRIDHARAN, K. and TEWARI, A. (2010). Smoothness, low noise and fast rates. In *Advances in Neural Information Processing Systems* 2199–2207.
- TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **60**. Springer, Heidelberg. [MR3184689 https://doi.org/10.1007/978-3-642-54075-2](https://doi.org/10.1007/978-3-642-54075-2)
- TELGARSKY, M. J. and DASGUPTA, S. (2013). Moment-based uniform deviation bounds for k -means and friends. In *Advances in Neural Information Processing Systems* 2940–2948.
- VAPNIK, V. and CHERVONENKIS, A. (1974). *Theory of Pattern Recognition*. Nauka, Moscow.
- VERSHYNIN, R. (2016). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. [MR3837109 https://doi.org/10.1017/9781108231596](https://doi.org/10.1017/9781108231596)

ON THE RATE OF CONVERGENCE OF FULLY CONNECTED DEEP NEURAL NETWORK REGRESSION ESTIMATES

BY MICHAEL KOHLER^{*} AND SOPHIE LANGER[†]

Fachbereich Mathematik, Technische Universität Darmstadt, ^{*}kohler@mathematik.tu-darmstadt.de;
[†]langner@mathematik.tu-darmstadt.de

Recent results in nonparametric regression show that deep learning, that is, neural network estimates with many hidden layers, are able to circumvent the so-called curse of dimensionality in case that suitable restrictions on the structure of the regression function hold. One key feature of the neural networks used in these results is that their network architecture has a further constraint, namely the *network sparsity*. In this paper, we show that we can get similar results also for least squares estimates based on simple fully connected neural networks with ReLU activation functions. Here, either the number of neurons per hidden layer is fixed and the number of hidden layers tends to infinity suitably fast for sample size tending to infinity, or the number of hidden layers is bounded by some logarithmic factor in the sample size and the number of neurons per hidden layer tends to infinity suitably fast for sample size tending to infinity. The proof is based on new approximation results concerning deep neural networks.

REFERENCES

- [1] ANTHONY, M. and BARTLETT, P. L. (2009). *Neural Network Learning: Theoretical Foundations*, 1st ed. Cambridge Univ. Press, New York, NY, USA.
- [2] BARRON, A. R. (1991). Complexity regularization with application to artificial neural networks. In *Nonparametric Functional Estimation and Related Topics (Spetses, 1990)*. NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **335** 561–576. Kluwer Academic, Dordrecht. [MR1154352](#)
- [3] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. [MR1237720](#) <https://doi.org/10.1109/18.256500>
- [4] BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.
- [5] BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. [MR3953451](#) <https://doi.org/10.1214/18-AOS1747>
- [6] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. Springer, New York. [MR1383093](#) <https://doi.org/10.1007/978-1-4612-0711-5>
- [7] DEVROYE, L. P. and WAGNER, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239. [MR0560725](#)
- [8] ELDAN, R. and SHAMIR, O. (2016). The power of depth for feedforward neural networks. In *29th Annual Conference on Learning Theory* (V. Feldman, A. Rakhlin and O. Shamir, eds.). *Proc. Mach. Learn. Res. (PMLR)* **49** 907–940. PMLR.
- [9] EVCI, U., PEDREGOSA, F., GOMEZ, A. and ELSÉN, E. (2019). The difficulty of training sparse neural networks. CoRR, abs/1906.10732.
- [10] FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. [MR0650892](#)
- [11] GROHS, P., PEREKRESTENKO, D., ELBRÄCHTER, D. and BÖLCSKEI, H. (2019). Deep neural network approximation theory. *IEEE Trans. Inf. Theory*.
- [12] GYÖRFI, L., KOHLER, M., KRZYZAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer, New York. [MR1920390](#) <https://doi.org/10.1007/b97848>

MSC2020 subject classifications. Primary 62G08; secondary 41A25, 82C32.

Key words and phrases. Curse of dimensionality, deep learning, neural networks, nonparametric regression, rate of convergence.

- [13] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171](#) <https://doi.org/10.1214/aos/1176349020>
- [14] HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. [MR1134488](#)
- [15] HAYKIN, S. (1998). *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice Hall PTR, Upper Saddle River, NJ, USA.
- [16] HERTZ, J., KROGH, A. and PALMER, R. G. (1991). *Introduction to the Theory of Neural Computation. Santa Fe Institute Studies in the Sciences of Complexity. Lecture Notes, I.* Addison-Wesley, Redwood City, CA. With forewords by Jack Cowan and Christof Koch. [MR1096298](#)
- [17] HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. [MR2382659](#) <https://doi.org/10.1214/0090536070000000415>
- [18] IMAIZUMI, M. and FUKUMIZU, K. (2019). Deep neural networks learn non-smooth functions effectively. *Proc. Mach. Learn. Res.* **89** 869–878.
- [19] KOHLER, M. and KRZYŻAK, A. (2005). Adaptive regression estimation with multilayer feedforward neural networks. *J. Nonparametr. Stat.* **17** 891–913. [MR2192165](#) <https://doi.org/10.1080/10485250500309608>
- [20] KOHLER, M. and KRZYŻAK, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Trans. Inf. Theory* **63** 1620–1630. [MR3625984](#) <https://doi.org/10.1109/TIT.2016.2634401>
- [21] KOHLER, M. and LANGER, S. (2021). Supplement to “On the rate of convergence of fully connected deep neural network regression estimates.” <https://doi.org/10.1214/20-AOS2034SUPPA>, <https://doi.org/10.1214/20-AOS2034SUPPB>
- [22] KONG, E. and XIA, Y. (2007). Variable selection for the single-index model. *Biometrika* **94** 217–229. [MR2367831](#) <https://doi.org/10.1093/biomet/asm008>
- [23] LIU, Z., SUN, M., ZHOU, T., HUANG, G. and DARRELL, T. (2018). Rethinking the value of network pruning. CoRR, abs/1810.05270.
- [24] LU, J., SHEN, Z., YANG, H. and ZHANG, S. (2020). Deep network approximation for smooth functions. CoRR, [arXiv:2001.03040](#).
- [25] MCCAFFREY, D. F. and GALLANT, A. R. (1994). Convergence rates for single hidden layer feedforward networks. *Neural Netw.* **7** 147–158. [https://doi.org/10.1016/0893-6080\(94\)90063-9](https://doi.org/10.1016/0893-6080(94)90063-9)
- [26] MHASKAR, H. N. and POGGIO, T. (2016). Deep vs. shallow networks: An approximation theory perspective. *Anal. Appl. (Singap.)* **14** 829–848. [MR3564936](#) <https://doi.org/10.1142/S0219530516400042>
- [27] OHN, I. and KIM, Y. (2019). Smooth function approximation by deep neural networks with general activation functions. *Entropy* **21** Paper No. 627. [MR3988437](#) <https://doi.org/10.3390/e21070627>
- [28] OONO, K. and SUZUKI, T. (2019). Approximation and nonparametric estimation of ResNet-type convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.) **97** 4922–4931. Proc. Mach. Learn. Res. (PMLR), Long Beach, CA, USA.
- [29] RIPLEY, B. D. and HJORT, N. L. (1995). *Pattern Recognition and Neural Networks*, 1st ed. Cambridge Univ. Press, New York, NY, USA.
- [30] SCHMIDHUBER, J. (2015). Deep learning in neural networks: An overview. *Neural Netw.* **61** 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [31] SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. [MR4134774](#) <https://doi.org/10.1214/19-AOS1875>
- [32] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. [MR0673642](#)
- [33] STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. [MR0790566](#) <https://doi.org/10.1214/aos/1176349548>
- [34] STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. With discussion by Andreas Buja and Trevor Hastie and a rejoinder by the author. [MR1272079](#) <https://doi.org/10.1214/aos/1176325361>
- [35] SUZUKI, T. (2019). Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality. In *International Conference on Learning Representations*.
- [36] YAROTSKY, D. (2018). Optimal approximation of continuous functions by very deep ReLU networks. *COLT* **75** 639–649.
- [37] YAROTSKY, D. and ZHEVNERCHUK, A. (2019). The phase diagram of approximation rates for deep neural networks. CoRR, abs/1906.09477.
- [38] YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. [MR1951258](#) <https://doi.org/10.1198/016214502388618861>

INFINITE-DIMENSIONAL GRADIENT-BASED DESCENT FOR ALPHA-DIVERGENCE MINIMISATION

BY KAMÉLIA DAUDEL^{1,*}, RANDAL DOUC² AND FRANÇOIS PORTIER^{1,†}

¹LTCI, Télécom Paris, Institut Polytechnique de Paris, *kamelia.daudel@gmail.com; †francois.portier@telecom-paris.fr

²SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, randal.douc@telecom-sudparis.eu

This paper introduces the (α, Γ) -descent, an iterative algorithm which operates on measures and performs α -divergence minimisation in a Bayesian framework. This gradient-based procedure extends the commonly-used variational approximation by adding a prior on the variational parameters in the form of a measure. We prove that for a rich family of functions Γ , this algorithm leads at each step to a systematic decrease in the α -divergence and derive convergence results. Our framework recovers the Entropic Mirror Descent algorithm and provides an alternative algorithm that we call the Power Descent. Moreover, in its stochastic formulation, the (α, Γ) -descent allows to optimise the mixture weights of any given mixture model without any information on the underlying distribution of the variational parameters. This renders our method compatible with many choices of parameters updates and applicable to a wide range of Machine Learning tasks. We demonstrate empirically on both toy and real-world examples the benefit of using the Power Descent and going beyond the Entropic Mirror Descent framework, which fails as the dimension grows.

REFERENCES

- [1] BAMLER, R., ZHANG, C., OPPER, M. and MANDT, S. (2017). Perturbative black box variational inference. In *Advances in Neural Information Processing Systems* 30 (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) 5079–5088. Curran Associates.
- [2] BEAL, M. J. (2003). Variational algorithms for approximate Bayesian inference. PhD Thesis.
- [3] BECK, A. and TEOULLE, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **31** 167–175. [MR1967286](#) [https://doi.org/10.1016/S0167-6377\(02\)00231-6](https://doi.org/10.1016/S0167-6377(02)00231-6)
- [4] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. [MR3671776](#) <https://doi.org/10.1080/01621459.2017.1285773>
- [5] BLEI, D. M., NG, A. Y. and JORDAN, M. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research* **3** 993–1022.
- [6] BOTTOU, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* (Y. Lechevallier and G. Saporta, eds.) 177–186. Physica-Verlag/Springer, Heidelberg. [MR3362066](#)
- [7] BUBECK, S. (2015). Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.* **8** 231–357. <https://doi.org/10.1561/22000000050>
- [8] CHÉRIEF-ABDELLATIF, B.-E., ALQUIER, P. and KHAN, M. E. (2019). A generalization bound for online variational inference. In *Proceedings of the 29th International Conference on Machine Learning* **101** 662–677.
- [9] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. [MR2153989](#) <https://doi.org/10.1214/009053604000000698>
- [10] CICHOCKI, A. and AMARI, S. (2010). Families of alpha- beta- and gamma-divergences: Flexible and robust measures of similarities. *Entropy* **12** 1532–1568. [MR2659408](#) <https://doi.org/10.3390/e12061532>

MSC2020 subject classifications. Primary 62F15; secondary 62F30, 62F35, 62G07, 62L99.

Key words and phrases. Variational inference, alpha-divergence, Kullback–Leibler divergence, Mirror Descent.

- [11] CICHOCKI, A., CRUCES, S. and AMARI, S.-I. (2011). Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy* **13** 134–170. <https://doi.org/10.3390/e13010134>
- [12] DAUDEL, K., DOUC, R. and PORTIER, F. (2021). Supplement to “Infinite-dimensional gradient-based descent for alpha-divergence minimisation.” <https://doi.org/10.1214/20-AOS2035SUPP>
- [13] DEHAENE, G. and BARTHELMÉ, S. (2018). Expectation propagation in the large data limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 199–217. [MR3744718 https://doi.org/10.1111/rssb.12241](https://doi.org/10.1111/rssb.12241)
- [14] DELYON, B. and PORTIER, F. (2019). Safe and adaptive importance sampling: A mixture approach. Available online: <https://arxiv.org/abs/1903.08507> (accessed on 20 March 2020).
- [15] DIENG, A. B., TRAN, D., RANGANATH, R., PAISLEY, J. and BLEI, D. (2017). Variational inference via χ upper bound minimization. In *Advances in Neural Information Processing Systems* 30 (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) 2732–2741. Curran Associates.
- [16] DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2007). Convergence of adaptive mixtures of importance sampling schemes. *Ann. Statist.* **35** 420–448. [MR2332281 https://doi.org/10.1214/009053606000001154](https://doi.org/10.1214/009053606000001154)
- [17] GERSHMAN, S., HOFFMAN, M. D. and BLEI, D. M. (2012). Nonparametric variational inference. In *Proceedings of the 29th International Conference on Machine Learning*.
- [18] HELLINGER, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J. Reine Angew. Math.* **136** 210–271. [MR1580780 https://doi.org/10.1515/crll.1909.136.210](https://doi.org/10.1515/crll.1909.136.210)
- [19] HERNANDEZ-LOBATO, J., LI, Y., ROWLAND, M., BUI, T., HERNANDEZ-LOBATO, D. and TURNER, R. (2016). Black-box alpha divergence minimization. In *Proceedings of the 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 1511–1520. PMLR, New York, New York, USA.
- [20] HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. [MR3081926](https://doi.org/10.1214/1303081926)
- [21] HSIEH, Y.-P., LIU, C. and CEVHER, V. (2019). Finding mixed Nash equilibria of generative adversarial networks. In *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 2810–2819. PMLR, Long Beach, CA.
- [22] JAAKKOLA, T. S. and JORDAN, M. I. (1998). Improving the mean field approximation via the use of mixture distributions. In *Learning in Graphical Models* (Jordan, M. I., ed.). *NATO ASI Series (Series D: Behavioural and Social Sciences)* **89**. Springer.
- [23] JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233. <https://doi.org/10.1023/A:1007665907178>
- [24] KLOEK, T. and VAN DIJK, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* 1–19.
- [25] KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. [MR0039968 https://doi.org/10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
- [26] LI, Y., HERNÁNDEZ-LOBATO, J. M. and TURNER, R. E. (2015). Stochastic expectation propagation. In *Advances in Neural Information Processing Systems* 28 (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, eds.) 2323–2331. Curran Associates.
- [27] LI, Y. and TURNER, R. E. (2016). Rényi divergence variational inference. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 1073–1081. Curran Associates.
- [28] LINDSAY, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.* **22** 1081–1114. [MR1292557 https://doi.org/10.1214/aos/1176325512](https://doi.org/10.1214/aos/1176325512)
- [29] MINKA, T. (2004). Power EP. Technical Report No. MSR-TR-2004-149.
- [30] MINKA, T. (2005). Divergence measures and message passing. Technical Report No. MSR-TR-2005-173.
- [31] MINKA, T. P. (2001). Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. *UAI’01* 362–369. Morgan Kaufmann Publishers Inc., San Francisco, CA.
- [32] MORIMOTO, T. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* 85–108.
- [33] MORIMOTO, T. (1963). Markov processes and the H -theorem. *J. Phys. Soc. Jpn.* **18** 328–331. [MR0167200 https://doi.org/10.1143/JPSJ.18.328](https://doi.org/10.1143/JPSJ.18.328)
- [34] NEMIROVSKI, A. (2004). Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** 229–251. [MR2112984 https://doi.org/10.1137/S1052623403425629](https://doi.org/10.1137/S1052623403425629)

- [35] NEMIROVSKI, A., JUDITSKY, A., LAN, G. and SHAPIRO, A. (2008). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. [MR2486041](#) <https://doi.org/10.1137/070704277>
- [36] OH, M.-S. and BERGER, J. O. (1992). Adaptive importance sampling in Monte Carlo integration. *J. Stat. Comput. Simul.* **41** 143–168. [MR1276184](#) <https://doi.org/10.1080/00949659208810398>
- [37] OPPER, M. and WINTHER, O. (2000). Gaussian processes for classification: Mean-field algorithms. *Neural Comput.* **12** 2655–2684. <https://doi.org/10.1162/089976600300014881>
- [38] PAISLEY, J., BLEI, D. and JORDAN, M. (2012). Variational Bayesian inference with stochastic search. In *Proceedings of the 29th International Conference on Machine Learning* 1363–1370.
- [39] RANGANATH, R., GERRISH, S. and BLEI, D. (2014). Black box variational inference. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (S. Kaski and J. Corander, eds.). *Proceedings of Machine Learning Research* **33** 814–822. PMLR, Reykjavik, Iceland.
- [40] RANGANATH, R., TRAN, D. and BLEI, D. (2016). Hierarchical variational models. In *Proceedings of the 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 324–333. PMLR, New York, New York, USA.
- [41] RÉNYI, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 547–561. Univ. California Press, Berkeley, CA. [MR0132570](#)
- [42] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. [MR0042668](#) <https://doi.org/10.1214/aoms/1177729586>
- [43] SASON, I. (2018). On f -divergences: Integral representations, local behavior, and inequalities. *Entropy* **20** Paper No. 383, 32. [MR3862573](#) <https://doi.org/10.3390/e20050383>
- [44] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. [MR0673642](#)
- [45] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **60** 3797–3820. [MR3225930](#) <https://doi.org/10.1109/TIT.2014.2320500>
- [46] WANG, D., LIU, H. and LIU, Q. (2018). Variational inference with tail-adaptive f -divergence. In *Advances in Neural Information Processing Systems* 31 (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) 5737–5747. Curran Associates.
- [47] YIN, M. and ZHOU, M. (2018). Semi-implicit variational inference. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 5660–5669. PMLR, Stockholmsmässan, Stockholm Sweden.
- [48] ZHANG, C., BUTEPAGE, J., KJELLSTROM, H. and MANDT, S. (2019). Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 2008–2026. <https://doi.org/10.1109/TPAMI.2018.2889774>
- [49] ZHU, H. and ROHWER, R. (1995). Bayesian invariant measurements of generalization. *Neural Process. Lett.* **2** 28–31. <https://doi.org/10.1007/BF02309013>
- [50] ZHU, H. and ROHWER, R. (1995). Information geometric measurements of generalisation. Technical Report No. NCRG/4350.

MONITORING FOR A CHANGE POINT IN A SEQUENCE OF DISTRIBUTIONS

BY LAJOS HORVÁTH¹, PIOTR KOKOSZKA² AND SHIXUAN WANG³

¹*Department of Mathematics, University of Utah, horvath@math.utah.edu*

²*Department of Statistics, Colorado State University, piotr.kokoszka@colostate.edu*

³*Department of Economics, University of Reading, shixuan.wang@reading.ac.uk*

We propose a method for the detection of a change point in a sequence $\{F_i\}$ of distributions, which are available through a large number of observations at each $i \geq 1$. Under the null hypothesis, the distributions F_i are equal. Under the alternative hypothesis, there is a change point $i^* > 1$, such that $F_i = G$ for $i \geq i^*$ and some unknown distribution G , which is not equal to F_1 . The change point, if it exists, is unknown, and the distributions before and after the potential change point are unknown. The decision about the existence of a change point is made sequentially, as new data arrive. At each time i , the count of observations, N , can increase to infinity. The detection procedure is based on a weighted version of the Wasserstein distance. Its asymptotic and finite sample validity is established. Its performance is illustrated by an application to returns on stocks in the S&P 500 index.

REFERENCES

- ANG, A., HODRICK, R. J., XING, Y. and ZHANG, X. (2006). The cross-section of volatility and expected returns. *J. Finance* **61** 259–299.
- BARDSLEY, P., HORVÁTH, L., KOKOSZKA, P. and YOUNG, G. (2017). Change point tests in functional factor models with application to yield curves. *Econom. J.* **20** 86–117. [MR3636962 https://doi.org/10.1111/ectj.12075](https://doi.org/10.1111/ectj.12075)
- BARIGOZZI, M., CHO, H. and FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometrics* **206** 187–225. [MR3840788 https://doi.org/10.1016/j.jeconom.2018.05.003](https://doi.org/10.1016/j.jeconom.2018.05.003)
- BERTHET, P. and FORT, J. C. (2020). Exact rate of convergence of the expected W_2 distance between the empirical and true Gaussian distribution. *Electron. J. Probab.* **25** Paper No. 12, 16. [MR4059190 https://doi.org/10.1214/19-ejp410](https://doi.org/10.1214/19-ejp410)
- BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. [MR0898871 https://doi.org/10.1017/CBO9780511721434](https://doi.org/10.1017/CBO9780511721434)
- BLATTBERG, R. C. and GONEDES, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *J. Bus.* **47** 244–280.
- CARHART, M. M. (1997). On persistence in mutual fund performance. *J. Finance* **52** 57–82.
- CHEN, C. (2005). *Four Essays on the Equity Market: Decimalization, Volatility, Indexation and Cross-Sectional Returns*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, Univ. Illinois, Chicago, IL. [MR2708219](https://doi.org/10.1111/ectj.12075)
- CHEN, Y., HÄRDLE, W. K. and PIGORSCH, U. (2010). Localized realized volatility modeling. *J. Amer. Statist. Assoc.* **105** 1376–1393. [MR2796557 https://doi.org/10.1198/jasa.2010.ap09039](https://doi.org/10.1198/jasa.2010.ap09039)
- CHEN, Y., WANG, T. and SAMWORTH, R. J. (2020). High-dimensional, multiscale online changepoint detection. Technical Report, Univ. Cambridge. [arXiv:2003.03668](https://arxiv.org/abs/2003.03668).
- CHU, C.-S. J., STINCHCOMBE, M. and WHITE, H. (1996). Monitoring structural change. *Econometrica* **64** 1045–1065.
- CONT, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* **1** 223–236.

MSC2020 subject classifications. Primary 62G30, 62L10; secondary 62G10, 62G20.

Key words and phrases. Change point detection, Empirical quantile function, Sequential monitoring, Wasserstein distance.

- CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, Chichester. With a foreword by David Kendall. **MR1215046**
- CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882–894. **MR0501290**
- CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Probability and Mathematical Statistics. Academic Press, New York. **MR0666546**
- DEL BARRIO, E., GINÉ, E. and MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009–1071. **MR1698999** <https://doi.org/10.1214/aop/1022677394>
- DEL BARRIO, E., GINÉ, E. and UTZET, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. **MR2121458** <https://doi.org/10.3150/bj/1110228245>
- DEL BARRIO, E., CUESTA-ALBERTOS, J. A., MATRÁN, C. and RODRÍGUEZ-RODRÍGUEZ, J. M. (1999). Tests of goodness of fit based on the L_2 -Wasserstein distance. *Ann. Statist.* **27** 1230–1239. **MR1740113** <https://doi.org/10.1214/aos/1017938923>
- DUBEY, P. and MÜLLER, H.-G. (2020). Fréchet change-point detection. *Ann. Statist.* **48** 3312–3335. **MR4185810** <https://doi.org/10.1214/19-AOS1930>
- FAMA, E. (1965). The behavior of stock-market prices. *J. Bus.* **38** 34–105.
- FAMA, E. and FRENCH, K. (1993). Common risk factors in the returns on bonds and stocks. *J. Financ. Econ.* **33** 3–56.
- FAMA, E. and FRENCH, K. (2015). A five-factor asset pricing model. *J. Financ. Econ.* **116** 1–22.
- GRANGER, C. W. J. and DING, Z. (1995). Some properties of absolute return: An alternative measure of risk. *Ann. Econ. Statist.* **40** 67–91. **MR1476513** <https://doi.org/10.2307/20076016>
- GROMENKO, O., KOKOSZKA, P. and REIMHERR, M. (2017). Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 29–50. **MR3597963** <https://doi.org/10.1111/rssb.12156>
- HERSKOVIC, B., KELLY, B., LUSTIG, H. and VAN NIEUWERBURGH, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *J. Financ. Econ.* **119**.
- HORVÁTH, L., KOKOSZKA, P. and WANG, S. (2021). Supplement to “Monitoring for a change point in a sequence of distributions.” <https://doi.org/10.1214/20-AOS2036SUPP>
- JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. **MR3405600** <https://doi.org/10.1214/15-AOS1347>
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. **MR0375412** <https://doi.org/10.1007/BF00533093>
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. **MR0402883** <https://doi.org/10.1007/BF00532688>
- LI, J. and JIN, B. (2018). Multi-threshold accelerated failure time model. *Ann. Statist.* **46** 2657–2682. **MR3851751** <https://doi.org/10.1214/17-AOS1632>
- MANDELBROT, B. (1997). *The Variation of Certain Speculative Prices*. Springer, Berlin.
- MITTNIK, S. and RACHEV, S. T. (1993). Modeling asset returns with alternative stable distributions. *Econometric Rev.* **12** 261–389. With comments by P. C. B. Phillips, F. X. Diebold and R. T. Baillie and a reply by the authors. **MR1249625** <https://doi.org/10.1080/07474939308800266>
- MORGENSON, G. and HARVEY, C. (2002). *The New York Times Dictionary of Money and Investing: The Essential A-to-Z Guide to the Language of the New Market*. Macmillan, New York.
- PADILLA, O. H. M., ATHEY, A., REINHART, A. and SCOTT, J. G. (2019). Sequential nonparametric tests for a change in distribution: An application to detecting radiological anomalies. *J. Amer. Statist. Assoc.* **114** 514–528. **MR3963159** <https://doi.org/10.1080/01621459.2018.1476245>
- PANARETOS, V. M. and ZEMEL, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44** 771–812. **MR3476617** <https://doi.org/10.1214/15-AOS1387>
- PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. **MR3939527** <https://doi.org/10.1146/annurev-statistics-030718-104938>
- PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. **MR3449766** <https://doi.org/10.1214/15-AOS1363>
- POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. **MR0773162** <https://doi.org/10.1214/aos/1176346587>
- 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** <https://doi.org/10.1214/09-AOS775>
- PRAETZ, P. (1972). The distribution of share price changes. *J. Bus.* **45** 49–55.

- SHARPE, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19** 425–442.
- SKOROKHOD, A. V. (1965). *Studies in the Theory of Random Processes*. Addison-Wesley, Reading, MA. Translated from the Russian by Scripta Technica, Inc. [MR0185620](#)
- WANG, J. (2012). *A State Space Model Approach to Functional Time Series and Time Series Driven by Differential Equations*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, Rutgers Univ., New Brunswick, NJ. [MR3152378](#)
- XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. [MR3099117](#) <https://doi.org/10.1214/13-AOS1094>
- YAKIR, B. (1997). A note on optimal detection of a change in distribution. *Ann. Statist.* **25** 2117–2126. [MR1474086](#) <https://doi.org/10.1214/aos/1069362390>

WHAT IS RESOLUTION? A STATISTICAL MINIMAX TESTING PERSPECTIVE ON SUPERRESOLUTION MICROSCOPY

BY GYTIS KULAITIS^{1,*}, AXEL MUNK^{1,†} AND FRANK WERNER²

¹*Institute for Mathematical Stochastics, University of Göttingen, * gytis.kulaitis@mathematik.uni-goettingen.de; munk@math.uni-goettingen.de*

²*Institute of Mathematics, University of Würzburg, frank.werner@mathematik.uni-wuerzburg.de*

As a general rule of thumb the resolution of a light microscope (i.e., the ability to discern objects) is predominantly described by the full width at half maximum (FWHM) of its point spread function (psf)—the diameter of the blurring density at half of its maximum. Classical wave optics suggests a linear relationship between FWHM and resolution also manifested in the well-known Abbe and Rayleigh criteria, dating back to the end of the 19th century. However, during the last two decades conventional light microscopy has undergone a shift from microscopic scales to nanoscales. This increase in resolution comes with the need to incorporate the random nature of observations (light photons) and challenges the classical view of discernability, as we argue in this paper. Instead, we suggest a statistical description of resolution obtained from such random data. Our notion of discernability is based on statistical testing whether one or two objects with the same total intensity are present. For Poisson measurements, we get linear dependence of the (minimax) detection boundary on the FWHM, whereas for a homogeneous Gaussian model the dependence of resolution is nonlinear. Hence, at small physical scales modeling by homogeneous gaussians is inadequate, although often implicitly assumed in many reconstruction algorithms. In contrast, the Poisson model and its variance stabilized Gaussian approximation seem to provide a statistically sound description of resolution at the nanoscale. Our theory is also applicable to other imaging setups, such as telescopes.

REFERENCES

- [1] ABBE, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. Mikrosk. Anat.* **9** 413–418. <https://doi.org/10.1007/BF02956173>
- [2] ACUÑA, C. O. and HOROWITZ, J. (1997). A statistical approach to the resolution of point sources. *J. Appl. Stat.* **24** 421–435. [MR1491309 https://doi.org/10.1080/02664769723620](https://doi.org/10.1080/02664769723620)
- [3] AIRY, G. B. (1835). On the diffraction of an object-glass with circular aperture. *Trans. Camb. Philos. Soc.* **5** 283–291.
- [4] ASH, E. A. and NICHOLLS, G. (1972). Super-resolution aperture scanning microscope. *Nature* **237** 510–512. <https://doi.org/10.1038/237510a0>
- [5] ASPELMEIER, T., EGNER, A. and MUNK, A. (2015). Modern statistical challenges in high-resolution fluorescence microscopy. *Annu. Rev. Stat. Appl.* **2** 163–202. <https://doi.org/10.1007/s00344-014-9442-0>
- [6] BANTERLE, N., BUI, K. H., LEMKE, E. A. and BECK, M. (2013). Fourier ring correlation as a resolution criterion for super-resolution microscopy. *J. Struct. Biol.* **183** 363–367. <https://doi.org/10.1016/j.jsb.2013.05.004>
- [7] BERTERO, M., BOCCACCI, P., DESIDERÀ, G. and VICIDOMINI, G. (2009). Image deblurring with Poisson data: From cells to galaxies. *Inverse Probl.* **25** 123006. [MR2565572 https://doi.org/10.1088/0266-5611/25/12/123006](https://doi.org/10.1088/0266-5611/25/12/123006)
- [8] BETZIG, E., DAVIDSON, M. W., PATTERSON, G. H., LIPPINCOTT-SCHWARTZ, J., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S. and HESS, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313** 1642–1645.

MSC2020 subject classifications. Primary 91B06, 94A12; secondary 60F05.

Key words and phrases. Microscopy, (super)resolution, nanoscopy, minimax, detection boundary, equivalence of experiments.

- [9] BORN, M. and WOLF, E. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. Cambridge Univ. Press, Cambridge.
- [10] CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2013). Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19** 1229–1254. [MR3132912 https://doi.org/10.1007/s00041-013-9292-3](https://doi.org/10.1007/s00041-013-9292-3)
- [11] CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956. [MR3193963 https://doi.org/10.1002/cpa.21455](https://doi.org/10.1002/cpa.21455)
- [12] COURJON, D. (2003). *Near-Field Microscopy and Near-Field Optics*. Imperial College Press, London.
- [13] CREMER, C. and MASTERS, B. R. (2013). Resolution enhancement techniques in microscopy. *Eur. Phys. J. H* **38** 281–344. <https://doi.org/10.1140/epjh/e2012-20060-1>
- [14] DEN DEKKER, A. J. and VAN DEN BOS, A. (1997). Resolution: A survey. *J. Opt. Soc. Amer. A* **14** 547–557. <https://doi.org/10.1364/JOSAA.14.000547>
- [15] DONOHO, D. L. (1992). Superresolution via sparsity constraints. *SIAM J. Math. Anal.* **23** 1309–1331. [MR1177792 https://doi.org/10.1137/0523074](https://doi.org/10.1137/0523074)
- [16] DU, C. and KOU, S. C. (2020). Statistical methodology in single-molecule experiments. *Statist. Sci.* **35** 75–91. [MR4071359 https://doi.org/10.1214/19-STS752](https://doi.org/10.1214/19-STS752)
- [17] DÜRIG, U., POHL, D. W. and ROHNER, F. (1986). Near-field optical-scanning microscopy. *J. Appl. Phys.* **59** 3318–3327. <https://doi.org/10.1063/1.336848>
- [18] EGNER, A., GEISLER, C. and SIEGMUND, R. (2020). Basic knowledge in STED nanoscopy. In *Nanoscale Photonic Imaging* (T. Salditt, A. Egner and R. Luke, eds.) Springer, Berlin.
- [19] EGNER, A., GEISLER, C., VON MIDDENDORFF, C., BOCK, H., WENZEL, D., MEDDA, R., ANDRESEN, M., STIEL, A. C., JAKOBS, S. et al. (2007). Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophys. J.* **93** 3285–3290. <https://doi.org/10.1529/biophysj.107.112201>
- [20] EHRENBERG, M. (2014). Super-resolved fluorescence microscopy. Scientific Background on Nobel Prize in Chemistry.
- [21] FERNANDEZ-GRANDA, C. (2015). Super-resolution of point sources via convex programming. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015* 41–44.
- [22] FERREIRA DA COSTA, M. and CHI, Y. (2020). On the stable resolution limit of total variation regularization for spike deconvolution. *IEEE Trans. Inf. Theory* **66** 7237–7252. [MR4173638](https://doi.org/10.1109/TIT.2020.3000000)
- [23] GOODMAN, J. W. (1985). *Statistical Optics*. Wiley-Interscience, New York.
- [24] HARRIS, J. L. (1964). Resolving power and decision theory. *J. Opt. Soc. Amer.* **54** 606–611. <https://doi.org/10.1364/JOSA.54.000606>
- [25] HEILEMANN, M., VAN DE LINDE, S., SCHÜTTPELZ, M., KASPER, R., SEEFELDT, B., MUKHERJEE, A., TINNEFELD, P. and SAUER, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew. Chem.* **47** 6172–6176. <https://doi.org/10.1002/anie.200802376>
- [26] HEINTZMANN, R. and FICZ, G. (2013). Breaking the resolution limit in light microscopy. *Methods Cell Biol.* **114** 525–544. <https://doi.org/10.1016/B978-0-12-407761-4.00022-1>
- [27] HELL, S. W. (2007). Far-field optical nanoscopy. *Science* **316** 1153–1158. <https://doi.org/10.1126/science.1137395>
- [28] HELL, S. W. and WICHMANN, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19** 780–782. <https://doi.org/10.1364/OL.19.000780>
- [29] HELSTROM, C. (1964). The detection and resolution of optical signals. *IEEE Trans. Inf. Theory* **10** 275–287.
- [30] HELSTROM, C. (1965). Correction to ‘The detection and resolution of optical signals’. *IEEE Trans. Inf. Theory* **11** 125. <https://doi.org/10.1109/TIT.1965.1053725>
- [31] HELSTROM, C. W. (1973). Resolution of point sources of light as analyzed by quantum detection theory. *IEEE Trans. Inf. Theory* **19** 389–398. <https://doi.org/10.1109/TIT.1973.1055052>
- [32] HESS, S. T., GIRIRAJAN, T. P. K. and MASON, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91** 4258–4272. <https://doi.org/10.1529/biophysj.106.091116>
- [33] HOHAGE, T. and WERNER, F. (2016). Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms. *Inverse Probl.* **32** 093001. [MR3543331 https://doi.org/10.1088/0266-5611/32/9/093001](https://doi.org/10.1088/0266-5611/32/9/093001)
- [34] HOUSTON, W. V. (1927). A compound interferometer for fine structure work. *Phys. Rev.* **29** 478–484. <https://doi.org/10.1103/PhysRev.29.478>
- [35] HUANG, B., BATES, M. and ZHUANG, X. (2009). Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* **78** 993–1016. <https://doi.org/10.1146/annurev.biochem.77.061906.092014>

- [36] KLAR, T. A., JAKOBS, S., DYBA, M., EGNER, A. and HELL, S. W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl. Acad. Sci. USA* **97** 8206–8210. <https://doi.org/10.1073/pnas.97.15.8206>
- [37] KULAITIS, G., MUNK, A. and WERNER, F. (2021). Supplement to “What is resolution? A statistical minimax testing perspective on superresolution microscopy.” <https://doi.org/10.1214/20-AOS2037SUPP>
- [38] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. [MR2135927](https://doi.org/10.1007/978-1-4939-9826-2)
- [39] LEONHARDT, U. (2010). *Essential Quantum Optics: From Quantum Measurements to Black Holes*. Cambridge Univ. Press, Cambridge. <https://doi.org/10.1017/CBO9780511806117>
- [40] LU, X.-M., KROVI, H., NAIR, R., GUHA, S. and SHAPIRO, J. H. (2018). Quantum-optimal detection of one-versus-two incoherent optical sources with arbitrary separation. *npj Quantum Inf.* **4**. <https://doi.org/10.1038/s41534-018-0114-y>
- [41] MIDDLETON, D. (1953). Statistical criteria for the detection of pulsed carriers in noise. *I. J. Appl. Phys.* **24** 371–378. [MR0055652](https://doi.org/10.1063/jrmp.1.2.371)
- [42] MILANFAR, P. and SHAKOURI, A. (2002). A statistical analysis of diffraction-limited imaging. In *IEEE International Conference on Image Processing* 864–867.
- [43] MINSKY, M. (1961). Microscopy apparatus. US Patent 3013467, filed 7 November 1957, granted 19 December 1961.
- [44] MORGENSHTERN, V. I. and CANDÈS, E. J. (2016). Super-resolution of positive sources: The discrete setup. *SIAM J. Imaging Sci.* **9** 412–444. [MR3477315](https://doi.org/10.1137/15M1016552) <https://doi.org/10.1137/15M1016552>
- [45] MUNK, A., STAUDT, T. and WERNER, F. (2020). Statistical foundations of nanoscale photonic imaging. In *Nanoscale Photonic Imaging* (T. Salditt, A. Egner and R. Luke, eds.) Springer, Berlin.
- [46] NAIR, R. and TSANG, M. (2016). Far-field superresolution of thermal electromagnetic sources at the quantum limit. *Phys. Rev. Lett.* **117** 190801. <https://doi.org/10.1103/PhysRevLett.117.190801>
- [47] ORFANIDIS, S. J. (2016). *Electromagnetic Waves and Antennas*. Rutgers Univ. Available at <http://ecceweb1.rutgers.edu/~orfanidi/ewa/>.
- [48] OSHIKANE, Y., KATAOKA, T., OKUDA, M., HARA, S., INOUE, H. and NAKANO, M. (2007). Observation of nanostructure by scanning near-field optical microscope with small sphere probe. *Sci. Technol. Adv. Mater.* **8** 181–185. <https://doi.org/10.1016/j.stam.2007.02.013>
- [49] PAWLEY, J. B. (2006). *Handbook of Biological Confocal Microscopy*, 3rd ed. Springer, New York.
- [50] RAY, K. and SCHMIDT-HIEBER, J. (2018). The Le Cam distance between density estimation, Poisson processes and Gaussian white noise. *Math. Stat. Learn.* **1** 101–170. [MR4050245](https://doi.org/10.4171/msl/1-2-1) <https://doi.org/10.4171/msl/1-2-1>
- [51] REIFFEN, B. and SHERMAN, H. (1963). An optimum demodulator for Poisson processes: Photon source detectors. *IEEE Trans. Inf. Theory* **51** 1316–1320.
- [52] RITTWEGER, E., HAN, K. Y., IRVINE, S. E., EGGELING, C. and HELL, S. W. (2009). STED microscopy reveals crystal colour centres with nanometric resolution. *Nat. Photonics* **3** 144–147.
- [53] RUST, M. J., BATES, M. and ZHUANG, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3** 793–795. <https://doi.org/10.1038/nmeth929>
- [54] SHAHRAM, M. (2005). Statistical and information-theoretic analysis of resolution in imaging and array processing. Ph.D. thesis, Univ. California, Santa Cruz.
- [55] SHAHRAM, M. and MILANFAR, P. (2004). Imaging below the diffraction limit: A statistical analysis. *IEEE Trans. Image Process.* **13** 677–689. <https://doi.org/10.1109/TIP.2004.826096>
- [56] SHAHRAM, M. and MILANFAR, P. (2006). Statistical and information-theoretic analysis of resolution in imaging. *IEEE Trans. Inf. Theory* **52** 3411–3437. [MR2242357](https://doi.org/10.1109/TIT.2006.878180) <https://doi.org/10.1109/TIT.2006.878180>
- [57] SMITH, S. T. (2005). Statistical resolution limits and the complexified Cramér–Rao bound. *IEEE Trans. Signal Process.* **53** 1597–1609. [MR2148598](https://doi.org/10.1109/TSP.2005.845426) <https://doi.org/10.1109/TSP.2005.845426>
- [58] STAUDT, T., ASPELMEIER, T., LAITENBERGER, O., GEISLER, C., EGNER, A. and MUNK, A. (2020). Statistical molecule counting in super-resolution fluorescence microscopy: Towards quantitative nanoscopy. *Statist. Sci.* **35** 92–111. [MR4071360](https://doi.org/10.1214/19-STS753) <https://doi.org/10.1214/19-STS753>
- [59] STRUTT, J. W. (1879). XXXI. Investigations in optics, with special reference to the spectroscope. *Philos. Mag.* **8** 261–274. <https://doi.org/10.1080/14786447908639684>
- [60] TENNE, R., ROSSMAN, U., REPHAEL, B., ISRAEL, Y., KRUPINSKI-PTASZEK, A., LAPKIEWICZ, R., SILBERBERG, Y. and ORON, D. (2019). Super-resolution enhancement by quantum image scanning microscopy. *Nat. Photonics* **13** 116–122. <https://doi.org/10.1038/s41566-018-0324-z>
- [61] TEREbizh, V. YU. (1995). Image restoration with minimum a priori information. *Usp. Fiz. Nauk* **165** 143–176. <https://doi.org/10.3367/ufnr.0165.199502b.0143>

- [62] THAM, W.-K., FERRETTI, H. and STEINBERG, A. M. (2017). Beating Rayleigh’s curse by imaging using phase information. *Phys. Rev. Lett.* **118** 070801. MR3678576 <https://doi.org/10.1103/PhysRevLett.118.070801>
- [63] TSANG, M., NAIR, R. and LU, X.-M. (2016). Quantum information for semiclassical optics. In *Proceedings of SPIE* 10029, *Quantum and Nonlinear Optics IV* 1002903. <https://doi.org/10.1117/12.2245733>
- [64] TSANG, M., NAIR, R. and LU, X. M. (2016). Quantum theory of superresolution for two incoherent optical point sources. *Phys. Rev. X* **6** 031033. <https://doi.org/10.1103/PhysRevX.6.031033>
- [65] VON DIEZMANN, A., SHECHTMAN, Y. and MOERNER, W. E. (2017). Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking. *Chem. Rev.* **117** 7244–7275. <https://doi.org/10.1021/acs.chemrev.6b00629>
- [66] WESTPHAL, V. and HELL, S. W. (2005). Nanoscale resolution in the focal plane of an optical microscope. *Phys. Rev. Lett.* **94** 143903. <https://doi.org/10.1103/PhysRevLett.94.143903>
- [67] WOUTERLOOT, F. G. (2012). *Cellular Imaging Techniques for Neuroscience and Beyond*. Academic Press, San Diego.
- [68] YAMAGATA, K., FUJIWARA, A. and GILL, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** 2197–2217. MR3127863 <https://doi.org/10.1214/13-AOS1147>

THE DISTRIBUTION OF THE LASSO: UNIFORM CONTROL OVER SPARSE BALLS AND ADAPTIVE PARAMETER TUNING

BY LÉO MIOLANE¹ AND ANDREA MONTANARI²

¹*Courant Institute of Mathematical Sciences and Center for Data Science, New York University, leo.miolane@gmail.com*

²*Department of Statistics and Department of Electrical Engineering, Stanford University, montanar@stanford.edu*

The Lasso is a popular regression method for high-dimensional problems in which the number of parameters $\theta_1, \dots, \theta_N$, is larger than the number n of samples: $N > n$. A useful heuristics relates the statistical properties of the Lasso estimator to that of a simple soft-thresholding denoiser, in a denoising problem in which the parameters $(\theta_i)_{i \leq N}$ are observed in Gaussian noise, with a carefully tuned variance. Earlier work confirmed this picture in the limit $n, N \rightarrow \infty$, pointwise in the parameters θ and in the value of the regularization parameter.

Here, we consider a standard random design model and prove exponential concentration of its empirical distribution around the prediction provided by the Gaussian denoising model. Crucially, our results are uniform with respect to θ belonging to ℓ_q balls, $q \in [0, 1]$, and with respect to the regularization parameter. This allows us to derive sharp results for the performances of various data-driven procedures to tune the regularization.

Our proofs make use of Gaussian comparison inequalities, and in particular of a version of Gordon’s minimax theorem developed by Thrampoulidis, Oymak and Hassibi, which controls the optimum value of the Lasso optimization problem. Crucially, we prove a stability property of the minimizer in Wasserstein distance that allows one to characterize properties of the minimizer itself.

REFERENCES

- [1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. [MR3311453](#) <https://doi.org/10.1093/imaiai/iau005>
- [2] BARBIER, J., DIA, M., MACRIS, N. and KRZAKALA, F. (2016). The mutual information in random linear estimation. In *54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 625–632. IEEE.
- [3] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2018). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA* **116** 5451–5460.
- [4] BAYATI, M., ERDOGDU, M. A. and MONTANARI, A. (2013). Estimating lasso risk and noise level. In *Advances in Neural Information Processing Systems* 944–952.
- [5] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. [MR2810285](#) <https://doi.org/10.1109/TIT.2010.2094817>
- [6] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inf. Theory* **58** 1997–2017. [MR2951312](#) <https://doi.org/10.1109/TIT.2011.2174612>
- [7] BELLEC, P. C. (2020). Out-of-sample error estimate for robust m-estimators with convex penalty. Available at [arXiv:2008.11840](https://arxiv.org/abs/2008.11840).
- [8] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference* **9** 33–79. [MR4079177](#) <https://doi.org/10.1093/imaiai/iy021>
- [9] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#) <https://doi.org/10.1214/08-AOS620>

- [10] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics.* Springer, Heidelberg. MR2807761 <https://doi.org/10.1007/978-3-642-20192-9>
- [11] CAI, T. T. and GUO, Z. (2018). Accuracy assessment for high-dimensional linear regression. *Ann. Statist.* **46** 1807–1836. MR3819118 <https://doi.org/10.1214/17-AOS1604>
- [12] CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. MR2382644 <https://doi.org/10.1214/009053606000001523>
- [13] CANDÈS, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inf. Theory* **51** 4203–4215. MR2243152 <https://doi.org/10.1109/TIT.2005.858979>
- [14] CELENTANO, M., MONTANARI, A. and WEI, Y. (2020). The lasso with general gaussian designs with applications to hypothesis testing. Available at [arXiv:2007.13716](https://arxiv.org/abs/2007.13716).
- [15] CHEN, S. and DONOHO, D. L. (1995). Examples of basis pursuit. In *Wavelet Applications in Signal and Image Processing III* **2569** 564–575. International Society for Optics and Photonics.
- [16] CHETVERIKOV, D., LIAO, Z. and CHERNOZHUKOV, V. (2016). On cross-validated lasso. Available at [arXiv:1605.02214](https://arxiv.org/abs/1605.02214).
- [17] 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 <https://doi.org/10.1007/s00440-015-0675-z>
- [18] DONOHO, D. L. (2006). High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.* **35** 617–652. MR2225676 <https://doi.org/10.1007/s00454-005-1220-0>
- [19] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p -balls for l_q -error. *Probab. Theory Related Fields* **99** 277–303. MR1278886 <https://doi.org/10.1007/BF01199026>
- [20] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [21] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inf. Theory* **57** 6920–6941. MR2882271 <https://doi.org/10.1109/TIT.2011.2165823>
- [22] DONOHO, D. L. and TANNER, J. (2005). Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA* **102** 9452–9457. MR2168716 <https://doi.org/10.1073/pnas.0502258102>
- [23] EFRON, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *J. Amer. Statist. Assoc.* **99** 619–642. MR2090899 <https://doi.org/10.1198/016214504000000692>
- [24] EL KAROUI, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. Available at [arXiv:1311.2445](https://arxiv.org/abs/1311.2445).
- [25] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 <https://doi.org/10.1007/s00440-016-0754-9>
- [26] GORDON, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. MR0800188 <https://doi.org/10.1007/BF02759761>
- [27] GORDON, Y. (1988). On Milman’s inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 <https://doi.org/10.1007/BFb0081737>
- [28] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
- [29] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 <https://doi.org/10.1109/TIT.2014.2343629>
- [30] JOHNSTONE, I. M. (2002). Function estimation and gaussian sequence models. Unpublished manuscript, 2(5.3):2.
- [31] MA, J., XU, J. and MALEKI, A. (2019). Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization. *IEEE Trans. Inf. Theory* **65** 3600–3629. MR3959008 <https://doi.org/10.1109/TIT.2019.2893254>
- [32] MIOLANE, L. and MONTANARI, A. (2021). Supplement to “The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning.” <https://doi.org/10.1214/20-AOS2038SUPP>
- [33] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. MR3472260 <https://doi.org/10.1109/TIT.2015.2457942>

- [34] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2017). Consistent parameter estimation for LASSO and approximate message passing. *Ann. Statist.* **45** 2427–2454. [MR3737897](#) <https://doi.org/10.1214/16-AOS1529>
- [35] NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M -estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. [MR3025133](#) <https://doi.org/10.1214/12-STS400>
- [36] PANAHI, A. and HASSIBI, B. (2017). A universal analysis of large-scale regularized least squares solutions. In *Advances in Neural Information Processing Systems* 3381–3390.
- [37] RANGAN, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *Information Theory Proceedings (ISIT)*, 2011 *IEEE International Symposium on* 2168–2172. IEEE, New York.
- [38] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. [MR2882274](#) <https://doi.org/10.1109/TIT.2011.2165799>
- [39] REEVES, G. and PFISTER, H. D. (2016). The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact. In *Information Theory (ISIT)*, 2016 *IEEE International Symposium on* 665–669. IEEE, New York.
- [40] SCHNITER, P. and RANGAN, S. (2015). Compressive phase retrieval via generalized approximate message passing. *IEEE Trans. Signal Process.* **63** 1043–1055. [MR3311635](#) <https://doi.org/10.1109/TSP.2014.2386294>
- [41] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [42] STOJNIC, M. (2010). Recovery thresholds for ℓ_1 optimization in binary compressed sensing. In *Information Theory Proceedings (ISIT)*, 2010 *IEEE International Symposium on* 1593–1597. IEEE, New York.
- [43] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. Preprint. Available at [arXiv:1303.7291](https://arxiv.org/abs/1303.7291).
- [44] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. [MR3984492](#) <https://doi.org/10.1073/pnas.1810420116>
- [45] TAKAHASHI, T. and KABASHIMA, Y. (2018). A statistical mechanics approach to de-biasing and uncertainty estimation in LASSO for random measurements. *J. Stat. Mech. Theory Exp.* **7** 073405, 25. [MR3845486](#) <https://doi.org/10.1088/1742-5468/aace2e>
- [46] THRAMOULIDIS, C., ABBASI, E. and HASSIBI, B. (2018). Precise error analysis of regularized M -estimators in high dimensions. *IEEE Trans. Inf. Theory* **64** 5592–5628. [MR3832326](#) <https://doi.org/10.1109/TIT.2018.2840720>
- [47] THRAMOULIDIS, C., OYMAK, S. and HASSIBI, B. (2015). Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory* 1683–1709.
- [48] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [49] TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. [MR2985948](#) <https://doi.org/10.1214/12-AOS1003>
- [50] TROPP, J. A. (2015). Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory, a Renaissance. Appl. Numer. Harmon. Anal.* 67–101. Birkhäuser/Springer, Cham. [MR3467419](#)
- [51] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. [MR3224285](#) <https://doi.org/10.1214/14-AOS1221>
- [52] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. [MR2576316](#) <https://doi.org/10.1214/09-EJS506>
- [53] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. [MR2459454](#) <https://doi.org/10.1007/978-3-540-71050-9>
- [54] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. [MR3153940](#) <https://doi.org/10.1111/rssb.12026>

OPTIMAL CHANGE-POINT ESTIMATION IN TIME SERIES

BY NGAI HANG CHAN^{1,*}, WAI LEONG NG², CHUN YIP YAU^{1,†} AND HAIHAN YU³

¹Department of Statistics, The Chinese University of Hong Kong, *nhchan@sta.cuhk.edu.hk; †cyyau@sta.cuhk.edu.hk

²Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, wlng@hsu.edu.hk

³Department of Statistics, Iowa State University, markyu@iastate.edu

This paper establishes asymptotic theory for optimal estimation of change points in general time series models under α -mixing conditions. We show that the Bayes-type estimator is asymptotically minimax for change-point estimation under squared error loss. Two bootstrap procedures are developed to construct confidence intervals for the change points. An approximate limiting distribution of the change-point estimator under small change is also derived. Simulations and real data applications are presented to investigate the finite sample performance of the Bayes-type estimator and the bootstrap procedures.

REFERENCES

- [1] ANTOCH, J. and HUŠKOVÁ, M. (2000). Bayesian like R - and M -estimators of change points. *Discuss. Math. Probab. Stat.* **20** 115–133. International Conference on Statistical Inference (Łagów, 1998). [MR1784846 https://doi.org/10.7151/dmps.1007](https://doi.org/10.7151/dmps.1007)
- [2] ANTOCH, J. and HUŠKOVÁ, M. (2000). Bayesian-type estimators of change points. *J. Statist. Plann. Inference* **91** 195–208. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). [MR1814780 https://doi.org/10.1016/S0378-3758\(00\)00178-6](https://doi.org/10.1016/S0378-3758(00)00178-6)
- [3] BAI, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* **15** 453–472. [MR1292161 https://doi.org/10.1111/j.1467-9892.1994.tb00204.x](https://doi.org/10.1111/j.1467-9892.1994.tb00204.x)
- [4] BARDET, J.-M. and WINTENBERGER, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* **37** 2730–2759. [MR2541445 https://doi.org/10.1214/08-AOS674](https://doi.org/10.1214/08-AOS674)
- [5] BEAULIEU, C., CHEN, J. and SARMIENTO, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **370** 1228–1249.
- [6] BERKES, I., HORVÁTH, L. and KOKOSZKA, P. (2003). GARCH processes: Structure and estimation. *Bernoulli* **9** 201–227. [MR1997027 https://doi.org/10.3150/bj/1068128975](https://doi.org/10.3150/bj/1068128975)
- [7] BOROVKOV, A. A. (1999). Asymptotically optimal solutions in the change-point problem. *Theory Probab. Appl.* **43** 539–561. [MR1692429 https://doi.org/10.1137/S0040585X97977112](https://doi.org/10.1137/S0040585X97977112)
- [8] BOSE, A. (1990). Bootstrap in moving average models. *Ann. Inst. Statist. Math.* **42** 753–768. [MR1089474 https://doi.org/10.1007/BF02481148](https://doi.org/10.1007/BF02481148)
- [9] BOUGEROL, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.* **31** 942–959. [MR1227540 https://doi.org/10.1137/0331041](https://doi.org/10.1137/0331041)
- [10] BOUSSAMA, F. (1998). Ergodicité, mélange et estimation dans les modeles GARCH. Ph.D. thesis, Université 7, Paris.
- [11] CARRASCO, M. and CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18** 17–39. [MR1885348 https://doi.org/10.1017/S0266466602181023](https://doi.org/10.1017/S0266466602181023)
- [12] CHAN, N. H. and KUTOYANTS, Y. A. (2012). On parameter estimation of threshold autoregressive models. *Stat. Inference Stoch. Process.* **15** 81–104. [MR2892589 https://doi.org/10.1007/s11203-011-9064-0](https://doi.org/10.1007/s11203-011-9064-0)
- [13] CHAN, N. H., NG, W. L., YAU, C. Y. and YU, H. (2021). Supplement to “Optimal change-point estimation in time series.” <https://doi.org/10.1214/20-AOS2039SUPP>
- [14] DATTA, S. (1995). Limit theory and bootstrap for explosive and partially explosive autoregression. *Stochastic Process. Appl.* **57** 285–304. [MR1334973 https://doi.org/10.1016/0304-4149\(94\)00078-8](https://doi.org/10.1016/0304-4149(94)00078-8)

MSC2020 subject classifications. Primary 62M10; secondary 60F25.

Key words and phrases. Bayes-type estimator, confidence interval, double-sided random process, piecewise stationary time series, structural break.

- [15] DATTA, S. (1996). On asymptotic properties of bootstrap for AR(1) processes. *J. Statist. Plann. Inference* **53** 361–374. [MR1407648](#) [https://doi.org/10.1016/0378-3758\(95\)00147-6](https://doi.org/10.1016/0378-3758(95)00147-6)
- [16] DAVIS, R. A., HUANG, D. W. and YAO, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.* **23** 282–304. [MR1331669](#) <https://doi.org/10.1214/aos/1176324468>
- [17] DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. [MR2268041](#) <https://doi.org/10.1198/016214505000000745>
- [18] DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. Springer, New York. [MR1312160](#) <https://doi.org/10.1007/978-1-4612-2642-0>
- [19] DOUKHAN, P. and WINTENBERGER, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* **118** 1997–2013. [MR2462284](#) <https://doi.org/10.1016/j.spa.2007.12.004>
- [20] DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* **19** 1471–1495. [MR1126333](#) <https://doi.org/10.1214/aos/1176348257>
- [21] EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. [MR3706768](#) <https://doi.org/10.3150/16-BEJ887>
- [22] ELLIOTT, G. and TIMMERMANN, A. (2005). Optimal forecast combination under regime switching. *Internat. Econom. Rev.* **46** 1081–1102. [MR2173623](#) <https://doi.org/10.1111/j.1468-2354.2005.00361.x>
- [23] FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. With 32 discussions by 47 authors and a rejoinder by the authors. [MR3210728](#) <https://doi.org/10.1111/rssb.12047>
- [24] GAO, Z. and LING, S. (2019). Statistical inference for structurally changed threshold autoregressive models. *Statist. Sinica* **29** 1803–1829. [MR3970336](#)
- [25] GRANGER, C. W. J. and ANDERSEN, A. (1978/79). On the invertibility of time series models. *Stochastic Process. Appl.* **8** 87–92. [MR0511877](#) [https://doi.org/10.1016/0304-4149\(78\)90069-8](https://doi.org/10.1016/0304-4149(78)90069-8)
- [26] HUŠKOVÁ, M. and KIRCH, C. (2008). Bootstrapping confidence intervals for the change-point of time series. *J. Time Series Anal.* **29** 947–972. [MR2464948](#) <https://doi.org/10.1111/j.1467-9892.2008.00589.x>
- [27] JEONG, M. (2017). Residual-based Garch bootstrap and second order asymptotic refinement. *Econometric Theory* **33** 779–790. [MR3637976](#) <https://doi.org/10.1017/S0266466616000104>
- [28] LAI, W. R., JOHNSON, M. D., KUCHERLAPATI, R. and PARK, P. J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* **21** 3763–3770.
- [29] LING, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.* **35** 1213–1237. [MR2341704](#) <https://doi.org/10.1214/009053606000001514>
- [30] LING, S. (2016). Estimation of change-points in linear and nonlinear time series models. *Econometric Theory* **32** 402–430. [MR3471122](#) <https://doi.org/10.1017/S0266466614000863>
- [31] LING, S. and MCALEER, M. (2010). A general asymptotic theory for time-series models. *Stat. Neerl.* **64** 97–111. [MR2830968](#) <https://doi.org/10.1111/j.1467-9574.2009.00447.x>
- [32] MCCORMAC, F. G., HOGG, A. G., BLACKWELL, P. G., BUCK, C. E., HIGHAM, T. F. G. and REIMER, P. J. (2004). Shcal04 southern hemisphere calibration, 0–11.0 cal kyr BP. *Radiocarbon* **46** 1087–1092.
- [33] MOKKADEM, A. (1988). Mixing properties of ARMA processes. *Stochastic Process. Appl.* **29** 309–315. [MR0958507](#) [https://doi.org/10.1016/0304-4149\(88\)90045-2](https://doi.org/10.1016/0304-4149(88)90045-2)
- [34] NG, W. L., PAN, S. and YAU, C. Y. (2017). Inference for multiple change-points in linear and non-linear time series models. [arXiv:1703.00647](#).
- [35] OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.
- [36] REIMER, P. J., BAILLIE, M. G. L., BARD, E., BAYLISS, A., BECK, J. W., BERTRAND, C. J. H., BLACKWELL, P. G., BUCK, C. E. and WEYHENMEYER, C. E. (2004). IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* **46** 1029–1058.
- [37] RITOV, Y. (1990). Asymptotic efficient estimation of the change point with unknown distributions. *Ann. Statist.* **18** 1829–1839. [MR1074438](#) <https://doi.org/10.1214/aos/1176347881>
- [38] RODGERS, K. B., MIKALOFF-FLETCHER, S. E., BIANCHI, D., BEAULIEU, C., GALBRAITH, E. D., GNANADESIKAN, A., HOGG, A. G., IUDICONE, D., LINTNER, B. R. et al. (2011). Interhemispheric gradient of atmospheric radiocarbon reveals natural variability of Southern Ocean winds. *Clim. Past* **7** 1123–1138.
- [39] RUKHIN, A. L. (1994). Asymptotic minimaxity in the change-point problem. In *Change-Point Problems (South Hadley, MA, 1992). Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23** 284–291. IMS, Hayward, CA. [MR1477931](#) <https://doi.org/10.1214/lnms/1215463131>

- [40] SHAO, Q. M. (1993). Complete convergence for α -mixing sequences. *Statist. Probab. Lett.* **16** 279–287. [MR1212051 https://doi.org/10.1016/0167-7152\(93\)90131-2](https://doi.org/10.1016/0167-7152(93)90131-2)
- [41] SHAO, X. and ZHANG, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. [MR2752617 https://doi.org/10.1198/jasa.2010.tm10103](https://doi.org/10.1198/jasa.2010.tm10103)
- [42] WINTENBERGER, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scand. J. Stat.* **40** 846–867. [MR3145121 https://doi.org/10.1111/sjos.12038](https://doi.org/10.1111/sjos.12038)
- [43] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215 https://doi.org/10.1073/pnas.0506715102](https://doi.org/10.1073/pnas.0506715102)
- [44] WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. [MR2353389 https://doi.org/10.1214/009117907000000060](https://doi.org/10.1214/009117907000000060)
- [45] YAU, C. Y. and ZHAO, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 895–916. [MR3534355 https://doi.org/10.1111/rssb.12139](https://doi.org/10.1111/rssb.12139)

PROPRIETY OF THE REFERENCE POSTERIOR DISTRIBUTION IN GAUSSIAN PROCESS MODELING

BY JOSEPH MURÉ

EDF Recherche et Développement and Université Paris Diderot, joseph.mure@edf.fr

In a seminal article, Berger, De Oliveira and Sansó [*J. Amer. Statist. Assoc.* **96** (2001) 1361–1374] compare several objective prior distributions for the parameters of Gaussian process models with isotropic correlation kernel. The reference prior distribution stands out among them insofar as it always leads to a proper posterior. They prove this result for rough correlation kernels: Spherical, Exponential with power $\rho < 2$, Matérn with smoothness $\nu < 1$. This paper provides a proof for smooth correlation kernels: Exponential with power $\rho = 2$, Matérn with smoothness $\nu \geq 1$, Rational Quadratic, along with tail rates of the reference prior for these kernels.

REFERENCES

- ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55*. U.S. Government Printing Office, Washington, D.C. [MR0167642](#)
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, Boca Raton.
- BERGER, J. O., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* **96** 1361–1374. [MR1946582](#) <https://doi.org/10.1198/016214501753382282>
- BERNARDO, J. M. (2005). Reference analysis. In *Bayesian Thinking: Modeling and Computation. Handbook of Statist.* **25** 17–90. Elsevier/North-Holland, Amsterdam. [MR2490522](#) [https://doi.org/10.1016/S0169-7161\(05\)25002-2](https://doi.org/10.1016/S0169-7161(05)25002-2)
- CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. [MR2471287](#) <https://doi.org/10.1214/08-EJS273>
- CLARKE, B. S. and BARRON, A. R. (1994). Jeffreys’ prior is asymptotically least favorable under entropy risk. *J. Statist. Plann. Inference* **41** 37–60. [MR1292146](#) [https://doi.org/10.1016/0378-3758\(94\)90153-8](https://doi.org/10.1016/0378-3758(94)90153-8)
- DE OLIVEIRA, V., KEDEM, B. and SHORT, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *J. Amer. Statist. Assoc.* **92** 1422–1433. [MR1615252](#) <https://doi.org/10.2307/2965412>
- DRUILHET, P. and MARIN, J.-M. (2007). Invariant HPD credible sets and MAP estimators. *Bayesian Anal.* **2** 681–691. [MR2361970](#) <https://doi.org/10.1214/07-BA227>
- GOWER, J. C. (1985). Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Appl.* **67** 81–97. [MR0787367](#) [https://doi.org/10.1016/0024-3795\(85\)90187-9](https://doi.org/10.1016/0024-3795(85)90187-9)
- GU, M., WANG, X. and BERGER, J. O. (2018). Robust Gaussian stochastic process emulation. *Ann. Statist.* **46** 3038–3066. [MR3851764](#) <https://doi.org/10.1214/17-AOS1648>
- HANDCOCK, M. S. and WALLIS, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Statist. Assoc.* **89** 368–390. [MR1294070](#)
- KAZIANKA, H. and PILZ, J. (2012). Objective Bayesian analysis of spatial data with uncertain nugget and range parameters. *Canad. J. Statist.* **40** 304–327. [MR2927748](#) <https://doi.org/10.1002/cjs.11132>
- MURÉ, J. (2019). Optimal compromise between incompatible conditional probability distributions, with application to objective Bayesian Kriging. *ESAIM Probab. Stat.* **23** 271–309. [MR3963529](#) <https://doi.org/10.1051/ps/2018023>
- MURÉ, J. (2021). Supplement to “Propriety of the reference posterior distribution in Gaussian process modeling.” <https://doi.org/10.1214/20-AOS2040SUPP>
- PAULO, R. (2005). Default priors for Gaussian processes. *Ann. Statist.* **33** 556–582. [MR2163152](#) <https://doi.org/10.1214/009053604000001264>

MSC2020 subject classifications. Primary 62F15; secondary 62M30, 60G15.

Key words and phrases. Gaussian process, Kriging, correlation function, Jeffreys prior, reference prior, integrated likelihood, posterior propriety.

- RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2514435](#)
- REN, C., SUN, D. and HE, C. (2012). Objective Bayesian analysis for a spatial model with nugget effects. *J. Statist. Plann. Inference* **142** 1933–1946. [MR2903403](#) <https://doi.org/10.1016/j.jspi.2012.02.034>
- REN, C., SUN, D. and SAHU, S. K. (2013). Objective Bayesian analysis of spatial models with separable correlation functions. *Canad. J. Statist.* **41** 488–507. [MR3101596](#) <https://doi.org/10.1002/cjs.11186>
- SCHOENBERG, I. J. (1937). On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math. (2)* **38** 787–793. [MR1503370](#) <https://doi.org/10.2307/1968835>
- SERRE, D. (2002). *Matrices: Theory and Applications. Graduate Texts in Mathematics* **216**. Springer, New York. [MR1923507](#)
- STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. [MR1697409](#) <https://doi.org/10.1007/978-1-4612-1494-6>
- VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. [MR2418663](#) <https://doi.org/10.1214/0090536070000000613>

COMMUNITY DETECTION WITH DEPENDENT CONNECTIVITY

BY YUBAI YUAN^{*} AND ANNIE QU[†]

Department of Statistics, University of California, Irvine, ^{*}yubaiy@uci.edu; [†]qu2@uci.edu

In network analysis, within-community members are more likely to be connected than between-community members, which is reflected in that the edges within a community are intercorrelated. However, existing probabilistic models for community detection such as the stochastic block model (SBM) are not designed to capture the dependence among edges. In this paper, we propose a new community detection approach to incorporate intra-community dependence of connectivities through the Bahadur representation. The proposed method does not require specifying the likelihood function, which could be intractable for correlated binary connectivities. In addition, the proposed method allows for heterogeneity among edges between different communities. In theory, we show that incorporating correlation information can achieve a faster convergence rate compared to the independent SBM, and the proposed algorithm has a lower estimation bias and accelerated convergence compared to the variational EM. Our simulation studies show that the proposed algorithm outperforms the existing multinet network community detection methods assuming conditional independence among edges. We also demonstrate the application of the proposed method to agricultural product trading networks from different countries and to brain fMRI imaging networks.

REFERENCES

- [1] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- [2] ALEXANDER-BLOCH, A., LAMBIOTTE, R., ROBERTS, B., GIEDD, J., GOGTAY, N. and BULLMORE, E. (2012). The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage* **59** 3889–3900.
- [3] AMELIO, A., MANGIONI, G. and TAGARELLI, A. (2020). Modularity in multilayer networks using redundancy-based resolution and projection-based inter-layer coupling. *IEEE Trans. Netw. Sci. Eng.* **7** 1198–1214. [MR4149995](#)
- [4] ANAGNOSTOPOULOS, A., KUMAR, R. and MAHDIAN, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 7–15. ACM.
- [5] BAHADUR, R. R. (1959). A representation of the joint distribution of responses to n dichotomous items. Technical Report.
- [6] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. [MR3611487](#) <https://doi.org/10.1214/16-AOS1435>
- [7] BALL, B., KARRER, B. and NEWMAN, M. E. (2011). Efficient and principled method for detecting communities in networks. *Phys. Rev. E* **84**.
- [8] BARIGOZZI, M., FAGIOLO, G. and GARLASCHELLI, D. (2010). Multinet network of international trade: A commodity-specific analysis. *Phys. Rev. E* **81** 046104.
- [9] BETZEL, R. F., BERTOLERO, M. A., GORDON, E. M., GRATTON, C., DOSENBACH, N. U. and BASSETT, D. S. (2018). The community structure of functional brain networks exhibits scale-specific patterns of variability across individuals and time. *bioRxiv* 413278.
- [10] BHATTACHARYYA, S. and CHATTERJEE, S. (2018). Spectral clustering for multiple sparse networks: I. *arXiv preprint arXiv:1805.10594*.

MSC2020 subject classifications. Primary 62R07; secondary 62E17.

Key words and phrases. Bahadur representation, high-order approximation, multiple networks, stochastic block model, variational EM.

- [11] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*.
- [12] 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**.
- [13] BORDENAVE, C., LELARGE, M. and MASSOULIÉ, L. (2015). Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 1347–1357. IEEE Computer Soc., Los Alamitos, CA. [MR3473374](#) <https://doi.org/10.1109/FOCS.2015.86>
- [14] CALHOUN, V. D., EICHELE, T. and PEARLSON, G. (2009). Functional brain networks in schizophrenia: A review. *Front. Human Neurosci.* **3** 17. <https://doi.org/10.3389/neuro.09.017.2009>
- [15] CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. [MR2988467](#) <https://doi.org/10.1214/12-EJS729>
- [16] CHEN, X. and YANG, Y. (2021). Hanson–Wright inequality in Hilbert spaces with application to K -means clustering for non-Euclidean data. *Bernoulli* **27** 586–614. [MR4177382](#) <https://doi.org/10.3150/20-BEJ1251>
- [17] CHENG, J., LEVINA, E., WANG, P. and ZHU, J. (2014). A sparse Ising model with covariates. *Biometrics* **70** 943–953. [MR3295755](#) <https://doi.org/10.1111/biom.12202>
- [18] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. [MR2931253](#) <https://doi.org/10.1093/biomet/asr053>
- [19] DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. [MR2390817](#) <https://doi.org/10.1007/s11222-007-9046-7>
- [20] DESIKAN, R. S., SÉGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31** 968–980.
- [21] DE DOMENICO, M., NICOSIA, V., ARENAS, A. and LATORA, V. (2015). Structural reducibility of multi-layer networks. *Nat. Commun.* **6** 1–9.
- [22] DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis. *Appl. Stat.* 49–93.
- [23] DONG, X., FROSSARD, P., VANDERGHEYNST, P. and NEFEDOV, N. (2014). Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. *IEEE Trans. Signal Process.* **62** 905–918. [MR3160322](#) <https://doi.org/10.1109/TSP.2013.2295553>
- [24] FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. [MR0860518](#)
- [25] FRAZIER, J. A., CHIU, S., BREEZE, J. L., MAKRI, N., LANGE, N., KENNEDY, D. N., HERBERT, M. R., BENT, E. K., KONERU, V. K. et al. (2005). Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am. J. Psychiatr.* **162** 1256–1265.
- [26] GENG, J., BHATTACHARYA, A. and PATI, D. (2019). Probabilistic community detection with unknown number of communities. *J. Amer. Statist. Assoc.* **114** 893–905. [MR3963189](#) <https://doi.org/10.1080/01621459.2018.1458618>
- [27] HADLEY, J. A., KRAGULJAC, N. V., WHITE, D. M., VER HOEF, L., TABORA, J. and LAHTI, A. C. (2016). Change in brain network topology as a function of treatment response in schizophrenia: A longitudinal resting-state fMRI study using graph theory. *Npj Schizophrenia* **2** 1–7.
- [28] HAN, Q., XU, K. and AIROLDI, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning* 1511–1520.
- [29] HANDCOCK, M. S. (2003). Statistical models for social networks. In *Dynamic Social Network Modeling and Analysis*. National Academies Press, Washington.
- [30] HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *J. Roy. Statist. Soc. Ser. A* **170** 301–354. [MR2364300](#) <https://doi.org/10.1111/j.1467-985X.2007.00471.x>
- [31] HOFF, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* 657–664.
- [32] HOFF, P. (2021). Additive and multiplicative effects network models. *Statist. Sci.* **36** 34–50. [MR4194202](#) <https://doi.org/10.1214/19-STS757>
- [33] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088](#) [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [34] HONEY, G. D., POMAROL-CLOTET, E., CORLETT, P. R., HONEY, R. A., MCKENNA, P. J., BULLMORE, E. T. and FLETCHER, P. C. (2005). Functional dysconnectivity in schizophrenia associated with attentional modulation of motor function. *Brain* **128** 2597–2611.

- [35] JAAKKOLA, T. (2001). Tutorial on variational approximation methods. In *Advanced Mean Field Methods: Theory and Practice*.
- [36] JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing* 665–674. ACM, New York. [MR3210828](#) <https://doi.org/10.1145/2488608.2488693>
- [37] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. [MR2788206](#) <https://doi.org/10.1103/PhysRevE.83.016107>
- [38] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. [MR2683452](#) <https://doi.org/10.1109/TIT.2010.2046205>
- [39] KIM, N., WILBURNE, D., PETROVIĆ, S. and RINALDO, A. (2016). On the geometry and extremal properties of the edge-degeneracy model. arXiv preprint [arXiv:1602.00180](#).
- [40] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *J. Complex Netw.* **2** 203–271.
- [41] LATORA, V. and MARCHIORI, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87** 198701. <https://doi.org/10.1103/PhysRevLett.87.198701>
- [42] LATOUCHE, P., BIRMELE, E. and AMBROISE, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **12** 93–115. [MR2953099](#) <https://doi.org/10.1177/1471082X1001200105>
- [43] LAURITZEN, S., RINALDO, A. and SADEGHI, K. (2018). Random networks, graphical models and exchangeability. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 481–508. [MR3798875](#) <https://doi.org/10.1111/rssb.12266>
- [44] LE, C. M., LEVIN, K. and LEVINA, E. (2018). Estimating a network from multiple noisy realizations. *Electron. J. Stat.* **12** 4697–4740. [MR3894068](#) <https://doi.org/10.1214/18-ejs1521>
- [45] LEE, K.-M., YANG, J.-S., KIM, G., LEE, J., GOH, K.-I. and KIM, I.-M. (2011). Impact of the topology of global macroeconomic network on the spreading of economic crises. *PLoS ONE* **6**.
- [46] LEVIN, K., ATHREYA, A., TANG, M., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference. arXiv preprint [arXiv:1705.09355](#).
- [47] LEVIN, K., LODHIA, A. and LEVINA, E. (2019). Recovering low-rank structure from multiple networks with unknown edge distributions. arXiv preprint [arXiv:1906.07265](#).
- [48] LIANG, M., ZHOU, Y., JIANG, T., LIU, Z., TIAN, L., LIU, H. and HAO, Y. (2006). Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *NeuroReport* **17** 209–213.
- [49] LYNALL, M.-E., BASSETT, D. S., KERWIN, R., MCKENNA, P. J., KITZBICHLER, M., MULLER, U. and BULLMORE, E. (2010). Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* **30** 9477–9487.
- [50] MARIADASSOU, M., ROBIN, S. and VACHER, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.* **4** 715–742. [MR2758646](#) <https://doi.org/10.1214/10-AOAS361>
- [51] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York. [MR3238997](#)
- [52] MEUNIER, D., LAMBIOTTE, R. and BULLMORE, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* **4** 200. <https://doi.org/10.3389/fnins.2010.00200>
- [53] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* 69–75. ACM.
- [54] NEWMAN, M. E. and REINERT, G. (2016). Estimating the number of communities in a network. *Phys. Rev. Lett.* **117**.
- [55] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256. [MR2010377](#) <https://doi.org/10.1137/S003614450342480>
- [56] NEWMAN, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* (3) **74** 036104, 19. [MR2282139](#) <https://doi.org/10.1103/PhysRevE.74.036104>
- [57] NOBILE, A. and FEARNSIDE, A. T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Stat. Comput.* **17** 147–162. [MR2380643](#) <https://doi.org/10.1007/s11222-006-9014-7>
- [58] PATTISON, P. and ROBINS, G. (2002). Neighborhood-based models for social networks. *Sociol. Method.* **32** 301–337.
- [59] PAUL, S. and CHEN, Y. (2020). A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Ann. Appl. Stat.* **14** 993–1029. [MR4117838](#) <https://doi.org/10.1214/20-AOAS1339>

- [60] PAVLOVIC, D. M. (2015). Generalised stochastic blockmodels and their applications in the analysis of brain networks. Ph.D. thesis, Univ. Warwick.
- [61] RAHMAN, M. S., DEY, L. R., HAIDER, S., UDDIN, M. A. and ISLAM, M. (2017). Link prediction by correlation on social network. In *Computer and Information Technology (ICCIT)*, 2017 20th International Conference of 1–6. IEEE.
- [62] ROBINS, G., PATTISON, P., KALISH, Y. and LUSHER, D. (2007). An introduction to exponential random graph (p^*) models for social networks. *Soc. Netw.* **29** 173–191.
- [63] ROBINS, G., SNIJDERS, T., WANG, P., HANDCOCK, M. and PATTISON, P. (2007). Recent developments in exponential random graph (p^*) models for social networks. *Soc. Netw.* **29** 192–215.
- [64] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856 https://doi.org/10.1214/11-AOS887](https://doi.org/10.1214/11-AOS887)
- [65] SAADE, A., KRZAKALA, F. and ZDEBOROVÁ, L. (2014). Spectral clustering of graphs with the Bethe Hessian. In *Advances in Neural Information Processing Systems* 406–414.
- [66] SALDAÑA, D. F., YU, Y. and FENG, Y. (2017). How many communities are there? *J. Comput. Graph. Statist.* **26** 171–181. [MR3610418 https://doi.org/10.1080/10618600.2015.1096790](https://doi.org/10.1080/10618600.2015.1096790)
- [67] STANLEY, N., SHAI, S., TAYLOR, D. and MUCHA, P. J. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **3** 95–105. [MR3515211 https://doi.org/10.1109/TNSE.2016.2537545](https://doi.org/10.1109/TNSE.2016.2537545)
- [68] TANG, W., LU, Z. and DHILLON, I. S. (2009). Clustering with multiple graphs. In 2009 Ninth IEEE International Conference on Data Mining 1016–1021. IEEE.
- [69] TAYLOR, S. F., CHEN, A. C., TSO, I. F., LIBERZON, I. and WELSH, R. C. (2011). Social appraisal in chronic psychosis: Role of medial frontal and occipital networks. *J. Psychiatr. Res.* **45** 526–538.
- [70] TAYLOR, S. F., DEMETER, E., PHAN, K. L., TSO, I. F. and WELSH, R. C. (2014). Abnormal GABAergic function and negative affect in schizophrenia. *Neuropsychopharmacology* **39** 1000–1008.
- [71] TORREGGIANI, S., MANGIONI, G., PUMA, M. J. and FAGIOLO, G. (2018). Identifying the community structure of the food-trade international multi-network. *Environ. Res. Lett.* **13** 054026.
- [72] TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. and JOLIOT, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15** 273–289.
- [73] VON DER MALSBERG, C. (1994). The correlation theory of brain function. In *Models of Neural Networks* 95–119. Springer.
- [74] WANG, S., ARROYO, J., VOGELSTEIN, J. T. and PRIEBE, C. E. (2019). Joint embedding of graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*
- [75] WASSERMAN, S., FAUST, K. et al. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge Univ. Press.
- [76] WHITFIELD-GABRIELI, S. and NIETO-CASTANON, A. (2012). Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* **2** 125–141.
- [77] WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103. [MR0684867 https://doi.org/10.1214/aos/1176346060](https://doi.org/10.1214/aos/1176346060)
- [78] XIA, M., WANG, J. and HE, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS ONE* **8** e68910.
- [79] YU, Q., ALLEN, E., SUI, J., ARBABSHIRANI, M., PEARLSON, G. and CALHOUN, V. (2012). Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. *Curr. Top. Med. Chem.* **12** 2415–2425.
- [80] YUAN, Y. and QU, A. (2021). Supplement to “Community detection with dependent connectivity.” <https://doi.org/10.1214/20-AOS2042SUPP>
- [81] ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann. Statist.* **48** 2575–2598. [MR4152113 https://doi.org/10.1214/19-AOS1898](https://doi.org/10.1214/19-AOS1898)
- [82] ZHANG, J., SUN, W. W. and LI, L. (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *J. Amer. Statist. Assoc.* **115** 2022–2036. [MR4189774 https://doi.org/10.1080/01621459.2019.1677242](https://doi.org/10.1080/01621459.2019.1677242)
- [83] 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 https://doi.org/10.1214/12-AOS1036](https://doi.org/10.1214/12-AOS1036)

CORRECTION NOTE: “STATISTICAL INFERENCE FOR THE MEAN OUTCOME UNDER A POSSIBLY NONUNIQUE OPTIMAL TREATMENT RULE”

BY ALEX LUEDTKE¹, AURÉLIEN BIBAUT^{2,*} AND MARK VAN DER LAAN^{2,†}

¹*Department of Statistics, University of Washington, aluedtke@uw.edu*

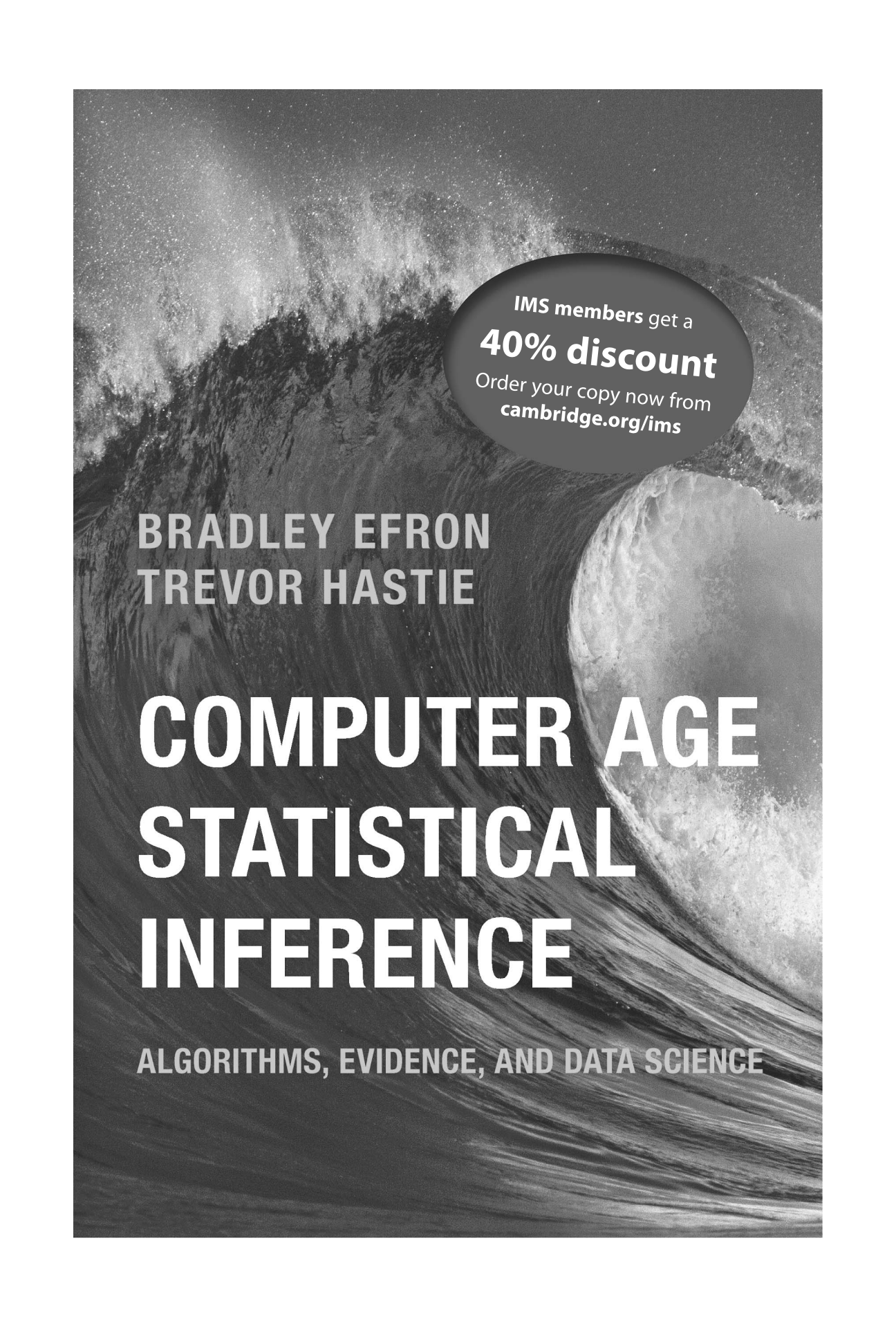
²*Division of Biostatistics, University of California, Berkeley, *aurelien.bibaut@berkeley.edu; †laan@berkeley.edu*

REFERENCES

- [1] BIBAUT, A. F., LUEDTKE, A. and VAN DER LAAN, M. J. (2020). Sufficient and insufficient conditions for the stochastic convergence of Cesàro means. Preprint. Available at [arXiv:2009.05974](https://arxiv.org/abs/2009.05974).
- [2] LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. [MR3476615 https://doi.org/10.1214/15-AOS1384](https://doi.org/10.1214/15-AOS1384)

MSC2020 subject classifications. Primary 62G05; secondary 62N99.

Key words and phrases. Efficient estimator, nonregular inference, online estimation, optimal treatment, path-wise differentiability, semiparametric model, optimal value.



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE