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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Vol. 47, No. 2—April 2019

THE ANNALS OF STATISTICS Vol. 47, No. 2, pp. 663–1202 April 2019

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 47, Number 2, April 2019. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2019, Vol. 47, No. 2, 663–690
https://doi.org/10.1214/17-AOS1602

COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES IN CYCLIC STRUCTURAL EQUATION MODELS

BY MATHIAS DRTON*, CHRISTOPHER FOX† AND Y. SAMUEL WANG*

University of Washington and University of Chicago†*

Software for computation of maximum likelihood estimates in linear structural equation models typically employs general techniques from nonlinear optimization, such as quasi-Newton methods. In practice, careful tuning of initial values is often required to avoid convergence issues. As an alternative approach, we propose a block-coordinate descent method that cycles through the considered variables, updating only the parameters related to a given variable in each step. We show that the resulting block update problems can be solved in closed form even when the structural equation model comprises feedback cycles. Furthermore, we give a characterization of the models for which the block-coordinate descent algorithm is well defined, meaning that for generic data and starting values all block optimization problems admit a unique solution. For the characterization, we represent each model by its mixed graph (also known as path diagram), which leads to criteria that can be checked in time that is polynomial in the number of considered variables.

REFERENCES

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

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York. MR0996025

CHAUDHURI, S., DRTON, M. and RICHARDSON, T. S. (2007). Estimation of a covariance matrix with zeros. *Biometrika* **94** 199–216. MR2307904

COLOMBO, D., MAATHUIS, M. H., KALISCH, M. and RICHARDSON, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *Ann. Statist.* **40** 294–321. MR3014308

DRTON, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. MR2502658

DRTON, M., EICHLER, M. and RICHARDSON, T. S. (2009). Computing maximum likelihood estimates in recursive linear models with correlated errors. *J. Mach. Learn. Res.* **10** 2329–2348. MR2563984

DRTON, M., FOX, C. and WANG, Y. S. (2018). Supplement to "Computation of maximum likelihood estimates in cyclic structural equation models." DOI:10.1214/17-AOS1602SUPP.

DRTON, M. and MAATHUIS, M. (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application* **4** 365–393.

DRTON, M. and RICHARDSON, T. S. (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* **91** 383–392. MR2081308

DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Birkhäuser, Basel. MR2723140

MSC2010 subject classifications. 62H12, 62F10.

Key words and phrases. Cyclic graph, feedback, linear structural equation model, graphical model, maximum likelihood estimation.

FOX, J. (2006). Structural equation modeling with the sem package in R. *Struct. Equ. Model.* **13** 465–486. MR2240110

FOX, C. (2014). Interpretation and inference of linear structural equation models. Ph.D. thesis, Univ. Chicago.

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

GRACE, J. B., ANDERSON, T. M., SEABLOOM, E. W., BORER, E. T., ADLER, P. B., HARPOLE, W. S., HAUTIER, Y., HILLEBRAND, H., LIND, E. M., et al. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529** 390–393.

HARARY, F. (1962). The determinant of the adjacency matrix of a graph. *SIAM Rev.* **4** 202–210. MR0144330

HOYLE, R. H., ed. (2012). *Handbook of Structural Equation Modeling*. Guilford Press, New York.

KLINE, R. B. (2015). *Principles and Practice of Structural Equation Modeling*, 4th ed. Guilford Press, New York.

LACERDA, G., SPIRTES, P., RAMSEY, J. and HOYER, P. (2008). Discovering cyclic causal models by independent components analysis. In *Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence* (*UAI*-08) 366–374. AUAI Press, Corvallis, OR.

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

MCDONALD, R. P. and HARTMANN, W. M. (1992). A procedure for obtaining initial values of parameters in the RAM model. *Multivar. Behav. Res.* **27** 57–76.

MOOIJ, J. M. and HESKES, T. (2013). Cyclic causal discovery from continuous equilibrium data. In *Proceedings of the* 29*th Annual Conference on Uncertainty in Artificial Intelligence* (*UAI*-13) (A. Nicholson and P. Smyth, eds.) 431–439. AUAI Press, Corvallis, OR.

NARAYANAN, A. (2012). A review of eight software packages for structural equation modeling. *Amer. Statist.* **66** 129–138.

NOWZOHOUR, C., MAATHUIS, M. and BÜHLMANN, P. (2015). Structure learning with bow-free acyclic path diagrams. Available at: arxiv:1508.01717.

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

PARK, G. and RASKUTTI, G. (2016). Identifiability assumptions and algorithm for directed graphical models with feedback. Available at: arxiv:1602.04418.

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286

R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

RICHARDSON, T. (1996). A discovery algorithm for directed cyclic graphs. In *Uncertainty in Artificial Intelligence* (*Portland*, *OR*, 1996) 454–461. Morgan Kaufmann, San Francisco, CA. MR1617227

RICHARDSON, T. (1997). A characterization of Markov equivalence for directed cyclic graphs. *Internat. J. Approx. Reason.* **17** 107–162. MR1462712

ROSSEEL, Y. (2012). lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **48** 1–36.

SACHS, K., PEREZ, O., PE'ER, D., LAUFFENBURGER, D. A. and NOLAN, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308** 523–529.

SILVA, R. (2013). A MCMC approach for learning the structure of Gaussian acyclic directed mixed graphs. In *Statistical Models for Data Analysis* (P. Giudici, S. Ingrassia and M. Vichi, eds.) 343–351. Springer, New York.

SPIRTES, P. (1995). Directed cyclic graphical representations of feedback models. In *Uncertainty in Artificial Intelligence*: *Proceedings of the* 11*th Conference* (P. Besnard and S. Hanks, eds.) 491–498. Morgan Kaufmann, San Francisco, CA.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. MIT Press, Cambridge, MA. MR1815675

STATACORP (2013). STATA structural equation modeling reference manual. StataCorp LP, College Station, TX, Release 13.

STEIGER, J. H. (2001). Driving fast in reverse. *J. Amer. Statist. Assoc.* **96** 331–338.

TRIANTAFILLOU, S. and TSAMARDINOS, I. (2016). Score-based vs constraint-based causal learning in the presence of confounders. In *UAI* 2016 *Workshop on Causation*: *Foundation to Application* (F. Eberhardt, E. Bareinboim, M. Maathuis, J. Mooij and R. Silva, eds.). *CEUR Workshop Proceedings* **1792** 59–67.

WERMUTH, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879. MR2817608

WRIGHT, S. (1921). Correlation and causation. *J. Agricultural Research* **20** 557–585.

WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.

The Annals of Statistics
2019, Vol. 47, No. 2, 691–719
https://doi.org/10.1214/17-AOS1624

FRÉCHET REGRESSION FOR RANDOM OBJECTS WITH EUCLIDEAN PREDICTORS

BY ALEXANDER PETERSEN AND HANS-GEORG MÜLLER[1]

University of California, Santa Barbara and University of California, Davis

Increasingly, statisticians are faced with the task of analyzing complex data that are non-Euclidean and specifically do not lie in a vector space. To address the need for statistical methods for such data, we introduce the concept of Fréchet regression. This is a general approach to regression when responses are complex random objects in a metric space and predictors are in $\mathcal{R}^p$, achieved by extending the classical concept of a Fréchet mean to the notion of a conditional Fréchet mean. We develop generalized versions of both global least squares regression and local weighted least squares smoothing. The target quantities are appropriately defined population versions of global and local regression for response objects in a metric space. We derive asymptotic rates of convergence for the corresponding fitted regressions using observed data to the population targets under suitable regularity conditions by applying empirical process methods. For the special case of random objects that reside in a Hilbert space, such as regression models with vector predictors and functional data as responses, we obtain a limit distribution. The proposed methods have broad applicability. Illustrative examples include responses that consist of probability distributions and correlation matrices, and we demonstrate both global and local Fréchet regression for demographic and brain imaging data. Local Fréchet regression is also illustrated via a simulation with response data which lie on the sphere.

REFERENCES

AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

ALLEN, E. A., DAMARAJU, E., PLIS, S. M., ERHARDT, E. B., EICHELE, T. and CALHOUN, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24** 663–676.

ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. MR2288028

BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** no. 25. MR3035753

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

MSC2010 subject classifications. Primary 62G05; secondary 62J99, 62G08.

Key words and phrases. Least squares regression, random objects, metric spaces, local linear regression, functional connectivity, densities as objects.

BHATTACHARYA, R. N., ELLINGSON, L., LIU, X., PATRANGENARU, V. and CRANE, M. (2012). Extrinsic analysis on manifolds is computationally faster than intrinsic analysis with applications to quality control by machine vision. *Appl. Stoch. Models Bus. Ind.* **28** 222–235. MR2944144

BORSDORF, R. and HIGHAM, N. J. (2010). A preconditioned Newton algorithm for the nearest correlation matrix. *IMA J. Numer. Anal.* **30** 94–107. MR2580548

BOUMAL, N., MISHRA, B., ABSIL, P.-A., SEPULCHRE, R. et al. (2014). Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.* **15** 1455–1459.

BRADLEY, J. V. (1968). *Distribution-Free Statistical Tests*. Prentice Hall, Englewood Cliffs, NJ.

CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306. MR0981451

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403. MR0516581

DAVIS, B. C., FLETCHER, P. T., BULLITT, E. and JOSHI, S. (2007). Population shape regression from random design data. In *IEEE* 11*th International Conference on Computer Vision*, *ICCV* 2007 1–7.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–121. MR1435485

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. MR1383587

FARAWAY, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39** 254–261. MR1462586

FARAWAY, J. J. (2014). Regression for non-Euclidean data using distance matrices. *J. Appl. Stat.* **41** 2342–2357. MR3256391

FERREIRA, L. K. and BUSATTO, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neurosci. Biobehav. Rev.* **37** 384–400.

FERREIRA, R., XAVIER, J., COSTEIRA, J. P. and BARROSO, V. (2013). Newton algorithms for Riemannian distance related problems on connected locally symmetric manifolds. *IEEE J. Sel. Top. Signal Process.* **7** 634–645.

FISHER, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge Univ. Press, Cambridge.

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge.

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* **22** 718–726.

HIGGINS, J. J. (2004). *An Introduction to Modern Nonparametric Statistics*. Brooks/Cole, Pacific Grove, CA.

HIGHAM, N. J. (2002). Computing the nearest correlation matrix—A problem from finance. *IMA J. Numer. Anal.* **22** 329–343.

HINKLE, J., MURALIDHARAN, P., FLETCHER, P. T. and JOSHI, S. (2012). Polynomial regression on Riemannian manifolds. In *Computer Vision—ECCV* 2012 1–14. Springer, Heidelberg.

LE, H. and BARDEN, D. (2014). On the measure of the cut locus of a Fréchet mean. *Bull. Lond. Math. Soc.* **46** 698–708. MR3239607

LEE, M., SMYSER, C. and SHIMONY, J. (2013). Resting-state fMRI: A review of methods and clinical applications. *Am. J. Neuroradiol.* **34** 1866–1872.

LEHMANN, E. L. and D'ABRERA, H. J. (2006). *Nonparametrics*: *Statistical Methods Based on Ranks*. Springer, New York.

LIN, L., THOMAS, B. S., ZHU, H. and DUNSON, D. B. (2015). Extrinsic local regression on manifold-valued data. Available at arXiv:1508.02201.

MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753.

MEVEL, K., LANDEAU, B., FOUQUET, M., LA JOIE, R., VILLAIN, N., MÉZENGE, F., PERROTIN, A., EUSTACHE, F., DESGRANGES, B. and CHÉTELAT, G. (2013). Age effect on the default mode network, inner thoughts, and cognitive abilities. *Neurobiol. Aging* **34** 1292–1301.

NIETHAMMER, M., HUANG, Y. and VIALARD, F.-X. (2011). Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2011 655–662. Springer, Berlin.

ONODA, K., ISHIHARA, M. and YAMAGUCHI, S. (2012). Decreased functional connectivity by aging is associated with cognitive decline. *J. Cogn. Neurosci.* **24** 2186–2198.

PANARETOS, V. M. and ZEMEL, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44** 771–812. MR3476617

PATRANGENARU, V. and ELLINGSON, L. (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL.

PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065

PETERSEN, A. and MÜLLER, H.-G. (2019). Supplement to "Fréchet regression for random objects with Euclidean predictors." DOI:10.1214/17-AOS1624SUPP.

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

PRENTICE, M. J. (1989). Spherical regression on matched pairs of orientation statistics. *J. Roy. Statist. Soc. Ser. B* **51** 241–248. MR1007456

QI, H. and SUN, D. (2006). A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.* **28** 360–385. MR2255334

SHELINE, Y. I. and RAICHLE, M. E. (2013). Resting state functional connectivity in preclinical Alzheimer's disease. *Biological Psychiatry* **74** 340–347.

SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. and ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2009 192–199. Springer, Berlin.

STEINKE, F. and HEIN, M. (2009). Non-parametric regression between manifolds. In *Advances in Neural Information Processing Systems* 1561–1568.

STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. MR2736019

SU, J., DRYDEN, I. L., KLASSEN, E., LE, H. and SRIVASTAVA, A. (2012). Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Image Vis. Comput.* **30** 428–442.

TAKATSU, A. (2011). Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **48** 1005–1026. MR2871291

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

WANG, H. and MARRON, J. S. (2007). Object oriented data analysis: Sets of trees. *Ann. Statist.* **35** 1849–1873. MR2363955

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956

ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the* 1974 *European Meeting of Statisticians* 591–602. Springer, Berlin.

The Annals of Statistics
2019, Vol. 47, No. 2, 720–757
https://doi.org/10.1214/17-AOS1633

DIVIDE AND CONQUER IN NONSTANDARD PROBLEMS AND THE SUPER-EFFICIENCY PHENOMENON

BY MOULINATH BANERJEE[*,1], CÉCILE DUROT[†,2] AND BODHISATTVA SEN[‡,3]

University of Michigan, Université Paris Nanterre† and Columbia University‡*

We study how the divide and conquer principle works in non-standard problems where rates of convergence are typically slower than $\sqrt{n}$ and limit distributions are non-Gaussian, and provide a detailed treatment for a variety of important and well-studied problems involving nonparametric estimation of a monotone function. We find that for a fixed model, the pooled estimator, obtained by averaging nonstandard estimates across mutually exclusive subsamples, outperforms the nonstandard monotonicity-constrained (global) estimator based on the entire sample in the sense of pointwise estimation of the function. We also show that, under appropriate conditions, if the number of subsamples is allowed to increase at appropriate rates, the pooled estimator is asymptotically normally distributed with a variance that is empirically estimable from the subsample-level estimates. Further, in the context of monotone regression, we show that this gain in efficiency under a fixed model comes at a price—the pooled estimator's performance, in a *uniform sense* (maximal risk) over a class of models worsens as the number of subsamples increases, leading to a version of the super-efficiency phenomenon. In the process, we develop analytical results for the order of the bias in isotonic regression, which are of independent interest.

REFERENCES

[1] BANERJEE, M. (2005). Likelihood ratio tests under local and fixed alternatives in monotone function problems. *Scand. J. Stat.* **32** 507–525. MR2232340

[2] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693

[3] BANERJEE, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statist. Sinica* 467–492.

[4] BANERJEE, M., DUROT, C. and SEN, B. (2019). Supplement to "Divide and conquer in nonstandard problems and the super-efficiency phenomenon." DOI:10.1214/17-AOS1633SUPP.

[5] BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. MR2336859

[6] BANERJEE, M. and WELLNER, J. A. (2005). Confidence intervals for current status data. *Scand. J. Stat.* **32** 405–424. MR2204627

[7] BROWN, L. D., LOW, M. G. and ZHAO, L. H. (1997). Superefficiency in nonparametric function estimation. *Ann. Statist.* **25** 2607–2625. MR1604424

MSC2010 subject classifications. Primary 62G20, 62G08; secondary 62F30.

Key words and phrases. Cube-root asymptotics, isotonic regression, local minimax risk, non-Gaussian limit, sample-splitting.

[8] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894

[9] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* 177–197. Cambridge Univ. Press, London.

[10] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

[11] DUROT, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122** 222–240. MR1894068

[12] DUROT, C. (2008). Monotone nonparametric regression with random design. *Math. Methods Statist.* **17** 327–341. MR2483461

[13] DUROT, C. and LOPUHAÄ, H. P. (2014). A Kiefer–Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electron. J. Stat.* **8** 2479–2513.

[14] DUROT, C. and THIÉBOT, K. (2006). Bootstrapping the shorth for regression. *ESAIM Probab. Stat.* **10** 216–235.

[15] GRENANDER, U. (1956). On the theory of mortality measurement, part II. *Skand. Aktuarietidskr.* **39** 125–153.

[16] 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, Cambridge.

[17] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

[18] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* 3–33. MR1334065

[19] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391

[20] LI, R., LIN, DENNIS, K. J. and LI, B. (2013). Statistical inference in massive data sets. *Appl. Stoch. Models Bus. Ind.* **29** 399–409.

[21] MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905

[22] MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

[23] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā, Ser. A* **31** 23–36.

[24] REVUZ, D. and YOR, M. (2013). *Continuous Martingales and Brownian Motion. Grundlehren der mathematischen Wissenschaften* **293**. Springer, Berlin.

[25] 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.

[26] ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. MR0770281

[27] SHI, C., LU, W. and SONG, R. (2018). A massive data framework for M-estimators with cubic-rate. *J. Amer. Statist. Assoc.* To appear.

[28] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original; translated by Vladimir Zaiats.

[29] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

[30] ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2013). Divide and conquer kernel ridge regression. In *Conference on Learning Theory* 592–617.

[31] ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.* **44** 1400–1437. MR3519928

The Annals of Statistics
2019, Vol. 47, No. 2, 758–782
https://doi.org/10.1214/17-AOS1634

RANK VERIFICATION FOR EXPONENTIAL FAMILIES

BY KENNETH HUNG AND WILLIAM FITHIAN[1]

University of California, Berkeley

Many statistical experiments involve comparing multiple population groups. For example, a public opinion poll may ask which of several political candidates commands the most support; a social scientific survey may report the most common of several responses to a question; or, a clinical trial may compare binary patient outcomes under several treatment conditions to determine the most effective treatment. Having observed the "winner" (largest observed response) in a noisy experiment, it is natural to ask whether that candidate, survey response or treatment is actually the "best" (stochastically largest response). This article concerns the problem of *rank verification*—post hoc significance tests of whether the orderings discovered in the data reflect the population ranks. For exponential family models, we show under mild conditions that an unadjusted two-tailed pairwise test comparing the first two-order statistics (i.e., comparing the "winner" to the "runner-up") is a valid test of whether the winner is truly the best. We extend our analysis to provide equally simple procedures to obtain lower confidence bounds on the gap between the winning population and the others, and to verify ranks beyond the first.

REFERENCES

ALIKHANI, L. (2011). Study: Tween TV today is all about fame. Available at http://thechart.blogs.cnn.com/2011/08/05/study-tweens-aim-for-fame-above-all-else/.

BERGER, R. L. (1980). Minimax subset selection for the multinomial distribution. *J. Statist. Plann. Inference* **4** 391–402. MR0596773

BERGER, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24** 295–300. MR0687187

BESAG, J. and CLIFFORD, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76** 633–642.

BOFINGER, E. (1991). Selecting "demonstrably best" or "demonstrably worst" exponential populations. *Aust. N. Z. J. Stat.* **33** 183–190. MR1144383

EDWARDS, D. G. and HSU, J. C. (1983). Multiple comparisons with the best treatment. *J. Amer. Statist. Assoc.* **78** 965–971.

FINNER, H. and STRASSBURGER, K. (2002). The partitioning principle: A powerful tool in multiple decision theory. *Ann. Statist.* **30** 1194–1213. MR1926174

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

FITHIAN, W., TAYLOR, J. E. and TIBSHIRANI, R. J. (2015). Selective sequential model selection. Preprint. Available at arXiv:1512.02565.

MSC2010 subject classifications. Primary 62F07; secondary 62F03.

Key words and phrases. Ranking, selective inference, exponential family, multiple comparison, sample best.

GUPTA, S. S., HUANG, D.-Y. and PANCHAPAKESAN, S. (1984). On some inequalities and monotonicity results in selection and ranking theory. In *Inequalities in Statistics and Probability* (*Lincoln*, *Neb.*, 1982) 211–227. IMS, Hayward, CA.

GUPTA, S. S. and LIANG, T. (1989). Selecting the best binomial population: Parametric empirical Bayes approach. *J. Statist. Plann. Inference* **23** 21–31. MR1029237

GUPTA, S. S. and NAGEL, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhyā* **29** 1–34.

GUPTA, S. S. and PANCHAPAKESAN, S. (1971). On multiple decision (subset selection) procedures. Technical report, Purdue Univ., West Lafayette, IN.

GUPTA, S. S. and PANCHAPAKESAN, S. (1985). Subset selection procedures: Review and assessment. *Amer. J. Math. Management Sci.* **5** 235–311. MR0859941

GUPTA, S. S. and WONG, W.-Y. (1976). On subset selection procedures for Poisson processes and some applications to the binomial and multinomial problems. Technical report.

GUTMANN, S. and MAYMIN, Z. (1987). Is the selected population the best? *Ann. Statist.* **15** 456–461. MR0885752

HSU, J. C. (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best. *Ann. Statist.* **12** 1136–1144. MR0751303

HSU, J. (1996). *Multiple Comparisons*: *Theory and Methods*. CRC Press, Boca Raton, FL.

KANNAN, N. and PANCHAPAKESAN, S. (2009). Does the selected normal population have the smallest variance? *Amer. J. Math. Management Sci.* **29** 109–123. MR2751256

MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2011). *Inequalities*: *Theory of Majorization and Its Applications*, 2nd ed. Springer, New York. MR2759813

MAYMIN, Z. and GUTMANN, S. (1992). Testing retrospective hypotheses. *Canad. J. Statist.* **20** 335–345. MR1190577

NETTLETON, D. (2009). Testing for the supremacy of a multinomial cell probability. *J. Amer. Statist. Assoc.* **104** 1052–1059. MR2750236

NG, H. K. T. and PANCHAPAKESAN, S. (2007). Is the selected multinomial cell the best?. *Sequential Anal.* **26** 415–423. MR2359863

QUINNIPIAC UNIVERSITY POLL INSTITUTE (2016). First-timers put Trump ahead in Iowa GOP caucus, Quinnipiac University poll finds; Sanders needs first-timers to tie Clinton in Dem caucus. Available at https://poll.qu.edu/images/polling/ia/ia02012016_Ifsmb28.pdf/.

STEFANSSON, G., KIM, W.-C. and HSU, J. C. (1988). On confidence sets in multiple comparisons. In *Statistical Decision Theory and Related Topics IV* **2** 89–104. Springer, New York. MR0927125

UHLS, Y. T. and GREENFIELD, P. M. (2012). The value of fame: Preadolescent perceptions of popular media and their relationship to future aspirations. *Dev. Psychol.* **48** 315–326.

The Annals of Statistics
2019, Vol. 47, No. 2, 783–794
https://doi.org/10.1214/17-AOS1639

SUB-GAUSSIAN ESTIMATORS OF THE MEAN OF A RANDOM VECTOR

BY GÁBOR LUGOSI[*,†,‡,1] AND SHAHAR MENDELSON[§,¶,2]

ICREA, Pompeu Fabra University*†*, Barcelona GSE*‡*, Technion, I.I.T.*§* and The Australian National University*¶*

We study the problem of estimating the mean of a random vector X given a sample of N independent, identically distributed points. We introduce a new estimator that achieves a purely sub-Gaussian performance under the only condition that the second moment of X exists. The estimator is based on a novel concept of a multivariate median.

REFERENCES

[1] ALON, N., MATIAS, Y. and SZEGEDY, M. (1999). The space complexity of approximating the frequency moments. *J. Comput. System Sci.* **58** 137–147. MR1688610

[2] ALOUPIS, G. (2006). Geometric measures of data depth. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **72** 147–158. Amer. Math. Soc., Providence, RI. MR2343118

[3] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford.

[4] CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185.

[5] COHEN, M. B., LEE, Y. T., MILLER, G., PACHOCKI, J. and SIDFORD, A. (2016). Geometric median in nearly linear time. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 9–21. ACM, New York. MR3536551

[6] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gausssian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558

[7] HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864

[8] HSU, D. (2010). Robust statistics. Available at http://www.inherentuncertainty.org/2010/12/robust-statistics.html.

[9] HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18. MR3491112

[10] 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

[11] JERRUM, M. R., VALIANT, L. G. and VAZIRANI, V. V. (1986). Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* **43** 169–188. MR0855970

[12] JOLY, E., LUGOSI, G. and OLIVEIRA, R. I. (2016). On the estimation of the mean of a random vector. Preprint.

MSC2010 subject classifications. Primary 62J02, 62G08; secondary 60G25.
Key words and phrases. Mean estimation, robust estimation, sub-Gaussian inequalities.

[13] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Space*. Springer, New York.
[14] LERASLE, M. and OLIVEIRA, R. I. (2012). Robust empirical mean estimators. Available at arXiv:1112.3914.
[15] LUGOSI, G. and MENDELSON, S. (2016). Risk minimization by median-of-means tournaments. Preprint.
[16] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335.
[17] NEMIROVSKY, A. S. and YUDIN, D. B. (1983). Problem complexity and method efficiency in optimization.
[18] VAN DER WAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
[19] VERSHYNIN, R. (2009). Lectures in geometric functional analysis. Available at https://www.math.uci.edu/~rvershyn/papers/GFA-book.pdf.

The Annals of Statistics
2019, Vol. 47, No. 2, 795–827
https://doi.org/10.1214/17-AOS1650

COMBINATORIAL INFERENCE FOR GRAPHICAL MODELS

BY MATEY NEYKOV[1], JUNWEI LU[1] AND HAN LIU[2]

Carnegie Mellon University, Princeton University and Northwestern University

We propose a new family of combinatorial inference problems for graphical models. Unlike classical statistical inference where the main interest is point estimation or parameter testing, combinatorial inference aims at testing the global structure of the underlying graph. Examples include testing the graph connectivity, the presence of a cycle of certain size, or the maximum degree of the graph. To begin with, we study the information-theoretic limits of a large family of combinatorial inference problems. We propose new concepts including structural packing and buffer entropies to characterize how the complexity of combinatorial graph structures impacts the corresponding minimax lower bounds. On the other hand, we propose a family of novel and practical structural testing algorithms to match the lower bounds. We provide numerical results on both synthetic graphical models and brain networks to illustrate the usefulness of these proposed methods.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012). Detection of correlations. *Ann. Statist.* **40** 412–435. MR3014312

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241. MR3322317

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

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877

ARIAS-CASTRO, E., BUBECK, S., LUGOSI, G. and VERZELEN, N. (2015). Detecting Markov random fields hidden in white noise. Preprint. Available at arXiv:1504.06984.

BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

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

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

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

CHEN, M., REN, Z., ZHAO, H. and ZHOU, H. (2016). Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *J. Amer. Statist. Assoc.* **111** 394–406. MR3494667

MSC2010 subject classifications. 62F03, 62F04, 62H15.

Key words and phrases. Graph structural inference, minimax testing, uncertainty assessment, multiple hypothesis testing, post-regularization inference.

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

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

DUBHASHI, D. and RANJAN, D. (1998). Balls and bins: A study in negative dependence. *Random Structures Algorithms* **13** 99–124. MR1642566

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

GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional Gaussian copula graphical models. Preprint. Available at arXiv:1502.02347.

INGSTER, Y. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problemy Peredachi Informatsii* **18** 61–73. MR0689340

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131

JANKOVÁ, J. and VAN DE GEER, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* **9** 1205–1229. MR3354336

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. MR0684886

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453

LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10** 2295–2328. MR2563983

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

NEYKOV, M., LU, J. and LIU, H. (2018a). Supplement to "Combinatorial inference for graphical models." DOI:10.1214/17-AOS1650SUPP.

NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2018b). A unified theory of confidence regions and testing for high dimensional estimating equations. *Statist. Sci* **33** 427–443. MR3843384

RASKUTTI, G., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. K. (2008). Model selection in Gaussian graphical models: High-dimensional consistency of ℓ_1-regularized MLE. In *Advances in Neural Information Processing Systems* 1329–1336.

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695

ROMANO, J. P. and WOLF, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108.

SANTHANAM, N. P. and WAINWRIGHT, M. J. (2012). Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inform. Theory* **58** 4117–4134.

VERZELEN, N. and VILLERS, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.* **38** 704–752. MR2604699

WANG, W., WAINWRIGHT, M. J. and RAMCHANDRAN, K. (2010). Information-theoretic bounds on model selection for Gaussian Markov random fields. In *ISIT* 1373–1377.

YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

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

The Annals of Statistics
2019, Vol. 47, No. 2, 828–856
https://doi.org/10.1214/17-AOS1652

ESTIMATION AND PREDICTION USING GENERALIZED WENDLAND COVARIANCE FUNCTIONS UNDER FIXED DOMAIN ASYMPTOTICS

BY MORENO BEVILACQUA[*,1,4], TARIK FAOUZI[†], REINHARD FURRER[‡,2] AND EMILIO PORCU[§,¶,3,4]

Universidad de Valparaiso, Universidad del BioBio†, University of Zurich‡, Newcastle University§ and Universidad de Atacama¶*

We study estimation and prediction of Gaussian random fields with covariance models belonging to the generalized Wendland (GW) class, under fixed domain asymptotics. As for the Matérn case, this class allows for a continuous parameterization of smoothness of the underlying Gaussian random field, being additionally compactly supported. The paper is divided into three parts: first, we characterize the equivalence of two Gaussian measures with GW covariance function, and we provide sufficient conditions for the equivalence of two Gaussian measures with Matérn and GW covariance functions. In the second part, we establish strong consistency and asymptotic distribution of the maximum likelihood estimator of the microergodic parameter associated to GW covariance model, under fixed domain asymptotics. The third part elucidates the consequences of our results in terms of (misspecified) best linear unbiased predictor, under fixed domain asymptotics. Our findings are illustrated through a simulation study: the former compares the finite sample behavior of the maximum likelihood estimation of the microergodic parameter with the given asymptotic distribution. The latter compares the finite-sample behavior of the prediction and its associated mean square error when using two equivalent Gaussian measures with Matérn and GW covariance models, using covariance tapering as benchmark.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series* **55**. Superintendent of Documents, U.S. Government Printing Office, Washington, DC. MR0177136

ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. MR0611857

AMBIKASARAN, S., FOREMAN-MACKEY, D., GREENGARD, L., HOGG, D. W. and O'NEIL, M. (2016). Fast Direct Methods for Gaussian Processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 252–265.

ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700

ASKEY, R. (1973). Radial characteristic functions. Technical Report, Research Center, Univ. Wisconsin-Madison, Madison, WI.

MSC2010 subject classifications. Primary 62M30; secondary 62F12, 60G25.

Key words and phrases. Compactly supported covariance, spectral density, large dataset, microergodic parameter.

BEVILACQUA, M., FAOUZI, T., FURRER, R. and PORCU, E. (2019). Supplement to "Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics." DOI:10.1214/17-AOS1652SUPP.

CHERNIH, A. and HUBBERT, S. (2014). Closed form representations and properties of the generalised Wendland functions. *J. Approx. Theory* **177** 17–33. MR3128401

CHERNIH, A., SLOAN, I. H. and WOMERSLEY, R. S. (2014). Wendland functions with increasing smoothness converge to a Gaussian. *Adv. Comput. Math.* **40** 17–33.

DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. MR2549562

FURRER, R., BACHOC, F. and DU, J. (2016). Asymptotic properties of multivariate tapering for estimation and prediction. *J. Multivariate Anal.* **149** 177–191.

FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523.

FURRER, R. and SAIN, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *J. Stat. Softw.* **36** 1–25.

GNEITING, T. (2002a). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508.

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

GOLUBOV, B. I. (1981). On Abel–Poisson type and Riesz means. *Anal. Math.* **7** 161–184.

HIRANO, T. and YAJIMA, Y. (2013). Covariance tapering for prediction of large spatial data sets in transformed random fields. *Ann. Inst. Statist. Math.* **65** 913–939. MR3105802

IBRAGIMOV, I. D. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes. Applications of Mathematics* **9**. Springer, New York. MR0543837

KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555.

KAUFMAN, C. G. and SHABY, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100** 473–484.

MARDIA, K. V. and MARSHALL, J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

MATÉRN, B. (1960). *Spatial Variation*. Meddelanden Från Statens Skogsforskningsinstitut, Band 49, Nr 5., Stockholm.

PORCU, E., ZASTAVNYI, V. P. and BEVILACQUA, M. (2017). Buhmann covariance functions, their compact supports, and their smoothness. *Dolomites Res. Notes Approx.* **10** 33–42. MR3708624

PUTTER, H. and YOUNG, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli* **7** 421–438. MR1836738

R DEVELOPMENT CORE TEAM (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SCHABACK, R. (2011). The missing Wendland functions. *Adv. Comput. Math.* **34** 67–81.

SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.* (2) **39** 811–841.

SKOROKHOD, A. V. and YADRENKO, M. I. (1973). On absolute continuity of measures corresponding to homogeneous Gaussian fields. *Theory Probab. Appl.* **18** 27–40.

STEIN, M. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856

STEIN, M. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872. MR1056340

STEIN, M. L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statist. Probab. Lett.* **17** 399–404. MR1237787

STEIN, M. L. (1999a). Predicting random fields with increasing dense observations. *Ann. Appl. Probab.* **9** 242–273. MR1682572

STEIN, M. L. (1999b). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

STEIN, M. L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *J. Statist. Plann. Inference* **123** 1–11. MR2058118

STEIN, M. L. (2013). Statistical properties of covariance tapers. *J. Comput. Graph. Statist.* **22** 866–885. MR3173747

WACKERNAGEL, H. (2003). *Multivariate Geostatistics*: *An Introduction with Applications*, 3rd ed. Springer, New York.

WANG, D. and LOH, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Stat.* **5** 238–269. MR2792553

WENDLAND, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4** 389–396. MR1366510

WU, Z. (1995). Compactly supported positive definite radial functions. *Adv. Comput. Math.* **18** 283–292.

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I*: *Basic Results*. Springer, New York.

ZASTAVNYI, V. P. (2006). On some properties of Buhmann functions. *Ukrainian Math. J.* **58** 1184–1208.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261.

ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

The Annals of Statistics
2019, Vol. 47, No. 2, 857–883
https://doi.org/10.1214/17-AOS1665

CHEBYSHEV POLYNOMIALS, MOMENT MATCHING, AND OPTIMAL ESTIMATION OF THE UNSEEN

BY YIHONG WU AND PENGKUN YANG

Yale University and University of Illinois at Urbana-Champaign

We consider the problem of estimating the support size of a discrete distribution whose minimum nonzero mass is at least $\frac{1}{k}$. Under the independent sampling model, we show that the sample complexity, that is, the minimal sample size to achieve an additive error of εk with probability at least 0.1 is within universal constant factors of $\frac{k}{\log k}\log^2\frac{1}{\varepsilon}$, which improves the state-of-the-art result of $\frac{k}{\varepsilon^2\log k}$ in [In *Advances in Neural Information Processing Systems* (2013) 2157–2165]. Similar characterization of the minimax risk is also obtained. Our procedure is a linear estimator based on the Chebyshev polynomial and its approximation-theoretic properties, which can be evaluated in $O(n+\log^2 k)$ time and attains the sample complexity within constant factors. The superiority of the proposed estimator in terms of accuracy, computational efficiency and scalability is demonstrated in a variety of synthetic and real datasets.

REFERENCES

[1] BAR-YOSSEF, Z., JAYRAM, T., KUMAR, R., SIVAKUMAR, D. and TREVISAN, L. (2002). Counting distinct elements in a data stream. In *Proceedings of the 6th Randomization and Approximation Techniques in Computer Science* 1–10. Springer, Berlin.

[2] BUNGE, J. and FITZPATRICK, M. (1993). Estimating the number of species: A review. *J. Amer. Statist. Assoc.* **88** 364–373.

[3] BURNHAM, K. P. and OVERTON, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60** 927–936.

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

[5] CHAO, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* 265–270. MR0793175

[6] CHAO, A. and LEE, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Amer. Statist. Assoc.* **87** 210–217. MR1158639

[7] CHARIKAR, M., CHAUDHURI, S., MOTWANI, R. and NARASAYYA, V. (2000). Towards estimation error guarantees for distinct values. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (*PODS*) 268–279. ACM, New York.

[8] DARROCH, J. and RATCLIFF, D. (1980). A note on capture-recapture estimation. *Biometrics* **36** 149–153.

[9] DZYADYK, V. K. and SHEVCHUK, I. A. (2008). *Theory of Uniform Approximation of Functions by Polynomials*. de Gruyter, Berlin.

MSC2010 subject classifications. Primary 62G05; secondary 62C20.

Key words and phrases. Support size estimation, large domain, polynomial approximation, high-dimensional statistics, nonparametric inference.

[10] EFRON, B. and THISTED, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.

[11] FISHER, R. A., CORBET, A. S. and WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12** 42–58.

[12] GANDOLFI, A. and SASTRI, C. C. A. (2004). Nonparametric estimations about species not observed in a random sample. *Milan J. Math.* **72** 81–105. MR2099128

[13] Global Language Monitor. Number of words in the English language. https://www.languagemonitor.com/global-english/no-of-words/. Accessed: 2016-02-16.

[14] GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264.

[15] GOOD, I. J. and TOULMIN, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63.

[16] GOTELLI, N. J. and COLWELL, R. K. (2011). Estimating species richness. *Biological Diversity*: *Frontiers in Measurement and Assessment* **12** 39–54.

[17] HARRIS, B. (1968). Statistical inference in the classical occupancy problem unbiased estimation of the number of classes. *J. Amer. Statist. Assoc.* **63** 837–847. MR0231480

[18] HUANG, S.-P. and WEIR, B. (2001). Estimating the total number of alleles using a sample coverage method. *Genetics* **159** 1365–1373.

[19] IBRAGIMOV, I. A., NEMIROVSKII, A. S. and KHAS'MINSKII, R. Z. (1987). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

[20] JIAO, J., VENKAT, K., HAN, Y. and WEISSMAN, T. (2015). Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inform. Theory* **61** 2835–2885.

[21] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253.

[22] LEWONTIN, R. C. and PROUT, T. (1956). Estimation of the number of different classes in a population. *Biometrics* **12** 211–223.

[23] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

[24] MAO, C. X. and LINDSAY, B. G. (2007). Estimating the number of classes. *Ann. Statist.* **35** 917–930. MR2336874

[25] MARCHAND, J. and SCHROECK JR, F. (1982). On the estimation of the number of equally likely classes in a population. *Comm. Statist. Theory Methods* **11** 1139–1146. MR0652619

[26] MCNEIL, D. R. (1973). Estimating an author's vocabulary. *J. Amer. Statist. Assoc.* **68** 92–96. MR0373124

[27] MITZENMACHER, M. and UPFAL, E. (2005). *Probability and Computing*: *Randomized Algorithms and Probabilistic Analysis*. Cambridge Univ. Press, Cambridge. MR2144605

[28] ORLITSKY, A., SURESH, A. T. and WU, Y. (2016). Optimal prediction of the number of unseen species. *Proc. Natl. Acad. Sci. USA* **113** 13283–13288.

[29] Oxford English Dictinary. http://public.oed.com/about/. Accessed: 2016-02-16.

[30] RASKHODNIKOVA, S., RON, D., SHPILKA, A. and SMITH, A. (2009). Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.* **39** 813–842.

[31] ROBBINS, H. E. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Stat.* **39** 256–257.

[32] SAMUEL, E. (1968). Sequential maximum likelihood estimation of the size of a population. *Ann. Math. Stat.* **39** 1057–1068.

[33] STEELE, J. M. (1986). An Efron–Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14** 753–758. MR0840528

[34] THISTED, R. and EFRON, B. (1987). Did Shakespeare write a newly-discovered poem? *Biometrika* **74** 445–455.

[35] TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable*. Pergamon Press, Elmsford, NY.

[36] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer, New York, NY. MR2724359

[37] VALIANT, G. (2017). Private communication.

[38] VALIANT, G. and VALIANT, P. (2010). A CLT and tight lower bounds for estimating entropy. In *Electronic Colloquium on Computational Complexity* (*ECCC*) **17** 179.

[39] VALIANT, G. and VALIANT, P. (2011). Estimating the unseen: An $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the* 43*rd Annual ACM Symposium on Theory of Computing* 685–694.

[40] VALIANT, G. and VALIANT, P. (2011). The power of linear estimators. In *Foundations of Computer Science* (*FOCS*), 2011 *IEEE* 52*nd Annual Symposium on* 403–412. IEEE, New York.

[41] VALIANT, P. and VALIANT, G. (2013). Estimating the unseen: Improved estimators for entropy and other properties. In *Advances in Neural Information Processing Systems* 2157–2165.

[42] WU, Y. and YANG, P. (2016). Sample complexity of the distinct elements problem. Available at arXiv:1612.03375.

[43] WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inform. Theory* **62** 3702–3720.

[44] WU, Y. and YANG, P. (2017). Supplement to "Chebyshev polynomials, moment matching and optimal estimation of the unseen." DOI:10.1214/17-AOS1665SUPP.

The Annals of Statistics
2019, Vol. 47, No. 2, 884–908
https://doi.org/10.1214/18-AOS1681

PARTIAL LEAST SQUARES PREDICTION IN HIGH-DIMENSIONAL REGRESSION

BY R. DENNIS COOK AND LILIANA FORZANI

University of Minnesota and Facultad de Ingeniería Química, UNL, Researcher of CONICET

We study the asymptotic behavior of predictions from partial least squares (PLS) regression as the sample size and number of predictors diverge in various alignments. We show that there is a range of regression scenarios where PLS predictions have the usual root-n convergence rate, even when the sample size is substantially smaller than the number of predictors, and an even wider range where the rate is slower but may still produce practically useful results. We show also that PLS predictions achieve their best asymptotic behavior in abundant regressions where many predictors contribute information about the response. Their asymptotic behavior tends to be undesirable in sparse regressions where few predictors contribute information about the response.

REFERENCES

[1] ABUDU, S., KING, P. and PAGANO, T. C. (2010). Application of partial least-squares regression in seasonal streamflow forecasting. *J. Hydrol. Eng.* **15** 612–623.

[2] BIANCOLILLO, A., BUCCI, R., MAGRÌ, A. L., MAGRÌ, A. D. and MARINI, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Anal. Chim. Acta* **820** 23–31.

[3] BOULESTEIX, A.-L. and STRIMMER, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.* **8** 32–44.

[4] BRO, R. and EELDÉN, L. (2009). PLS works. *J. Chemom.* **23** 69–71.

[5] CASTEJÒN, D., GARCÌA-SEGURA, J. M., ESCUDERO, R., HERRERA, A. and CAMBERO, M. I. (2015). Metabolomics of meat exudate: Its potential to evaluate beef meat conservation and aging. *Anal. Chim. Acta* **901** 1–11.

[6] CHUN, H. and KELEŞ, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 3–25. MR2751241

[7] COOK, R. D. (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In *Proceedings of the Section on Engineering and Physical Sciences* 18–25. Amer. Statist. Assoc., Alexandria, VA.

[8] COOK, R. D. (1998). *Regression Graphics*: *Ideas for Studying Regressions through Graphics*. Wiley, New York. MR1645673

[9] COOK, R. D. and FORZANI, L. (2017). Big data and partial least squares prediction. *Canad. J. Statist.* **46** 62–78.

[10] COOK, R. D. and FORZANI, L. (2018). Supplement to "Partial least squares prediction in high-dimensional regression." DOI:10.1214/18-AOS1681SUPP.

MSC2010 subject classifications. Primary 62J05; secondary 62F12.
Key words and phrases. Abundant regressions, dimension reduction, sparse regressions.

[11] COOK, R. D., FORZANI, L. and ROTHMAN, A. J. (2013). Prediction in abundant high-dimensional linear regression. *Electron. J. Stat.* **7** 3059–3088. MR3151762

[12] COOK, R. D., HELLAND, I. S. and SU, Z. (2013). Envelopes and partial least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 851–877. MR3124794

[13] COOK, R. D., LI, B. and CHIAROMONTE, F. (2007). Dimension reduction in regression without matrix inversion. *Biometrika* **94** 569–584. MR2410009

[14] COOK, R. D., LI, B. and CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica* **20** 927–960. MR2729839

[15] DE JONG, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **18** 251–263.

[16] DELAIGLE, A. and HALL, P. (2012). Methodology and theory for partial least squares applied to functional data. *Ann. Statist.* **40** 322–352. MR3014309

[17] ECK, D. J. and COOK, R. D. (2017). Weighted envelope estimation to handle variability in model selection. *Biometrika* **104** 743–749. MR3694595

[18] FRANK, I. E. and FRIDEMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 102–246.

[19] GARTHWAITE, P. H. (1994). An interpretation of partial least squares. *J. Amer. Statist. Assoc.* **89** 122–127. MR1266290

[20] GOICOECHEA, H. C. and OLIVER, A. C. (1999). Enhanced synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Anal. Chem.* **71** 4361–4368.

[21] HELLAND, I. S. (1990). Partial least squares regression and statistical models. *Scand. J. Stat.* **17** 97–114. MR1085924

[22] HELLAND, I. S. (1992). Maximum likelihood regression on relevant components. *J. Roy. Statist. Soc. Ser. B* **54** 637–647. MR1160488

[23] HELLAND, I. S. (2001). Some theoretical aspects of partial least squares regression. *Chemom. Intell. Lab. Syst.* **58** 97–107.

[24] KANDEL, T. A., GISLUM, R., JØRGENSEN, U. and LÆRKE, P. E. (2013). Prediction of biogas yield and its kinetics in reed canary grass using near infrared reflectance spectroscopy and chemometrics. *Bioresour. Technol.* **146** 282–287.

[25] KOCH, C., POSCH, A. E., GOICOECHEA, H. C., HERWIG, C. and LENDLA, B. (2013). Multi-analyte quantification in bioprocesses by Fourier-transform-infrared spectroscopy by partial least squares regression and multivariate curve resolution. *Anal. Chim. Acta* **807** 103–110.

[26] LI, W., CHENG, Z., WANG, Y. and QU, H. (2013). Quality control of *Lonicerae Japonicae Flos* using near infrared spectroscopy and chemometrics. *J. Pharm. Biomed. Anal.* **72** 33–39.

[27] LOBAUGH, N. J., WEST, R. and MCINTOSH, A. R. (2001). Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* **38** 517–530.

[28] MARTENS, H. and NÆS, T. (1992). *Multivariate Calibration*. Wiley, Chichester. MR1029523

[29] NÆS, T. and HELLAND, I. S. (1993). Relevant components in regression. *Scand. J. Stat.* **20** 239–250. MR1241390

[30] NAIK, P. and TSAI, C.-L. (2000). Partial least squares estimator for single-index models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 763–771. MR1796290

[31] NGUYEN, D. V. and ROCKE, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* **18** 39–50.

[32] NGUYEN, D. V. and ROCKE, D. M. (2004). On partial least squares dimension reduction for microarray-based classification: A simulation study. *Comput. Statist. Data Anal.* **46** 407–425. MR2067030

[33] SCHWARTZ, R. W., KEMBHAVI, A., HARWOOD, D. and DAVIS, L. S. (2009). Human detection using partial least squares analysis. In 2009 *IEEE* 12*th International Conference on Computer Vision* 24–31.

[34] TER BRAAK, C. J. F. and DE JONG, S. (1998). The objective function of partial least squares regression. *J. Chemom*. **12** 41–54.

[35] WOLD, S., MARTENS, H. and WOLD, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In *Proceedings of the Conference on Matrix Pencils* (A. Ruhe and B. Kågström, eds.). *Lecture Notes in Math*. **973** 286–293. Springer, Heidelberg.

[36] WORSLEY, K. J. (1997). An overview and some new developments in the statistical analysis of PET and fMRI data. *Hum. Brain Mapp*. **5** 254–258.

The Annals of Statistics
2019, Vol. 47, No. 2, 909–935
https://doi.org/10.1214/18-AOS1682

SIGNAL ALIASING IN GAUSSIAN RANDOM FIELDS FOR EXPERIMENTS WITH QUALITATIVE FACTORS

BY MING-CHUNG CHANG*, SHAO-WEI CHENG[†,1] AND CHING-SHUI CHENG[‡,§]

National Central University, National Tsing Hua University†, Academia Sinica‡ and University of California, Berkeley§*

Signal aliasing is an inevitable consequence of using fractional factorial designs. Unlike linear models with fixed factorial effects, for Gaussian random field models advocated in some Bayesian design and computer experiment literature, the issue of signal aliasing has not received comparable attention. In the present article, this issue is tackled for experiments with qualitative factors. The signals in a Gaussian random field can be characterized by the random effects identified from the covariance function. The aliasing severity of the signals is determined by two key elements: (i) the aliasing pattern, which depends only on the chosen design, and (ii) the effect priority, which is related to the variances of the random effects and depends on the model parameters. We first apply this framework to study the signal-aliasing problem for regular fractional factorial designs. For general factorial designs including nonregular ones, we propose an aliasing severity index to quantify the severity of signal aliasing. We also observe that the aliasing severity index is highly correlated with the prediction variance.

REFERENCES

BAILEY, R. A. (1982). The decomposition of treatment degrees of freedom in quantitative factorial experiments. *J. Roy. Statist. Soc. Ser. B* **44** 63–70. MR0655375

BAILEY, R. A. (1996). Orthogonal partitions in designed experiments. *Des. Codes Cryptogr.* **8** 45–77. MR1393974

BAILEY, R. A. (2008). *Design of Comparative Experiments*. *Cambridge Series in Statistical and Probabilistic Mathematics* **25**. Cambridge Univ. Press, Cambridge. MR2422352

CHANG, M. C. (2015). Effect aliasing in Gaussian random field models. Ph.D. dissertation, Institute of Statistics, National Tsing Hua Univ.

CHANG, M. C. and CHENG, S. W. (2018). Signal aliasing in Gaussian random fields for experiments with quantitative factors. *Ann. Statist.* To appear.

CHENG, C. S. (2014). *Theory of Factorial Design*: *Single- and Multi-Stratum Experiments*. CRC Press, Boca Raton, FL.

CHENG, C.-S., STEINBERG, D. M. and SUN, D. X. (1999). Minimum aberration and model robustness for two-level fractional factorial designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 85–93. MR1664104

CHENG, S.-W. and YE, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Ann. Statist.* **32** 2168–2185. MR2102507

MSC2010 subject classifications. 62K15.

Key words and phrases. Bayesian design, computer experiment, eigen-decomposition, fixed-effect model, fractional factorial, kriging model, random-effect model.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. Revised reprint of the 1991 edition. MR1239641

DENG, X., HUNG, Y. and LIN, C. D. (2015). Design for computer experiments with qualitative and quantitative factors. *Statist. Sinica* **25** 1567–1581. MR3409081

FRIES, A. and HUNTER, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22** 601–608. MR0596803

HAN, G., SANTNER, T. J., NOTZ, W. I. and BARTEL, D. L. (2009). Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics* **51** 278–288. MR2751072

JOSEPH, V. R. (2006). A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics* **48** 219–229. MR2277676

JOSEPH, V. R. and DELANEY, J. D. (2007). Functionally induced priors for the analysis of experiments. *Technometrics* **49** 1–11. MR2345447

JOSEPH, V. R., HUNG, Y. and SUDJIANTO, A. (2008). Blind kriging: A new method for developing metamodels. *J. Mech. Des.* **130** 1–8.

KERR, M. K. (2001). Bayesian optimal fractional factorials. *Statist. Sinica* **11** 605–630. MR1863153

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications. Mathematics in Science and Engineering* **143**. Academic Press, New York. MR0552278

MITCHELL, T. J., MORRIS, M. D. and YLVISAKER, D. (1995). Two-level fractional factorials and Bayesian prediction. *Statist. Sinica* **5** 559–573. MR1347607

QIAN, P. Z. G., WU, H. and WU, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50** 383–396. MR2457574

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. MR2160708

SEBER, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley, Hoboken, NJ. MR2365265

STEINBERG, D. M. and BURSZTYN, D. (2004). Data analytic tools for understanding random field regression models. *Technometrics* **46** 411–420. MR2101509

WELCH, W. J., BUCK, R. J., SACKS, J., WYNN, H. P., MITCHELL, T. J. and MORRIS, M. D. (1992). Screening, predicting and computer experiments. *Technometrics* **34** 15–25.

WU, C. F. J. and HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd ed. Wiley, Hoboken, NJ. MR2583259

XU, H. and WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs [corrected republication of MR1863969]. *Ann. Statist.* **29** 1066–1077. MR1869240

ZHOU, Q., QIAN, P. Z. G. and ZHOU, S. (2011). A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics* **53** 266–273. MR2857704

The Annals of Statistics
2019, Vol. 47, No. 2, 936–964
https://doi.org/10.1214/18-AOS1694

CROSS: EFFICIENT LOW-RANK TENSOR COMPLETION

BY ANRU ZHANG[1]

University of Wisconsin-Madison

The completion of tensors, or high-order arrays, attracts significant attention in recent research. Current literature on tensor completion primarily focuses on recovery from a set of uniformly randomly measured entries, and the required number of measurements to achieve recovery is not guaranteed to be optimal. In addition, the implementation of some previous methods are NP-hard. In this article, we propose a framework for low-rank tensor completion via a novel tensor measurement scheme that we name Cross. The proposed procedure is efficient and easy to implement. In particular, we show that a third-order tensor of Tucker rank-(r_1, r_2, r_3) in p_1-by-p_2-by-p_3 dimensional space can be recovered from as few as $r_1r_2r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)$ noiseless measurements, which matches the sample complexity lower bound. In the case of noisy measurements, we also develop a theoretical upper bound and the matching minimax lower bound for recovery error over certain classes of low-rank tensors for the proposed procedure. The results can be further extended to fourth or higher-order tensors. Simulation studies show that the method performs well under a variety of settings. Finally, the procedure is illustrated through a real dataset in neuroimaging.

REFERENCES

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

BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In 29*th Annual Conference on Learning Theory* 417–445.

BHOJANAPALLI, S. and SANGHAVI, S. (2015). A new sampling technique for tensors. Preprint. Available at arXiv:1502.05023.

CAI, T., CAI, T. T. and ZHANG, A. (2016). Structured matrix completion with applications to genomic data integration. *J. Amer. Statist. Assoc.* **111** 621–633. MR3538692

CAI, T. T. and ZHOU, W.-X. (2016). Matrix completion via max-norm constrained optimization. *Electron. J. Stat.* **10** 1493–1525. MR3507371

CAIAFA, C. F. and CICHOCKI, A. (2010). Generalizing the column-row matrix decomposition to multi-way arrays. *Linear Algebra Appl.* **433** 557–573. MR2653820

CAIAFA, C. F. and CICHOCKI, A. (2015). Stable, robust, and super fast reconstruction of tensors using multi-way projections. *IEEE Trans. Signal Process.* **63** 780–793. MR3299461

CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory* **57** 2342–2359. MR2809094

MSC2010 subject classifications. Primary 62H12; secondary 62C20.

Key words and phrases. Cross tensor measurement, denoising, minimax rate-optimal, neuroimaging, tensor completion.

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **56** 2053–2080. MR2723472

CAO, Y. and XIE, Y. (2016). Poisson matrix recovery and completion. *IEEE Trans. Signal Process.* **64** 1609–1620. MR3548877

CAO, Y., ZHANG, A. and LI, H. (2017). Multi-sample Estimation of Bacterial Composition Matrix in Metagenomics Data. Preprint. Available at arXiv:1706.02380.

GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010. MR2765628

GUHANIYOGI, R., QAMAR, S. and DUNSON, D. B. (2017). Bayesian tensor regression. *J. Mach. Learn. Res.* **18** 79. MR3714242

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** 45. MR3144915

JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems* **1** 1431–1439. MIT Press, Cambridge, MA.

JIANG, X., RASKUTTI, G. and WILLETT, R. (2015). Minimax optimal rates for Poisson inverse problems with physical constraints. *IEEE Trans. Inform. Theory* **61** 4458–4474. MR3372365

JOHNDROW, J. E., BHATTACHARYA, A. and DUNSON, D. B. (2017). Tensor decompositions and sparse log-linear models. *Ann. Statist.* **45** 1–38. MR3611485

KARATZOGLOU, A., AMATRIAIN, X., BALTRUNAS, L. and OLIVER, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 79–86. ACM, New York.

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inform. Theory* **56** 2980–2998. MR2683452

KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056

KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869

KRESSNER, D., STEINLECHNER, M. and VANDEREYCKEN, B. (2014). Low-rank tensor completion by Riemannian optimization. *BIT* **54** 447–468. MR3223510

KRISHNAMURTHY, A. and SINGH, A. (2013). Low-rank matrix and tensor completion via adaptive sampling. In *Advances in Neural Information Processing Systems* 836–844.

LI, N. and LI, B. (2010). Tensor completion for on-board compression of hyperspectral images. In 2010 *IEEE International Conference on Image Processing* 517–520. IEEE, New York.

LI, L. and ZHANG, X. (2017). Parsimonious tensor response regression. *J. Amer. Statist. Assoc.* **112** 1131–1146. MR3735365

LI, X., ZHOU, H. and LI, L. (2013). Tucker tensor regression and neuroimaging analysis. Preprint. Available at arXiv:1304.5637.

LI, L., CHEN, Z., WANG, G., CHU, J. and GAO, H. (2014). A tensor PRISM algorithm for multi-energy CT reconstruction and comparative studies. *J. X-Ray Sci. Technol.* **22** 147–163.

LIU, J., MUSIALSKI, P., WONKA, P. and YE, J. (2013). Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 208–220.

MAHONEY, M. W., MAGGIONI, M. and DRINEAS, P. (2008). Tensor-CUR decompositions for tensor-based data. *SIAM J. Matrix Anal. Appl.* **30** 957–987. MR2447438

MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML* 73–81.

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

NOWAK, R. D. and KOLACZYK, E. D. (2000). A statistical multiscale framework for Poisson inverse problems. *IEEE Trans. Inform. Theory* **46** 1811–1825. MR1790322

OSELEDETS, I. V., SAVOSTIANOV, D. V. and TYRTYSHNIKOV, E. E. (2008). Tucker dimensionality reduction of three-dimensional arrays in linear time. *SIAM J. Matrix Anal. Appl.* **30** 939–956. MR2447437

OSELEDETS, I. V. and TYRTYSHNIKOV, E. E. (2009). Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.* **31** 3744–3759. MR2556560

PIMENTEL-ALARCÓN, D. L., BOSTON, N. and NOWAK, R. D. (2016). A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE J. Sel. Top. Signal Process.* **10** 623–636.

RASKUTTI, G., YUAN, M. and CHEN, H. (2017). Convex regularization for high-dimensional multi-response tensor regression. Preprint. Available at arXiv:1512.01215v2.

RAUHUT, H., SCHNEIDER, R. and STOJANAC, Ž. (2017). Low rank tensor recovery via iterative hard thresholding. *Linear Algebra Appl.* **523** 220–262. MR3624675

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

RENDLE, S. and SCHMIDT-THIEME, L. (2010). Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* 81–90. ACM, New York.

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342

RUDELSON, M. and VERSHYNIN, R. (2007). Sampling from large matrices: An approach through geometric functional analysis. *J. ACM* **54** 21. MR2351844

SEMERCI, O., HAO, N., KILMER, M. E. and MILLER, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23** 1678–1693. MR3191324

SHAH, P., RAO, N. and TANG, G. (2015). Optimal low-rank tensor recovery from separable measurements: Four contractions suffice. Preprint. Available at arXiv:1505.04085.

SREBRO, N. and SHRAIBMAN, A. (2005). Rank, trace-norm and max-norm. In *Learning Theory. Lecture Notes in Computer Science* **3559** 545–560. Springer, Berlin. MR2203286

SUN, W. W. and LI, L. (2016). Sparse low-rank tensor response regression. Preprint. Available at arXiv:1609.04523.

SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 899–916. MR3641413

TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395

WAGNER, A. and ZUK, O. (2015). Low-rank matrix recovery from row-and-column affine measurements. In *Proceedings of the* 32*nd International Conference on Machine Learning* (*ICML*-15) 2012–2020.

WANG, Y. and SINGH, A. (2015). Provably correct algorithms for matrix column subset selection with selectively sampled data. Preprint. Available at arXiv:1505.04343.

WETZSTEIN, G., LANMAN, D., HIRSCH, M. and RASKAR, R. (2012). Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting.

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132

YUAN, M. and ZHANG, C.-H. (2017). Incoherent tensor norms and their applications in higher order tensor completion. *IEEE Trans. Inform. Theory* **63** 6753–6766. MR3707566

ZHANG, A. (2019). Supplement to "Cross: Efficient low-rank tensor completion." DOI:10.1214/18-AOS1694SUPP.

ZHANG, A. and XIA, D. (2017). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inform. Theory* **64** 7311–7338.

ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.* **108** 540–552. MR3174640

The Annals of Statistics
2019, Vol. 47, No. 2, 965–993
https://doi.org/10.1214/18-AOS1698

COVARIATE BALANCING PROPENSITY SCORE BY TAILORED LOSS FUNCTIONS[1]

BY QINGYUAN ZHAO

University of Pennsylvania

In observational studies, propensity scores are commonly estimated by maximum likelihood but may fail to balance high-dimensional pretreatment covariates even after specification search. We introduce a general framework that unifies and generalizes several recent proposals to improve covariate balance when designing an observational study. Instead of the likelihood function, we propose to optimize special loss functions—covariate balancing scoring rules (CBSR)—to estimate the propensity score. A CBSR is uniquely determined by the link function in the GLM and the estimand (a weighted average treatment effect). We show CBSR does not lose asymptotic efficiency in estimating the weighted average treatment effect compared to the Bernoulli likelihood, but CBSR is much more robust in finite samples. Borrowing tools developed in statistical learning, we propose practical strategies to balance covariate functions in rich function classes. This is useful to estimate the maximum bias of the inverse probability weighting (IPW) estimators and construct honest confidence intervals in finite samples. Lastly, we provide several numerical examples to demonstrate the tradeoff of bias and variance in the IPW-type estimators and the tradeoff in balancing different function classes of the covariates.

REFERENCES

[1] ABADIE, A. and IMBENS, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74** 235–267. MR2194325

[2] ATHEY, S., IMBENS, G. W., WAGER, S. et al. (2016). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623.

[3] AUSTIN, P. C. and STUART, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat. Med.* **34** 3661–3679. MR3422140

[4] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[5] BUJA, A., STUETZLE, W. and SHEN, Y. (2005). Loss functions for binary class probability estimation and classification: Structure and applications. Working Draft.

[6] CALIENDO, M. and KOPEINIG, S. (2008). Some practical guidance for the implementation of propensity score matching. *J. Econ. Surv.* **22** 31–72.

[7] CHAN, K. C. G., YAM, S. C. P. and ZHANG, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 673–700. MR3506798

MSC2010 subject classifications. Primary 62P10; secondary 62C99.

Key words and phrases. Convex optimization, kernel method, inverse probability weighting, proper scoring rule, regularized regression, statistical decision theory.

[8] COCHRAN, W. G. (1953). Matching in analytical studies. *Am. J. Public Health Nation's Health* **43** 684–691.

[9] COCHRAN, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24** 295–313. MR0228136

[10] CRUMP, R., HOTZ, V. J., IMBENS, G. and MITNIK, O. (2006). Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand. Technical Report No. 330, National Bureau of Economic Research, Cambridge, MA.

[11] DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

[12] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

[13] FAN, J., IMAI, K., LIU, H., NING, Y. and YANG, X. (2016). Improving covariate balancing propensity score: A doubly robust and efficient approach. Technical Report, Princeton Univ., Princeton, NJ.

[14] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting. *Ann. Statist.* **28** 337–407. MR1790002

[15] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328

[16] GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

[17] GRAHAM, B. S., DE XAVIER PINTO, C. C. and EGEL, D. (2012). Inverse probability tilting for moment condition model with missing data. *Rev. Econ. Stud.* **79** 1053–1079. MR2986390

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

[19] HAINMUELLER, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46. DOI:10.1093/pan/mpr025.

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

[21] HAZLETT, C. (2016). Kernel balancing: A flexible non-parametric weighting procedure for estimating causal effects. Available at arXiv:1605.00155.

[22] HECKMAN, J. J., ICHIMURA, H. and TODD, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econ. Stud.* **64** 605–654.

[23] HIRANO, K. and IMBENS, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Serv. Outcomes Res. Methodol.* **2** 259–278. DOI:10.1023/A:1020371312283.

[24] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826

[25] HOFMANN, T., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Kernel methods in machine learning. *Ann. Statist.* **36** 1171–1220. MR2418654

[26] IMAI, K., KING, G. and STUART, E. A. (2008). Misunderstanding between experimentalists and observationalists about causal inference. *J. Roy. Statist. Soc. Ser. A* **171** 481–502. MR2427345

[27] IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. MR3153941

[28] IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

[29] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951

[30] JANSON, L., FOYGEL BARBER, R. and CANDÈS, E. (2017). EigenPrism: Inference for high dimensional signal-to-noise ratios. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1037–1065. MR3689308

[31] KALLUS, N. (2016). Generalized optimal matching methods for causal inference. Available at arXiv:1612.08321.

[32] KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458

[33] LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2016). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473

[34] LUNCEFORD, J. K. and DAVIDIAN, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Stat. Med.* **23** 2937–2960.

[35] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London. MR3223057

[36] MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938

[37] NORMAND, S.-L. T., LANDRUM, M. B., GUADAGNOLI, E., AYANIAN, J. Z., RYAN, T. J., CLEARY, P. D. and MCNEIL, B. J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *J. Clin. Epidemiol.* **54** 387–398.

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

[39] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

[40] ROSENBAUM, P. R. and RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79** 516–524.

[41] ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33–38.

[42] RUBIN, D. B. (1973). Matching to remove bias in observational studies. *Biometrics* 159–183.

[43] RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2** 808–804. MR2516795

[44] RUBIN, D. B. (2009). Author's reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Stat. Med.* **28** 1420–1423. MR2724703

[45] SAVAGE, L. J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801. MR0331571

[46] SMITH, J. A. and TODD, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *J. Econometrics* **125** 305–353. MR2143379

[47] STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812

[48] WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442

[49] WANG, Y. and ZUBIZARRETA, J. R. (2017). Approximate balancing weights: Characterizations from a shrinkage estimation perspective. Available at arXiv:1705.00998.

[50] WONG, R. K. W. and CHAN, K. C. G. (2018). Kernel-based covariate functional balancing for observational studies. *Biometrika* **105** 199–213. MR3768874

[51] ZHAO, Q. (2018). Supplement to "Covariate balancing propensity score by tailored loss functions." DOI:10.1214/18-AOS1698SUPP.

[52] ZHAO, Q. and PERCIVAL, D. (2017). Entropy balancing is doubly robust. *J. Causal Inference* **5**.

[53] ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672

The Annals of Statistics
2019, Vol. 47, No. 2, 994–1024
https://doi.org/10.1214/18-AOS1701

THE GEOMETRY OF HYPOTHESIS TESTING OVER CONVEX CONES: GENERALIZED LIKELIHOOD RATIO TESTS AND MINIMAX RADII

BY YUTING WEI, MARTIN J. WAINWRIGHT[1] AND ADITYANAND GUNTUBOYINA[2]

University of California, Berkeley

We consider a compound testing problem within the Gaussian sequence model in which the null and alternative are specified by a pair of closed, convex cones. Such cone testing problem arises in various applications, including detection of treatment effects, trend detection in econometrics, signal detection in radar processing and shape-constrained inference in nonparametric statistics. We provide a sharp characterization of the GLRT testing radius up to a universal multiplicative constant in terms of the geometric structure of the underlying convex cones. When applied to concrete examples, this result reveals some interesting phenomena that do not arise in the analogous problems of estimation under convex constraints. In particular, in contrast to estimation error, the testing error no longer depends purely on the problem complexity via a volume-based measure (such as metric entropy or Gaussian complexity); other geometric properties of the cones also play an important role. In order to address the issue of optimality, we prove information-theoretic lower bounds for the minimax testing radius again in terms of geometric quantities. Our general theorems are illustrated by examples including the cases of monotone and orthant cones, and involve some results of independent interest.

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

[2] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[3] 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, New York. MR0326887

[4] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46** 36–48. MR0104312

[5] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. II. *Biometrika* **46** 328–335. MR0112204

[6] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554

MSC2010 subject classifications. Primary 62F03; secondary 52A05.

Key words and phrases. Hypothesis testing, closed convex cone, likelihood ratio test, minimax rate, Gaussian complexity.

[7] BATU, T., KUMAR, R. and RUBINFELD, R. (2004). Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing* 381–390. ACM, New York. MR2121623

[8] BESSON, O. (2006). Adaptive detection of a signal whose signature belongs to a cone. In *Proceedings SAM Conference*.

[9] BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. MR3102549

[10] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982

[11] CHEN, D. and PLEMMONS, R. J. (2010). Nonnegativity constraints in numerical analysis. In *The Birth of Numerical Analysis* 109–139. World Sci. Publ., Hackensack, NJ. MR2604144

[12] CHETVERIKOV, D. (2012). Testing regression monotonicity in econometric models Technical report, UCLA. Available at arXiv:1212.6757.

[13] DYKSTRA, R. L. and ROBERTSON, T. (1983). On testing monotone tendencies. *J. Amer. Statist. Assoc.* **78** 342–350. MR0711109

[14] EDWARDS, A. W. F. (1984). *Likelihood*. Cambridge Univ. Press, Cambridge. Reprint of the 1972 original. MR0788494

[15] ERMAKOV, M. S. (1991). Minimax detection of a signal in Gaussian white noise. *Theory Probab. Appl.* **35** 667–679. MR1090496

[16] FAN, J. and JIANG, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *TEST* **16** 409–444. MR2365172

[17] FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

[18] GRECO, M., GINI, F. and FARINA, A. (2008). Radar detection and classification of jamming signals belonging to a cone class. *IEEE Trans. Signal Process.* **56** 1984–1993. MR2518695

[19] HU, X. and WRIGHT, F. T. (1994). Likelihood ratio tests for a class of nonoblique hypotheses. *Ann. Inst. Statist. Math.* **46** 137–145. MR1272754

[20] INGSTER, YU. I. (1987). A minimax test of nonparametric hypotheses on the density of a distribution in L_p metrics. *Theory Probab. Appl.* **31** 333–337. MR0851000

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

[22] KOLTCHINSKII, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory* **47** 1902–1914. MR1842526

[23] KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50** 403–418. MR0163386

[24] LEHMANN, E. L. (2012). On likelihood ratio tests. In *Selected Works of E. L. Lehmann* (J. Rojo, ed.) 209–216. Springer, New York. MR2920510

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

[26] LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. MR1681702

[27] LEPSKI, O. V. and TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117** 17–48. MR1759508

[28] MCCOY, M. B. and TROPP, J. A. (2014). From Steiner formulas for cones to concentration of intrinsic volumes. *Discrete Comput. Geom.* **51** 926–963. MR3216671

[29] MEINSHAUSEN, N. (2013). Sign-constrained least squares estimation for high-dimensional regression. *Electron. J. Stat.* **7** 1607–1631. MR3066380

[30] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Dominance of likelihood ratio tests under cone constraints. *Ann. Statist.* **20** 2087–2099. MR1193327

[31] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Testing nonoblique hypotheses. *Comm. Statist. Theory Methods* **21** 471–484. MR1158571

[32] MENÉNDEZ, J. A. and SALVADOR, B. (1991). Anomalies of the likelihood ratio tests for testing restricted hypotheses. *Ann. Statist.* **19** 889–898. MR1105850

[33] MEYER, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika* **90** 223–232. MR1966562

[34] MOREAU, J.-J. (1962). Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *C. R. Acad. Sci. Paris* **255** 238–240. MR0139919

[35] PERLMAN, M. D. and WU, L. (1999). The emperor's new tests. *Statist. Sci.* **14** 355–381. MR1765215

[36] PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and Analysis* (*Varenna*, 1985). *Lecture Notes in Math.* **1206** 167–241. Springer, Berlin. MR0864714

[37] RAUBERTAS, R. F., LEE, C.-I. C. and NORDHEIM, E. V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. Theory Methods* **15** 2809–2833. MR0855765

[38] ROBERTSON, T. (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc.* **73** 197–202. MR0488300

[39] ROBERTSON, T. (1986). On testing symmetry and unimodality. In *Advances in Order Restricted Statistical Inference* (*Iowa City*, *Iowa*, 1985). *Lecture Notes in Statistics* **37** 231–248. Springer, Berlin. MR0875656

[40] ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505. MR0471147

[41] ROBERTSON, T. and WRIGHT, F. T. (1982). On measuring the conformity of a parameter set to a trend, with applications. *Ann. Statist.* **10** 1234–1245. MR0673658

[42] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester. MR0961262

[43] SCHARF, L. L. (1991). *Statistical Signal Processing*: *Detection*, *Estimation and Time Series Analysis*. Addison-Wesley, Reading, MA.

[44] SEN, B. and MEYER, M. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 423–448. MR3611753

[45] SHAPIRO, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Stat. Rev.* **56** 49–62. MR0963140

[46] SLAWSKI, M. and HEIN, M. (2013). Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electron. J. Stat.* **7** 3004–3056. MR3151760

[47] SPOKOINY, V. G. (1998). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* **7** 245–273. MR1651777

[48] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

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

[50] WARRACK, G. and ROBERTSON, T. (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Amer. Statist. Assoc.* **79** 881–886. MR0770282

[51] WEI, Y. and WAINWRIGHT, M. J. (2016). Sharp minimax bounds for testing discrete monotone distributions. In *IEEE International Symposium on Information Theory* (*ISIT*) 2684–2688. IEEE, Los Alamitos, CA.

[52] WEI, Y., WAINWRIGHT, M. J and GUNTUBOYINA, A. (2019). Supplement to "The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii." DOI:10.1214/18-AOS1701SUPP.

[53] ZARANTONELLO, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. In *Contributions to Nonlinear Functional Analysis* (*Proc. Sympos.*, *Math. Res. Center*, *Univ. Wisconsin*, *Madison*, *Wis.*, 1971) 237–424. Academic Press, New York. MR0388177

The Annals of Statistics
2019, Vol. 47, No. 2, 1025–1060
https://doi.org/10.1214/18-AOS1703

NONPARAMETRIC IMPLIED LÉVY DENSITIES[1]

BY LIKUAN QIN AND VIKTOR TODOROV

Northwestern University

This paper develops a nonparametric estimator for the Lévy density of an asset price, following an Itô semimartingale, implied by short-maturity options. The asymptotic setup is one in which the time to maturity of the available options decreases, the mesh of the available strike grid shrinks and the strike range expands. The estimation is based on aggregating the observed option data into nonparametric estimates of the conditional characteristic function of the return distribution, the derivatives of which allow to infer the Fourier transform of a known transform of the Lévy density in a way which is robust to the level of the unknown diffusive volatility of the asset price. The Lévy density estimate is then constructed via Fourier inversion. We derive an asymptotic bound for the integrated squared error of the estimator in the general case as well as its probability limit in the special Lévy case. We further show rate optimality of our Lévy density estimator in a minimax sense. An empirical application to market index options reveals relative stability of the left tail decay during high and low volatility periods.

REFERENCES

[1] ANDERSEN, T., FUSARI, N. and TODOROV, V. (2017). The pricing of short-term market risk: Evidence from weekly options. *J. Finance* **72** 1335–1386.

[2] BELOMESTNY, D. (2011). Spectral estimation of the Lévy density in partially observed affine models. *Stochastic Process. Appl.* **121** 1217–1244. MR2794974

[3] BELOMESTNY, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39** 2205–2242. MR2893866

[4] BELOMESTNY, D. and REISS, M. (2006). Spectral calibration of exponential Lévy models. *Finance Stoch.* **10** 449–474. MR2276314

[5] BELOMESTNY, D. and REISS, M. (2015). *Lévy Matters IV: Estimation for Discretely Observed Lévy Processes. Lecture Notes in Mathematics* **2128**. Springer, Berlin.

[6] BENTATA, A. and CONT, R. (2012). Short-time asymptotics for marginal distributions of semimartingales. Working paper.

[7] CARR, P., GEMAN, H., MADAN, D. and YOR, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus.* **75** 305–332.

[8] CARR, P. and MADAN, D. (2001). Optimal positioning in derivative securities. *Quant. Finance* **1** 19–37. MR1810014

[9] COMTE, F. and GENON-CATALOT, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.* **119** 4088–4123. MR2565560

MSC2010 subject classifications. Primary 62G07, 62M05; secondary 60H10, 60J75.

Key words and phrases. Itô semimartingale, Fourier inversion, Lévy density, nonparametric density estimation, options, stochastic volatility.

[10] COMTE, F. and GENON-CATALOT, V. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.* **39** 803–837. MR2816339

[11] CONT, R. and TANKOV, P. (2004). Nonparametric calibration of jump-diffusion option pricing models. *J. Comput. Finance* **7** 1–49.

[12] DUFFIE, D. (2001). *Dynamic Asset Pricing Theory*, 3rd ed. Princeton Univ. Press, Princeton, NJ.

[13] DUFFIE, D., PAN, J. and SINGLETON, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68** 1343–1376. MR1793362

[14] FIGUEROA-LÓPEZ, J. E. (2009). Nonparametric estimation of time-changed Lévy models under high-frequency data. *Adv. in Appl. Probab.* **41** 1161–1188. MR2663241

[15] FIGUEROA-LÓPEZ, J. E., GONG, R. and HOUDRÉ, C. (2012). Small-time expansions of the distributions, densities, and option prices of stochastic volatility models with Lévy jumps. *Stochastic Process. Appl.* **122** 1808–1839. MR2914773

[16] FIGUEROA-LÓPEZ, J. E. and ÓLAFSSON, S. (2016). Short-term asymptotics for the implied volatility skew under a stochastic volatility model with Lévy jumps. *Finance Stoch.* **20** 973–1020. MR3551858

[17] FIGUEROA-LÓPEZ, J. E. and ÓLAFSSON, S. (2016). Short-time expansions for close-to-the-money options under a Lévy jump model with stochastic volatility. *Finance Stoch.* **20** 219–265. MR3441292

[18] GUGUSHVILI, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21** 321–343. MR2530929

[19] JACOD, J. and PROTTER, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096

[20] JACOD, J. and TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42** 1029–1069. MR3210995

[21] JONGBLOED, G., VAN DER MEULEN, F. H. and VAN DER VAART, A. W. (2005). Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes. *Bernoulli* **11** 759–791. MR2172840

[22] KAPPUS, J. and REISS, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.* **64** 314–328. MR2683463

[23] MERTON, R. (1976). Option pricing when underlying asset returns are discontinuous. *J. Financ. Econ.* **3** 125–144.

[24] MUHLE-KARBE, J. and NUTZ, M. (2011). Small-time asymptotics of option prices and first absolute moments. *J. Appl. Probab.* **48** 1003–1020. MR2896664

[25] NEUMANN, M. H. and REISS, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805

[26] NICKL, R., REISS, M., SÖHL, J. and TRABS, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields* **164** 61–108. MR3449386

[27] QIN, L. and TODOROV, V. (2019). Supplement to "Nonparametric implied Lévy densities." DOI:10.1214/18-AOS1703SUPP.

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

[29] SÖHL, J. (2014). Confidence sets in nonparametric calibration of exponential Lévy models. *Finance Stoch.* **18** 617–649. MR3232018

[30] SÖHL, J. and TRABS, M. (2014). Option calibration of exponential Lévy models: Confidence intervals and empirical results. *J. Comput. Finance* **18** 91–119.

[31] TRABS, M. (2014). Calibration of self-decomposable Lévy models. *Bernoulli* **20** 109–140. MR3160575

[32] TRABS, M. (2015). Quantile estimation for Lévy measures. *Stochastic Process. Appl.* **125** 3484–3521. MR3357617

The Annals of Statistics
2019, Vol. 47, No. 2, 1061–1087
https://doi.org/10.1214/18-AOS1706

ON MODEL SELECTION FROM A FINITE FAMILY OF POSSIBLY MISSPECIFIED TIME SERIES MODELS

BY HSIANG-LING HSU[*,1], CHING-KANG ING[†,2] AND HOWELL TONG[‡,§,3]

National University of Kaohsiung, National Tsing Hua University†, University of Electronic Science & Technology‡ and London School of Economics§*

Consider finite parametric time series models. "I have n observations and k models, which model should I choose on the basis of the data alone" is a frequently asked question in many practical situations. This poses the key problem of selecting a model from a collection of candidate models, none of which is necessarily the true data generating process (DGP). Although existing literature on model selection is vast, there is a serious lacuna in that the above problem does not seem to have received much attention. In fact, existing model selection criteria have avoided addressing the above problem directly, either by assuming that the true DGP is included among the candidate models and aiming at choosing this DGP, or by assuming that the true DGP can be asymptotically approximated by an increasing sequence of candidate models and aiming at choosing the candidate having the best predictive capability in some asymptotic sense. In this article, we propose a misspecification-resistant information criterion (MRIC) to address the key problem directly. We first prove the asymptotic efficiency of MRIC whether the true DGP is among the candidates or not, within the fixed-dimensional framework. We then extend this result to the high-dimensional case in which the number of candidate variables is much larger than the sample size. In particular, we show that MRIC can be used in conjunction with a high-dimensional model selection method to select the (asymptotically) best predictive model across several high-dimensional misspecified time series models.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19** 716–723. MR0423716

AKAIKE, H. (1978). On the likelihood of a time series model. *J. Roy. Statist. Soc. Ser. D* 217–235.

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

BOZDOGAN, H. (2000). Akaike's information criterion and recent developments in information complexity. *J. Math. Psych.* **44** 62–91. MR1770002

BURNHAM, K. P. and ANDERSON, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer, New York. MR1919620

MSC2010 subject classifications. Primary 63M30; secondary 62F07, 62F12.

Key words and phrases. AIC, BIC, misspecification-resistant information criterion, multistep prediction error, high-dimensional misspecified models, orthogonal greedy algorithm.

CHAN, N. H. and ING, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39** 1526–1550. MR2850211

DAVIES, P. L. (2008). Approximating data (with discussion). *J. Korean Statist. Soc.* **37** 191–240.

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

FINDLEY, D. F. (1991). Counterexamples to parsimony and BIC. *Ann. Inst. Statist. Math.* **43** 505–514. MR1143638

FINDLEY, D. F. and WEI, C. Z. (1993). Moment bounds for deriving time series CLTs and model selection procedures. *Statist. Sinica* **3** 453–480. MR1243396

FINDLEY, D. F. and WEI, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *J. Multivariate Anal.* **83** 415–450. MR1945962

GREENAWAY-MCGREVY, R. (2013). Multistep prediction of panel vector autoregressive processes. *Econometric Theory* **29** 699–734. MR3092461

GREENAWAY-MCGREVY, R. (2015). Evaluating panel data forecasts under independent realization. *J. Multivariate Anal.* **136** 108–125. MR3321483

HSU, H.-L., ING, C.-K. and TONG, H. (2019). Supplement to "On model selection from a finite family of possibly misspecified time series models." DOI:10.1214/18-AOS1706SUPP.

ING, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory* **19** 254–279. MR1966030

ING, C.-K. (2007). Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series. *Ann. Statist.* **35** 1238–1277. MR2341705

ING, C.-K. and LAI, T. L. (2011). A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Statist. Sinica* **21** 1473–1513. MR2895106

ING, C.-K. and WEI, C.-Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.* **85** 130–155. MR1978181

ING, C.-K. and WEI, C.-Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474. MR2211091

INOUE, A. and KILIAN, L. (2006). On the selection of forecasting models. *J. Econometrics* **130** 273–306. MR2211796

KONISHI, S. and KITAGAWA, G. (1996). Generalised information criteria in model selection. *Biometrika* **83** 875–890. MR1440051

LI, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15** 958–975. MR0902239

LIU, W. and YANG, Y. (2011). Parametric or nonparametric? A parametricness index for model selection. *Ann. Statist.* **39** 2074–2102. MR2893862

LV, J. and LIU, J. S. (2014). Model selection principles in misspecified models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 141–167. MR3153937

MALLOWS, C. L. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12** 758–765. MR0740928

RAO, C. R. and WU, Y. H. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76** 369–374. MR1016028

SCHORFHEIDE, F. (2005). VAR forecasting under misspecification. *J. Econometrics* **128** 99–136. MR2022928

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SHAO, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. MR1466682

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126. MR0403130

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164. MR0557560

SHIBATA, R. (1981). An optimal selection of regression variables. *Biometrika* **68** 45–54. MR0614940

SIN, C.-Y. and WHITE, H. (1996). Information criteria for selecting possibly misspecified parametric models. *J. Econometrics* **71** 207–225. MR1381082

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. MR0443204

TAKEUCHI, K. (1976). The distribution of information statistic and the criterion of the adequacy of a model. *Suri-Kagaku* (*Mathematical Sciences*) **3** 12–18.

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

VAN ERVEN, T., GRÜNWALD, P. and DE ROOIJ, S. (2012). Catching up faster by switching sooner: A predictive approach to adaptive estimation with an application to the AIC–BIC dilemma. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 361–417. MR2925369

WANG, H. (2009). Forward regression for ultra-high dimensional variable screening. *J. Amer. Statist. Assoc.* **104** 1512–1524. MR2750576

WEI, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20** 1–42. MR1150333

WU, W.-B. and WU, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186

XIA, Y. and TONG, H. (2011). Feature matching (with discussion). *Statist. Sci.* **26** 21–46.

YANG, Y. (2007). Prediction/estimation with simple linear models: Is it really that simple? *Econometric Theory* **23** 1–36. MR2338950

ZHANG, Y. and YANG, Y. (2015). Cross-validation for selecting a model selection procedure. *J. Econometrics* **187** 95–112. MR3347297

The Annals of Statistics
2019, Vol. 47, No. 2, 1088–1112
https://doi.org/10.1214/18-AOS1707

ESTIMATING THE ALGORITHMIC VARIANCE OF RANDOMIZED ENSEMBLES VIA THE BOOTSTRAP[1]

BY MILES E. LOPES

University of California, Davis

Although the methods of bagging and random forests are some of the most widely used prediction methods, relatively little is known about their algorithmic convergence. In particular, there are not many theoretical guarantees for deciding when an ensemble is "large enough"—so that its accuracy is close to that of an ideal infinite ensemble. Due to the fact that bagging and random forests are randomized algorithms, the choice of ensemble size is closely related to the notion of "algorithmic variance" (i.e., the variance of prediction error due only to the training algorithm). In the present work, we propose a bootstrap method to estimate this variance for bagging, random forests and related methods in the context of classification. To be specific, suppose the training dataset is fixed, and let the random variable ERR_t denote the prediction error of a randomized ensemble of size t. Working under a "first-order model" for randomized ensembles, we prove that the centered law of ERR_t can be consistently approximated via the proposed method as $t \to \infty$. Meanwhile, the computational cost of the method is quite modest, by virtue of an extrapolation technique. As a consequence, the method offers a practical guideline for deciding when the algorithmic fluctuations of ERR_t are negligible.

REFERENCES

ARLOT, S. and GENUER, R. (2014). Analysis of purely random forests bias. arXiv:1407.3939.

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

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

BICKEL, P. J. and YAHAV, J. A. (1988). Richardson extrapolation and the bootstrap. *J. Amer. Statist. Assoc.* **83** 387–393. MR0971364

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

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

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

BREZINSKI, C. and ZAGLIA, M. R. (2013). *Extrapolation Methods*: *Theory and Practice*. North-Holland, Amsterdam.

BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. MR1926165

BUJA, A. and STUETZLE, W. (2000). Smoothing effects of bagging. Preprint, AT&T Labs-Research, Florham Park, NJ.

MSC2010 subject classifications. Primary 62F40; secondary 65B05, 68W20, 60G25.
Key words and phrases. Bootstrap, random forests, bagging, randomized algorithms.

BUJA, A. and STUETZLE, W. (2006). Observations on bagging. *Statist. Sinica* **16** 323–351. MR2267238

BÜRGISSER, P. and CUCKER, F. (2013). *Condition: The Geometry of Numerical Algorithms. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **349**. Springer, Heidelberg. MR3098452

BYRD, R. H., CHIN, G. M., NOCEDAL, J. and WU, Y. (2012). Sample size selection in optimization methods for machine learning. *Math. Program.* **134** 127–155. MR2947555

CANNINGS, T. I. and SAMWORTH, R. J. (2017). Random-projection ensemble classification. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 959–1035. With discussions and a reply by the authors. MR3689307

DIETTERICH, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **40** 139–157.

EFRON, B. (2014). Estimation and accuracy after model selection. *J. Amer. Statist. Assoc.* **109** 991–1007. MR3265671

GENUER, R. (2012). Variance reduction in purely random forests. *J. Nonparametr. Stat.* **24** 543–562. MR2968888

HALL, P. and SAMWORTH, R. J. (2005). Properties of bagged nearest neighbour classifiers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 363–379. MR2155343

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York. MR1851606

HERNÁNDEZ-LOBATO, D., MARTÍNEZ-MUÑOZ, G. and SUÁREZ, A. (2013). How large should ensembles of classifiers be? *Pattern Recognit.* **46** 1323–1336.

HO, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 832–844.

KALLENBERG, O. (2006). *Foundations of Modern Probability*. Springer, Berlin.

LAM, L. and SUEN, C. Y. (1997). Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum.* **27** 553–568.

LATINNE, P., DEBEIR, O. and DECAESTECKER, C. (2001). Limiting the number of trees in random forests. In *Multiple Classifier Systems* (*Cambridge*, 2001). *Lecture Notes in Computer Science* **2096** 178–187. Springer, Berlin. MR2043269

LIAW, A. and WIENER, M. (2002). Classification and regression by randomForest. *R News* **2** 18–22.

LICHMAN, M. (2013). UCI machine learning repository.

LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176

LOPES, M. E. (2016). A sharp bound on the computation-accuracy tradeoff for majority voting ensembles. Preprint, arXiv:1303.0727.

LOPES, M. E. (2019). Supplement to "Estimating the algorithmic variance of randomized ensembles via the bootstrap." DOI:10.1214/18-AOS1707SUPP.

LOPES, M. E., WANG, S. and MAHONEY, M. W. (2017). A bootstrap method for error estimation in randomized matrix multiplication. Preprint, arXiv:1708.01945.

LOPES, M. E., WANG, S. and MAHONEY, M. W. (2018). Error estimation for randomized least-squares algorithms via the bootstrap. In *Proceedings of the* 35*th International Conference on Machine Learning* **80** 3217–3226.

MENTCH, L. and HOOKER, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.* **17** Paper No. 26, 41. MR3491120

NG, A. Y. and JORDAN, M. I. (2001). Convergence rates of the Voting Gibbs classifier, with application to Bayesian feature selection. In *Proceedings of the* 18*th International Conference on Machine Learning* 377–384.

OSHIRO, T. M., PEREZ, P. S. and BARANAUSKAS, J. A. (2012). How many trees in a random forest? In *Machine Learning and Data Mining in Pattern Recognition* 154–168. Springer, Berlin.

SCHAPIRE, R. E. and FREUND, Y. (2012). *Boosting*: *Foundations and Algorithms*. MIT Press, Cambridge, MA. MR2920188

SCORNET, E. (2016a). On the asymptotics of random forests. *J. Multivariate Anal.* **146** 72–83. MR3477650

SCORNET, E. (2016b). Random forests and kernel methods. *IEEE Trans. Inform. Theory* **62** 1485–1500. MR3472261

SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876

SEXTON, J. and LAAKE, P. (2009). Standard errors for bagged and random forest estimators. *Comput. Statist. Data Anal.* **53** 801–811. MR2654590

SIDI, A. (2003). *Practical Extrapolation Methods*: *Theory and Applications*. *Cambridge Monographs on Applied and Computational Mathematics* **10**. Cambridge Univ. Press, Cambridge. MR1994507

SIMON, L. (1983). *Lectures on Geometric Measure Theory*. *Proceedings of the Centre for Mathematical Analysis*, *Australian National University* **3**. Australian National Univ., Centre for Mathematical Analysis, Canberra. MR0756417

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

WAGER, S., HASTIE, T. and EFRON, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15** 1625–1651. MR3225243

WHITE, B. (2016). Introduction to minimal surface theory. In *Geometric Analysis*. *IAS/Park City Math. Ser.* **22** 387–438. Amer. Math. Soc., Providence, RI. MR3524221

The Annals of Statistics
2019, Vol. 47, No. 2, 1113–1147
https://doi.org/10.1214/18-AOS1708

EFFICIENT NONPARAMETRIC BAYESIAN INFERENCE FOR *X*-RAY TRANSFORMS

BY FRANÇOIS MONARD[*,1], RICHARD NICKL[†,2] AND GABRIEL P. PATERNAIN[†,3]

University of California, Santa Cruz and University of Cambridge†*

We consider the statistical inverse problem of recovering a function $f : M \to \mathbb{R}$, where M is a smooth compact Riemannian manifold with boundary, from measurements of general X-ray transforms $I_a(f)$ of f, corrupted by additive Gaussian noise. For M equal to the unit disk with "flat" geometry and $a = 0$ this reduces to the standard Radon transform, but our general setting allows for anisotropic media M and can further model local "attenuation" effects—both highly relevant in practical imaging problems such as SPECT tomography. We study a nonparametric Bayesian inference method based on standard Gaussian process priors for f. The posterior reconstruction of f corresponds to a Tikhonov regulariser with a reproducing kernel Hilbert space norm penalty that does not require the calculation of the singular value decomposition of the forward operator I_a. We prove Bernstein–von Mises theorems for a large family of one-dimensional linear functionals of f, and they entail that posterior-based inferences such as credible sets are valid and optimal from a frequentist point of view. In particular we derive the asymptotic distribution of smooth linear functionals of the Tikhonov regulariser, which attains the semiparametric information lower bound. The proofs rely on an invertibility result for the "Fisher information" operator $I_a^* I_a$ between suitable function spaces, a result of independent interest that relies on techniques from microlocal analysis. We illustrate the performance of the proposed method via simulations in various settings.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A., eds. (1966). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York. MR0208797

[2] BOUTET DE MONVEL, L. (1966). Comportement d'un opérateur pseudo-différentiel sur une variété à bord. II. Pseudo-noyaux de Poisson. *J. Anal. Math.* **17** 255–304. MR0239255

[3] BOUTET DE MONVEL, L. (1971). Boundary problems for pseudo-differential operators. *Acta Math.* **126** 11–51. MR0407904

[4] BROWN, D. A., SAIBABA, A. and VALLÉLIAN, S. (2017). Low rank independence samplers in bayesian inverse problems. ArXiv preprint. Available at arXiv:1609.07180v2.

[5] BUDINGER, T. F., GULLBERG, G. T. and HUESMAN, R. H. (1979). Emission computed tomography. In *Image Reconstruction from Projections* (G. T. Herman, ed.). *Topics in Applied Physics* **32** 147–246. Springer, Berlin.

MSC2010 subject classifications. Primary 62G20; secondary 58J40, 65R10, 62F15.

Key words and phrases. Inverse problem, Bernstein–von Mises theorem, MAP estimate, Tikhonov regulariser, Gaussian prior, Radon transform, semiparametric efficiency.

[6] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477

[7] CASTILLO, I. (2017). Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2074–2102. MR3729648

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

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

[10] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597

[11] CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Probl.* **24** Article ID 034004. MR2421941

[12] CAVALIER, L., GOLUBEV, G. K., PICARD, D. and TSYBAKOV, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.* **30** 843–874. MR1922543

[13] CHAZARAIN, J. and PIRIOU, A. (1982). *Introduction to the Theory of Linear Partial Differential Equations. Studies in Mathematics and Its Applications* **14**. North-Holland, Amsterdam. Translated from the French. MR0678605

[14] CREAGER, K. C. (1992). Anisotropy of the inner core from differential travel times of the phases PKP and PKIPK. *Nature* **356** 309–314.

[15] DAIRBEKOV, N. S., PATERNAIN, G. P., STEFANOV, P. and UHLMANN, G. (2007). The boundary rigidity problem in the presence of a magnetic field. *Adv. Math.* **216** 535–609. MR2351370

[16] DASHTI, M., LAW, K. J. H., STUART, A. M. and VOSS, J. (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.* **29** Article ID 095017. MR3104933

[17] DASHTI, M. and STUART, A. (2016). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification* (R. Ghanem, D. Higdon and H. Owhadi, eds.) Springer, Berlin.

[18] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. Revised reprint of the 1989 original. MR1932358

[19] DUVALL, T. L. JR, JEFFFERIES, S. M., HARVEY, J. W. and POMERANTZ, M. A. (1993). Time–distance helioseismology. *Nature* **362** 430–432.

[20] ESKIN, G. I. (1981). *Boundary Value Problems for Elliptic Pseudodifferential Equations. Translations of Mathematical Monographs* **52**. Amer. Math. Soc., Providence, RI. Translated from the Russian by S. Smith. MR0623608

[21] FRIGYIK, B., STEFANOV, P. and UHLMANN, G. (2008). The X-ray transform for a generic family of curves and weights. *J. Geom. Anal.* **18** 89–108. MR2365669

[22] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782

[23] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Univ. Press, New York. MR3588285

[24] GOLDENSHLUGER, A. and PEREVERZEV, S. V. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields* **118** 169–186. MR1790080

[25] GRUBB, G. (2014). Local and nonlocal boundary conditions for μ-transmission and fractional elliptic pseudodifferential operators. *Anal. PDE* **7** 1649–1682. MR3293447

[26] GRUBB, G. (2015). Fractional Laplacians on domains, a development of Hörmander's theory of μ-transmission pseudodifferential operators. *Adv. Math.* **268** 478–528. MR3276603

[27] GUILLEMIN, V. and STERNBERG, S. (1977). *Geometric Asymptotics*. *Mathematical Surveys* **14**. Amer. Math. Soc., Providence, RI. MR0516965

[28] HELGASON, S. (1999). *The Radon Transform*, 2nd ed. *Progress in Mathematics* **5**. Birkhäuser, Boston, MA. MR1723736

[29] HÖRMANDER, L. (1965). *Ch II*, *Boundary Problems for "Classical" Pseudo-Differential Operators*. *Lectures Notes at IAS*, Princeton. Available at http://www.math.ku.dk/~grubb/LH65.pdf.

[30] HÖRMANDER, L. (1985). *The Analysis of Linear Partial Differential Operators*. *III*. *Pseudodifferential operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **274**. Springer, Berlin. MR0781536

[31] JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18** 251–280. MR1041393

[32] KERKYACHARIAN, G., KYRIAZIS, G., LE PENNEC, E., PETRUSHEV, P. and PICARD, D. (2010). Inversion of noisy Radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.* **28** 24–45. MR2563258

[33] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881

[34] KUCHMENT, P. (2014). *The Radon Transform and Medical Imaging*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **85**. SIAM, Philadelphia, PA. MR3137435

[35] LOUIS, A. K. (1984). Orthogonal function series expansions and the null space of the Radon transform. *SIAM J. Math. Anal.* **15** 621–633. MR0740700

[36] MCLEAN, W. (2000). *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge Univ. Press, Cambridge. MR1742312

[37] MONARD, F. (2014). Numerical implementation of geodesic X-ray transforms and their inversion. *SIAM J. Imaging Sci.* **7** 1335–1357. MR3217215

[38] MUHOMETOV, R. G. (1977). The reconstruction problem of a two-dimensional Riemannian metric, and integral geometry. *Dokl. Akad. Nauk SSSR* **232** 32–35. MR0431074

[39] MUNK, W. and WUNSCH, C. (1979). Ocean acoustic tomography: A scheme for large scale monitoring. *Deep-Sea Res.* **26A** 123–161.

[40] NATTERER, F. (1986). *The Mathematics of Computerized Tomography*. Wiley, Chichester. MR0856916

[41] NICKL, R. (2017). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. ArXiv preprint. Available at arxiv:1707.01764.

[42] NICKL, R. and SÖHL, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693. MR3670192

[43] NOVIKOV, R. G. (2002). An inversion formula for the attenuated X-ray transformation. *Ark. Mat.* **40** 145–167. MR1948891

[44] PATERNAIN, G. P., SALO, M., UHLMANN, G. and ZHOU, H. (2018). The geodesic X-ray transform with matrix weights. *Am. J. Math.* To appear. Available at arXiv:1605.07894.

[45] PESTOV, L. and UHLMANN, G. (2004). On characterization of the range and inversion formulas for the geodesic X-ray transform. *Int. Math. Res. Not.* **80** 4331–4347. MR2126628

[46] PESTOV, L. and UHLMANN, G. (2005). Two dimensional compact simple Riemannian manifolds are boundary distance rigid. *Ann. of Math.* (2) **161** 1093–1110. MR2153407

[47] RADON, J. Über die Bestimmung Von Funktionen Durch Ihre Integralwerte Längs Gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse* **69** 262–277.

[48] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. MR3117105

[49] SALO, M. and UHLMANN, G. (2011). The attenuated ray transform on simple surfaces. *J. Differential Geom.* **88** 161–187. MR2819758

[50] SHARAFUTDINOV, V., SKOKAN, M. and UHLMANN, G. (2005). Regularity of ghosts in tensor tomography. *J. Geom. Anal.* **15** 499–542. MR2190243

[51] SHARAFUTDINOV, V. A. (1994). *Integral Geometry of Tensor Fields*. VSP, Utrecht. MR1374572

[52] SHARAFUTDINOV, V. A. (2005). Geometric symbol calculus for pseudodifferential operators. I. [Translation of *Mat. Tr.* **7**(2) (2004) 159–206; MR2124544]. *Siberian Adv. Math.* **15** 81–125. MR2191866

[53] STEFANOV, P. and UHLMANN, G. (2004). Stability estimates for the X-ray transform of tensor fields and boundary rigidity. *Duke Math. J.* **123** 445–467. MR2068966

[54] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785

[55] TAYLOR, M. E. (1996). *Partial Differential Equations. I. Basic Theory. Applied Mathematical Sciences* **115**. Springer, New York. MR1395148

[56] THORBERGSSON, G. (1978). Closed geodesics on non-compact Riemannian manifolds. *Math. Z.* **159** 249–258. MR0493872

[57] TRÈVES, F. (1980). *Introduction to Pseudodifferential and Fourier Integral Operators. Vol.* 1. *Pseudodifferential Operators*. Plenum Press, New York. MR0597144

[58] UHLMANN, G. and VASY, A. (2016). The inverse problem for the local geodesic ray transform. *Invent. Math.* **205** 83–120. MR3514959

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

[60] 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

[61] VIŠIK, M. I. and ÈSKIN, G. I. (1965). Convolution equations in a bounded region. *Uspekhi Mat. Nauk* **20** 89–152. MR0185273

[62] Seppecher's FEM_beta package. Available at http://math.univ-lyon1.fr/homes-www/seppecher/codes.php.

The Annals of Statistics
2019, Vol. 47, No. 2, 1148–1178
https://doi.org/10.1214/18-AOS1709

GENERALIZED RANDOM FORESTS

BY SUSAN ATHEY*, JULIE TIBSHIRANI† AND STEFAN WAGER*

Stanford University and Elasticsearch BV†*

We propose generalized random forests, a method for nonparametric statistical estimation based on random forests (Breiman [*Mach. Learn.* **45** (2001) 5–32]) that can be used to fit any quantity of interest identified as the solution to a set of local moment equations. Following the literature on local maximum likelihood estimation, our method considers a weighted set of nearby training examples; however, instead of using classical kernel weighting functions that are prone to a strong curse of dimensionality, we use an adaptive weighting function derived from a forest designed to express heterogeneity in the specified quantity of interest. We propose a flexible, computationally efficient algorithm for growing generalized random forests, develop a large sample theory for our method showing that our estimates are consistent and asymptotically Gaussian and provide an estimator for their asymptotic variance that enables valid confidence intervals. We use our approach to develop new methods for three statistical tasks: nonparametric quantile regression, conditional average partial effect estimation and heterogeneous treatment effect estimation via instrumental variables. A software implementation, `grf` for `R` and `C++`, is available from `CRAN`.

REFERENCES

ABADIE, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *J. Econometrics* **113** 231–263. MR1960380

AMIT, Y. and GEMAN, D. (1997). Shape quantization and recognition with randomized trees. *Neural Comput*. **9** 1545–1588.

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

ANGRIST, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *AER* 313–336.

ANGRIST, J. D. and EVANS, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *AER* 450–477.

ARLOT, S. and GENUER, R. (2014). Analysis of purely random forests bias. ArXiv preprint. Available at arXiv:1407.3939.

ATHEY, S. and IMBENS, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113** 7353–7360. MR3531135

ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2018). Supplement to "Generalized random forests." DOI:10.1214/18-AOS1709SUPP.

BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429. MR3001131

MSC2010 subject classifications. 62G05.
Key words and phrases. Asymptotic theory, causal inference, instrumental variable.

BEYGELZIMER, A. and LANGFORD, J. (2009). The offset tree for learning with partial labels. In *Proceedings of KDD* 129–138. ACM.

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

BIAU, G. and DEVROYE, L. (2010). On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. *J. Multivariate Anal.* **101** 2499–2518. MR2719877

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

BIAU, G. and SCORNET, E. (2016). A random forest guided tour. *TEST* **25** 197–227. MR3493512

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

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

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. MR1926165

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68.

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172

DAROLLES, S., FAN, Y., FLORENS, J. P. and RENAULT, E. (2011). Nonparametric instrumental regression. *Econometrica* **79** 1541–1565. MR2883763

DENIL, M., MATHESON, D. and DE FREITAS, N. (2014). Narrowing the Gap: Random forests in theory and in practice. In *Proceedings of ICML* 665–673.

DIETTERICH, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **40** 139–157.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR0659849

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. MR0615434

FAN, J., FARMEN, M. and GIJBELS, I. (1998). Local maximum likelihood estimation and inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 591–608. MR1626013

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

GEURTS, P., ERNST, D. and WEHENKEL, L. (2006). Extremely randomized trees. *Mach. Learn.* **63** 3–42.

GORDON, L. and OLSHEN, R. A. (1985). Tree-structured survival analysis. *Cancer Treat. Rep.* **69** 1065–1069.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

HANSEN, B. E. (1992). Testing for parameter instability in linear models. *J. Policy Model.* **14** 517–533.

HARTFORD, J., LEWIS, G., LEYTON-BROWN, K. and TADDY, M. (2017). Deep IV: A flexible approach for counterfactual prediction. In *Proceedings of ICML* 1414–1423.

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

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546

HJORT, N. L. and KONING, A. (2002). Tests for constancy of model parameters over time. *J. Nonparametr. Stat.* **14** 113–132. MR1905588

HO, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 832–844.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294

HONORÉ, B. E. and KYRIAZIDOU, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* **68** 839–874. MR1771585

HOTHORN, T., LAUSEN, B., BENNER, A. and RADESPIEL-TRÖGER, M. (2004). Bagging survival trees. *Stat. Med.* **23** 77–91.

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–475.

ISHWARAN, H. and KOGALUR, U. B. (2010). Consistency of random survival forests. *Statist. Probab. Lett.* **80** 1056–1064. MR2651045

KALLUS, N. (2017). Recursive Partitioning for Personalization using Observational Data. In *Proceedings of ICML*. 1789–1798.

KLEIBER, C. and ZEILEIS, A. (2008). *Applied Econometrics with R*. Springer Science & Business Media.

LEBLANC, M. and CROWLEY, J. (1992). Relative risk trees for censored survival data. *Biometrics* 411–425.

LEWBEL, A. (2007). A local generalized method of moments estimator. *Econom. Lett.* **94** 124–128. MR2287404

LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176

LOADER, C. (1999). *Local Regression and Likelihood*. Springer, New York. MR1704236

MALLOWS, C. L. (1973). Some comments on Cp. *Technometrics* **15** 661–675.

MEINSHAUSEN, N. (2006). Quantile regression forests. *J. Mach. Learn. Res.* **7** 983–999. MR2274394

MENTCH, L. and HOOKER, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.* **17** 26. MR3491120

MOLINARO, A. M., DUDOIT, S. and VAN DER LAAN, M. J. (2004). Tree-based multivariate regression and density estimation with right-censored data. *J. Multivariate Anal.* **90** 154–177. MR2064940

NEWEY, W. K. (1994a). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10** 233–253. MR1293201

NEWEY, W. K. (1994b). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237

NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578. MR2000257

NEYMAN, J. (1979). $C(\alpha)$ tests and their use. *Sankhya*, *Ser. A* **41** 1–21. MR0615037

NYBLOM, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR0999682

PLOBERGER, W. and KRÄMER, W. (1992). The CUSUM test with OLS residuals. *Econometrica* **60** 271–285. MR1162619

POTERBA, J. M., VENTI, S. F. and WISE, D. A. (1996). How retirement saving programs increase saving. *J. Electron. Publ.* **10** 91–112.

ROBINS, J. M. and RITOV, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Stat. Med.* **16**.

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151. MR0856811

SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876

SEXTON, J. and LAAKE, P. (2009). Standard errors for bagged and random forest estimators. *Comput. Statist. Data Anal.* **53** 801–811. MR2654590

STANISWALIS, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84** 276–283. MR0999689

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. MR0443204

SU, L., MURTAZASHVILI, I. and ULLAH, A. (2013). Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling. *J. Bus. Econom. Statist.* **31** 184–207. MR3055331

SU, X., TSAI, C.-L., WANG, H., NICKERSON, D. M. and LI, B. (2009). Subgroup analysis via recursive partitioning. *J. Mach. Learn. Res.* **10** 141–158.

TIBSHIRANI, R. and HASTIE, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.* **82** 559–567. MR0898359

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

VARIAN, H. R. (2014). Big data: New tricks for econometrics. *J. Electron. Publ.* **28** 3–27.

WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242.

WAGER, S., HASTIE, T. and EFRON, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15** 1625–1651. MR3225243

WAGER, S. and WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. ArXiv preprint. Available at arXiv:1503.06388.

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

WRIGHT, M. N. and ZIEGLER, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17.

ZEILEIS, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Rev.* **24** 445–466. MR2208835

ZEILEIS, A. and HORNIK, K. (2007). Generalized M-fluctuation tests for parameter instability. *Stat. Neerl.* **61** 488–508. MR2351461

ZEILEIS, A., HOTHORN, T. and HORNIK, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist.* **17** 492–514. MR2439970

ZHU, R., ZENG, D. and KOSOROK, M. R. (2015). Reinforcement learning trees. *J. Amer. Statist. Assoc.* **110** 1770–1784. MR3449072

The Annals of Statistics
2019, Vol. 47, No. 2, 1179–1202
https://doi.org/10.1214/18-AOS1723

A CLASSIFICATION CRITERION FOR DEFINITIVE SCREENING DESIGNS

BY ERIC D. SCHOEN[*,‡,1], PIETER T. EENDEBAK[†] AND PETER GOOS[*,†]

KU Leuven, University of Antwerp*[†] *and TNO*[‡]

A conference design is a rectangular matrix with orthogonal columns, one zero in each column, at most one zero in each row and -1's and $+1$'s elsewhere. A definitive screening design can be constructed by folding over a conference design and adding a row vector of zeroes. We prove that, for a given even number of rows, there is just one isomorphism class for conference designs with two or three columns. Next, we derive all isomorphism classes for conference designs with four columns. Based on our results, we propose a classification criterion for definitive screening designs founded on projections into four factors. We illustrate the potential of the criterion by studying designs with 24 and 82 factors.

REFERENCES

DELSARTE, P., GOETHALS, J. M. and SEIDEL, J. J. (1971). Orthogonal matrices with zero diagonal. II. *Canad. J. Math.* **23** 816–832. MR0325418

DENG, L.-Y. and TANG, B. (1999). Generalized resolution and minimum aberration criteria for Plackett–Burman and other nonregular factorial designs. *Statist. Sinica* **9** 1071–1082. MR1744824

DENG, L.-Y. and TANG, B. (2002). Design selection and classification for Hadamard matrices using generalized minimum aberration criteria. *Technometrics* **44** 173–184. MR1938726

DOUGHERTY, S., SIMPSON, J. R., HILL, R. R., PIGNATIELLO, J. J. and WHITE, E. D. (2015). Effect of heredity and sparsity on second-order screening design performance. *Qual. Reliab. Eng. Int.* **31** 355–368.

FIDALEO, M., LAVECCHIA, R., PETRUCCI, E. and ZUORRO, A. (2016). Application of a novel definitive screening design to decolorization of an azo dye on boron-doped diamond electrodes. *Int. J. Environ. Sci. Technol.* **13** 835–842.

GEORGIOU, S. D., STYLIANOU, S. and AGGARWAL, M. (2014). Efficient three-level screening designs using weighing matrices. *Statistics* **48** 815–833. MR3234064

GREIG, M., HAANPÄÄ, H. and KASKI, P. (2006). On the coexistence of conference matrices and near resolvable 2-$(2k+1, k, k-1)$ designs. *J. Combin. Theory Ser. A* **113** 703–711. MR2216463

HURKENS, C. A. J. and SEIDEL, J. J. (1985). Conference matrices from projective planes of order 9. *European J. Combin.* **6** 49–57. MR0793487

JONES, B. and NACHTSHEIM, C. J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. *J. Qual. Technol.* **43** 1–15.

JONES, B. and NACHTSHEIM, C. J. (2013). Definitive screening designs with added two-level categorical factors. *J. Qual. Technol.* **45** 121–129.

JONES, B. and NACHTSHEIM, C. J. (2017). Effective design-based model selection for definitive screening designs. *Technometrics* **59** 319–329. MR3677963

MSC2010 subject classifications. Primary 62K20; secondary 05B20, 94C30.
Key words and phrases. Conference matrix, isomorphism class, J_4-characteristic, projection.

MEE, R. W., SCHOEN, E. D. and EDWARDS, D. J. (2017). Selecting an orthogonal or nonorthogonal two-level design for screening. *Technometrics* **59** 305–318. MR3677962

NACHTSHEIM, A. C., SHEN, W. and LIN, D. K. J. (2017). Two-level augmented definitve screening designs. *J. Qual. Technol.* **49** 93–107.

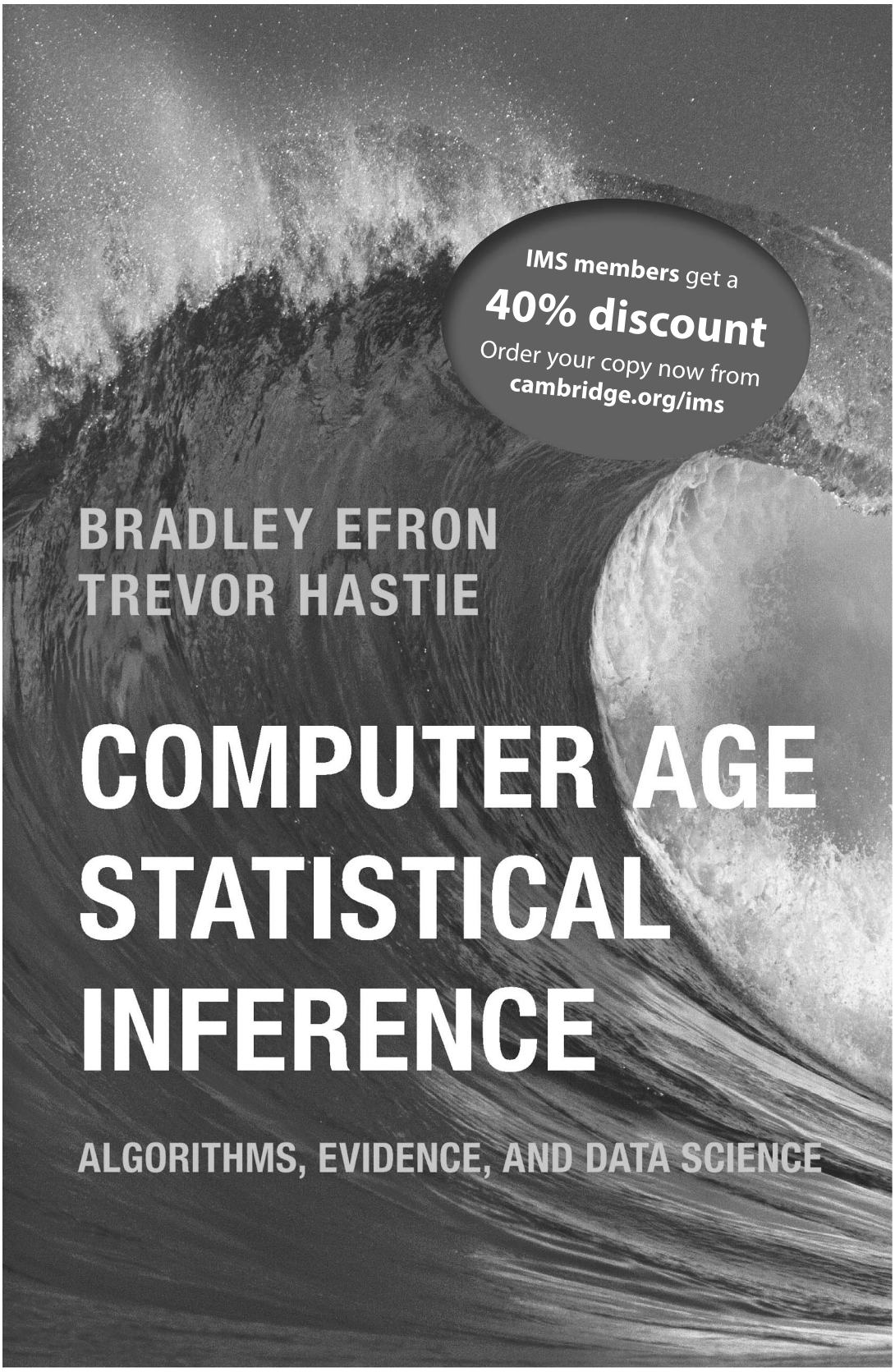
NGUYEN, N. K. and PHAM, T. D. (2016). Small mixed-level screening designs with orthogonal quadratic effects. *J. Qual. Technol.* **48** 405–414.

PATIL, M. V. (2017). Multi response simulation and optimization of gas tungsten arc welding. *Appl. Math. Model.* **42** 540–553.

SCHOEN, E. D., EENDEBAK, P. T. and GOOS, P. (2019). Supplement to "A classification criterion for definitive screening designs". DOI:10.1214/18-AOS1723SUPP.

SCHOEN, E. D., VO-THANH, N. and GOOS, P. (2017). Two-level orthogonal screening designs with 24, 28, 32, and 36 runs. *J. Amer. Statist. Assoc.* **112** 1354–1369. MR3735383

XIAO, L., LIN, D. K. and BAI, F. (2012). Constructing definitive screening designs using conference matrices. *J. Qual. Technol.* **44** 1–7.


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

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS