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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Exact post-selection inference, with application to the lasso
JASON D. LEE, DENNIS L. SUN, YUEKAI SUN AND JONATHAN E. TAYLOR 907
Nonparametric eigenvalue-regularized precision or covariance matrix estimator
CLIFFORD LAM 928
Global rates of convergence of the MLEs of log-concave and s-concave densities
CHARLES R. DOSS AND JON A. WELLNER 954
Classification in general finite dimensional spaces with the k-nearest neighbor rule
SÉBASTIEN GADAT, THIERRY KLEIN AND CLÉMENT MARTEAU 982
A new prior for discrete DAG models with a restricted set of directions
HÉLÈNE MASSAM AND JACEK WESOŁOWSKI 1010
SLOPE is adaptive to unknown sparsity and asymptotically minimax
WEIJIE SU AND EMMANUEL CANDÈS 1038
Supremum norm posterior contraction and credible sets for nonparametric multivariate regression.................... WILLIAM WEIMIN YOO AND SUBHASHIS GHOSAL 1069
Optimal designs for comparing curves .... HOLGER DETTE AND KIRSTEN SCHORNING 1103
Randomization-based models for multitiered experiments: I. A chain of randomizations
R. A. BAILEY AND C. J. BRIEN 1131
Vector quantile regression: An optimal transport approach
GUILLAUME CARLIER, VICTOR CHERNOZHUKOV AND ALFRED GALICHON 1165
Structure identification in panel data analysis
YUAN KE, JIALIANG LI AND WENYANG ZHANG 1193
Inference for single-index quantile regression models with profile optimization
SHUJIE MA AND XUMING HE 1234
Theoretical analysis of nonparametric filament estimation
WANLI QIAO AND WOLFGANG POLONIK 1269
Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data................................ YU-RU SU AND JANE-LING WANG 1298
Approximation and estimation of s-concave densities via Rényi divergences
QIYANG HAN AND JON A. WELLNER 1332

Correction

Limit theorems for empirical processes of cluster functionals
HOLGER DREES AND HOLGER ROOTZÉN 1360

Vol. 44, No. 3—June 2016

The Annals of Statistics
2016, Vol. 44, No. 3, 907–927
DOI: 10.1214/15-AOS1371

EXACT POST-SELECTION INFERENCE, WITH APPLICATION TO THE LASSO

BY JASON D. LEE[*,1], DENNIS L. SUN[+,2], YUEKAI SUN[*,3] AND JONATHAN E. TAYLOR[‡,4]

University of California, Berkeley, California Polytechnic State University[+] and Stanford University[‡]*

We develop a general approach to valid inference after model selection. At the core of our framework is a result that characterizes the distribution of a post-selection estimator conditioned on the *selection event*. We specialize the approach to model selection by the lasso to form valid confidence intervals for the selected coefficients and test whether all relevant variables have been included in the model.

REFERENCES

BENJAMINI, Y., HELLER, R. and YEKUTIELI, D. (2009). Selective inference in complex research. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4255–4271. MR2546387

BENJAMINI, Y. and YEKUTIELI, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.* **100** 71–93. MR2156820

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

COX, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika* **62** 441–444. MR0378189

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

FISHER, R. (1956). On a test of significance in Pearson's Biometrika Tables (No. 11). *J. Roy. Statist. Soc. Ser. B.* **18** 56–60. MR0082773

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

JAVANMARD, A. and MONTANARI, A. (2013). Confidence intervals and hypothesis testing for high-dimensional regression. Preprint. Available at arXiv:1306.3171.

KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787

LEEB, H. and PÖTSCHER, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21** 21–59. MR2153856

LEEB, H. and PÖTSCHER, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Statist.* **34** 2554–2591. MR2291510

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

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. and TIBSHIRANI, R. (2014). A significance test for the lasso (with discussion). *Ann. Statist.* **42** 413–468. MR3210970

MSC2010 subject classifications. Primary 62F03, 62J07; secondary 62E15.
Key words and phrases. Lasso, confidence interval, hypothesis test, model selection.

MILLER, A. (2002). *Subset Selection in Regression*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2001193

NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133

PÖTSCHER, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7** 163–185. MR1128410

PÖTSCHER, B. M. and SCHNEIDER, U. (2010). Confidence sets based on penalized maximum likelihood estimators in Gaussian regression. *Electron. J. Stat.* **4** 334–360. MR2645488

ROBINSON, G. K. (1979). Conditional properties of statistical procedures. *Ann. Statist.* **7** 742–755. MR0532239

SAMPSON, A. R. and SILL, M. W. (2005). Drop-the-losers design: Normal case. *Biom. J.* **47** 257–268. MR2145117

SILL, M. W. and SAMPSON, A. R. (2009). Drop-the-losers design: Binomial case. *Comput. Statist. Data Anal.* **53** 586–595. MR2654594

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

TAYLOR, J., LOCKHART, R., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). Post-selection adaptive inference for least angle regression and the lasso. Preprint. Available at arXiv:1401.3889.

TIAN, X. and TAYLOR, J. (2015). Asymptotics of selective inference. Preprint. Available at arXiv:1501.03588.

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

TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. MR3066375

TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2015). Uniform asymptotic inference and the bootstrap after model selection. Preprint. Available at arXiv:1506.06266.

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2013). On asymptotically optimal confidence regions and tests for high-dimensional models. Preprint. Available at arXiv:1303.0518.

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

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

ZHONG, H. and PRENTICE, R. L. (2008). Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9** 621–634.

ZOLLNER, S. and PRITCHARD, J. K. (2007). Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **80** 605–615.

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

The Annals of Statistics
2016, Vol. 44, No. 3, 928–953
DOI: 10.1214/15-AOS1393

NONPARAMETRIC EIGENVALUE-REGULARIZED PRECISION OR COVARIANCE MATRIX ESTIMATOR

BY CLIFFORD LAM

London School of Economics and Political Science

We introduce nonparametric regularization of the eigenvalues of a sample covariance matrix through splitting of the data (NERCOME), and prove that NERCOME enjoys asymptotic optimal nonlinear shrinkage of eigenvalues with respect to the Frobenius norm. One advantage of NERCOME is its computational speed when the dimension is not too large. We prove that NERCOME is positive definite almost surely, as long as the true covariance matrix is so, even when the dimension is larger than the sample size. With respect to the Stein's loss function, the inverse of our estimator is asymptotically the optimal precision matrix estimator. Asymptotic efficiency loss is defined through comparison with an ideal estimator, which assumed the knowledge of the true covariance matrix. We show that the asymptotic efficiency loss of NERCOME is almost surely 0 with a suitable split location of the data. We also show that all the aforementioned optimality holds for data with a factor structure. Our method avoids the need to first estimate any unknowns from a factor model, and directly gives the covariance or precision matrix estimator, which can be useful when factor analysis is not the ultimate goal. We compare the performance of our estimators with other methods through extensive simulations and real data analysis.

REFERENCES

ABADIR, K. M., DISTASO, W. and ŽIKEŠ, F. (2014). Design-free estimation of variance matrices. *J. Econometrics* **181** 165–180. MR3209862

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

BICKEL, P. J. and LEVINA, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

BICKEL, P. J. and LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607

DEMIGUEL, V. and NOGALES, F. J. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* **55** 798–812.

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991

MSC2010 subject classifications. Primary 62H12; secondary 62G20, 15B52.

Key words and phrases. High dimensional data analysis, covariance matrix, Stieltjes transform, data splitting, nonlinear shrinkage, factor model.

FAN, J., LIAO, Y. and MINCHEVA, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **39** 3320–3356. MR3012410

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

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

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob. Contributions to the Theory of Statistics* **1** 361–379. Univ. California Press, Berkeley, CA. MR0133191

LAM, C. (2015). Supplement to "Nonparametric eigenvalue-regularized precision or covariance matrix estimator." DOI:10.1214/15-AOS1393SUPP.

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

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

LAM, C., YAO, Q. and BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98** 901–918. MR2860332

LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718

LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88** 365–411. MR2026339

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942

LEDOIT, O. and WOLF, M. (2013a). Optimal estimation of a large-dimensional covariance matrix under Stein's loss. ECON—Working Papers 122, Dept. Economics, Univ. Zürich.

LEDOIT, O. and WOLF, M. (2013b). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. ECON—Working Papers 105, Dept. Economics, Univ. Zürich.

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

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

POURAHMADI, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika* **94** 1006–1013. MR2376812

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. MR2504372

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

STEIN, C. (1975). Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting IMS*. Atlanta, GA.

STEIN, C. (1986). Lectures on the theory of estimation of many parameters. *J. Sov. Math.* **34** 1373–1403.

STOCK, J. and WATSON, M. (2005). Implications of dynamic factor models for var analysis. NBER working papers. No. 11467.

WON, J.-H., LIM, J., KIM, S.-J. and RAJARATNAM, B. (2013). Condition-number-regularized covariance estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 427–450. MR3065474

The Annals of Statistics
2016, Vol. 44, No. 3, 954–981
DOI: 10.1214/15-AOS1394

GLOBAL RATES OF CONVERGENCE OF THE MLES OF LOG-CONCAVE AND s-CONCAVE DENSITIES

BY CHARLES R. DOSS AND JON A. WELLNER

University of Minnesota and University of Washington

We establish global rates of convergence for the Maximum Likelihood Estimators (MLEs) of log-concave and s-concave densities on $\mathbb{R}$. The main finding is that the rate of convergence of the MLE in the Hellinger metric is no worse than $n^{-2/5}$ when $-1 < s < \infty$ where $s = 0$ corresponds to the log-concave case. We also show that the MLE does not exist for the classes of s-concave densities with $s < -1$.

REFERENCES

[1] ADLER, R. J., FELDMAN, R. E. and TAQQU, M. S., eds. (1998). *A Practical Guide to Heavy Tails*. Birkhäuser, Boston, MA. MR1652283

[2] BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann*. *Statist*. **37** 1299–1331. MR2509075

[3] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann*. *Statist*. **25** 970–981. MR1447736

[4] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719

[5] BORELL, C. (1974). Convex measures on locally convex spaces. *Ark*. *Mat*. **12** 239–252. MR0388475

[6] BORELL, C. (1975). Convex set functions in d-space. *Period*. *Math*. *Hungar.* **6** 111–136. MR0404559

[7] BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J*. *Funct*. *Anal*. **22** 366–389. MR0450480

[8] BRONŠTEĬN, E. M. (1976). ϵ-entropy of convex sets and functions. *Sibirsk*. *Mat*. *Zh*. **17** 508–514, 715. MR0415155

[9] CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron*. *J*. *Stat*. **4** 254–270. MR2645484

[10] CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **72** 545–607. MR2758237

[11] DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality*, *Convexity*, *and Applications*. Academic Press, Boston, MA. MR0954608

[12] DOSS, C. R. and WELLNER, J. A. (2015). Supplement to "Global rates of convergence of the MLEs of log-concave and s-concave densities." DOI:10.1214/15-AOS1394SUPP.

MSC2010 subject classifications. Primary 62G07; secondary 62G05, 62G20.

Key words and phrases. Bracketing entropy, consistency, empirical processes, global rate, Hellinger metric, log-concave, s-concave.

[13] DOSS, C. R. (2013). Shape-constrained inference for concave-transformed densities and their modes. Ph.D. thesis, Dept. Statistics, Univ. Washington, Seattle, WA.

[14] DOSS, C. R. and WELLNER, J. A. (2013). Global rates of convergence of the MLEs of log-concave and s-concave densities. Available at arXiv:1306.1438.

[15] DOSS, C. R. and WELLNER, J. A. (2015). Inference for the mode of a log-concave density. Technical report, Univ. Washington, Seattle, WA.

[16] DRYANOV, D. (2009). Kolmogorov entropy for classes of convex functions. *Constr. Approx.* **30** 137–153. MR2519658

[17] DUDLEY, R. M. (1984). A course on empirical processes. In *École D'été de Probabilités de Saint-Flour, XII—*1982. *Lecture Notes in Math*. **1097** 1–142. Springer, Berlin. MR0876079

[18] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712

[19] DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

[20] DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist*. **39** 702–730. MR2816336

[21] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist*. **40** 385–411. MR3014311

[22] GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inform. Theory* **59** 1957–1965. MR3043776

[23] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621

[24] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359.

[25] KIM, A. K. H. and SAMWORTH, R. J. (2015). Global rates of convergence in log-concave density estimation. Available at arXiv:1404.2298v2.

[26] KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist*. **38** 2998–3027. MR2722462

[27] PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function$_2$. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

[28] PRÉKOPA, A. (1995). *Stochastic Programming*. *Mathematics and Its Applications* **324**. Kluwer Academic, Dordrecht. MR1375234

[29] RESNICK, S. I. (2007). *Heavy-Tail Phenomena*: *Probabilistic and Statistical Modeling*. Springer, New York. MR2271424

[30] RINOTT, Y. (1976). On convexity of measures. *Ann. Probab.* **4** 1020–1026. MR0428540

[31] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

[32] SCHUHMACHER, D., HÜSLER, A. and DÜMBGEN, L. (2011). Multivariate log-concave distributions as a nearly parametric model. *Stat. Risk Model*. **28** 277–295. MR2838319

[33] SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist*. **38** 3751–3781. MR2766867

[34] VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist*. **21** 14–44. MR1212164

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

[36] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

[37] Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

[38] Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570

The Annals of Statistics
2016, Vol. 44, No. 3, 982–1009
DOI: 10.1214/15-AOS1395

CLASSIFICATION IN GENERAL FINITE DIMENSIONAL SPACES WITH THE *k*-NEAREST NEIGHBOR RULE

BY SÉBASTIEN GADAT*, THIERRY KLEIN† AND CLÉMENT MARTEAU‡

Toulouse School of Economics, Université Paul Sabatier†
and Université Lyon 1‡*

Given an n-sample of random vectors $(X_i, Y_i)_{1\le i\le n}$ whose joint law is unknown, the long-standing problem of supervised classification aims to *optimally* predict the label Y of a given new observation X. In this context, the k-nearest neighbor rule is a popular flexible and intuitive method in non-parametric situations. Even if this algorithm is commonly used in the machine learning and statistics communities, less is known about its prediction ability in general finite dimensional spaces, especially when the support of the density of the observations is $\mathbb{R}^d$. This paper is devoted to the study of the statistical properties of the k-nearest neighbor rule in various situations. In particular, attention is paid to the marginal law of X, as well as the smoothness and margin properties of the *regression function* $\eta(X) = \mathbb{E}[Y|X]$. We identify two necessary and sufficient conditions to obtain uniform consistency rates of classification and derive sharp estimates in the case of the k-nearest neighbor rule. Some numerical experiments are proposed at the end of the paper to help illustrate the discussion.

REFERENCES

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

[2] AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist*. **35** 608–633. MR2336861

[3] BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman & Hall, London. MR1010226

[4] BOTEV, Z. I., GROTOWSKI, J. F. and KROESE, D. P. (2010). Kernel density estimation via diffusion. *Ann. Statist*. **38** 2916–2957. MR2722460

[5] BOUCHERON, S., BOUSQUET, O. and LUGOSI, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM Probab. Stat*. **9** 323–375. MR2182250

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

[7] CANNINGS, T. (2013). Nearest neighbour classification in the tails of a distribution. Preprint.

[8] CÉROU, F. and GUYADER, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat*. **10** 340–355 (electronic). MR2247925

[9] CHAUDHURI, K. and DASGUPTA, S. (2014). Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) **27** 3437–3445. Curran Associates, Red Hook, NY.

MSC2010 subject classifications. Primary 62G05, 62F15; secondary 62G20.
Key words and phrases. Supervised classification rates, k-nearest neighbor, plug-in rules.

[10] DEVROYE, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.* **9** 1310–1319. MR0630113

[11] DEVROYE, L., GYÖRFI, L., KRZYŻAK, A. and LUGOSI, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Statist.* **22** 1371–1385. MR1311980

[12] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093

[13] FIX, E. and HODGES, J. L. (1951). Discriminatory analysis, nonparametric discrimination, consistency properties. Randolph Field, Texas, Project 21-49-004, Report 4.

[14] FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. MR1473055

[15] GADAT, S., KLEIN, T. and MARTEAU, C. (2015). Supplement to "Classification in general finite dimensional spaces with the k-nearest neighbor rule." DOI:10.1214/15-AOS1395SUPP.

[16] GOLDENSHLUGER, A. and LEPSKI, O. (2014). On Adaptive Minimax Density Estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001

[17] GYŐRFI, L. (1978). On the Rate of Convergence of Nearest Neighbor Rules. *IEEE Trans. Inform. Theory* **24** 509–512. MR0501595

[18] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York. MR1920390

[19] HALL, P., PARK, B. U. and SAMWORTH, R. J. (2008). Choice of neighbor order in nearest-neighbor classification. *Ann. Statist.* **36** 2135–2152. MR2458182

[20] HÜSLER, J., LIU, R. Y. and SINGH, K. (2002). A formula for the tail probability of a multivariate normal distribution and its applications. *J. Multivariate Anal.* **82** 422–430. MR1921635

[21] LECUÉ, G. (2007). Simultaneous adaptation to the margin and to complexity in classification. *Ann. Statist.* **35** 1698–1721. MR2351102

[22] LI, W. V. and SHAO, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes*: *Theory and Methods*. *Handbook of Statist.* **19** 533–597. North-Holland, Amsterdam. MR1861734

[23] LIAN, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electron. J. Stat.* **5** 31–40. MR2773606

[24] LOUSTAU, S. and MARTEAU, C. (2015). Minimax fast rates for discriminant analysis with errors in variables. *Bernoulli* **21** 176–208. MR3322316

[25] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618

[26] REYNAUD-BOURET, P., RIVOIRARD, V. and TULEAU-MALOT, C. (2011). Adaptive density estimation: A curse of support? *J. Statist. Plann. Inference* **141** 115–139. MR2719482

[27] RODRÍGUEZ CASAL, A. (2007). Set estimation under convexity type assumptions. *Ann. Inst. Henri Poincaré Probab. Stat.* **43** 763–774. MR3252430

[28] SAMWORTH, R. J. (2012). Optimal weighted nearest neighbour classifiers. *Ann. Statist.* **40** 2733–2763. MR3097618

[29] STEINWART, I. (2005). Consistency of support vector machines and other regularized kernel classifiers. *IEEE Trans. Inform. Theory* **51** 128–142. MR2234577

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

[31] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002

[32] VAPNIK, V. N. (1998). *Statistical Learning Theory*. Wiley, New York. MR1641250

The Annals of Statistics
2016, Vol. 44, No. 3, 1010–1037
DOI: 10.1214/15-AOS1396

A NEW PRIOR FOR DISCRETE DAG MODELS WITH A RESTRICTED SET OF DIRECTIONS

BY HÉLÈNE MASSAM[1] AND JACEK WESOŁOWSKI[2]

York University and Politechnika Warszawska

In this paper, we first develop a new family of conjugate prior distributions for the cell probability parameters of discrete graphical models Markov with respect to a set $\mathcal{P}$ of moral directed acyclic graphs with skeleton a given decomposable graph G. This family, which we call the $\mathcal{P}$-Dirichlet, is a generalization of the hyper Dirichlet given in [*Ann. Statist.* **21** (1993) 1272–1317]: it keeps the directed strong hyper Markov property for every DAG in $\mathcal{P}$ but increases the flexibility in the choice of its parameters, that is, the hyper parameters.

Our second contribution is a characterization of the $\mathcal{P}$-Dirichlet, which yields, as a corollary, a characterization of the hyper Dirichlet and a characterization of the Dirichlet also. Like the characterization of the Dirichlet given in [*Ann. Statist.* **25** (1997) 1344–1369], our characterization of the $\mathcal{P}$-Dirichlet is based on local and global independence of the probability parameters and also a separability property explicitly defined here but implicitly used in that paper through the choice of two particular DAGs. Another advantage of our approach is that we need not make the assumption of the existence of a positive density function. We use the method of moments for our proofs.

REFERENCES

[1] ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. MR1439312

[2] BOBECKA, K. and WESOŁOWSKI, J. (2007). The Dirichlet distribution and process through neutralities. *J. Theoret. Probab.* **20** 295–308. MR2324532

[3] BOBECKA, K. and WESOŁOWSKI, J. (2009). Moments method approach to characterizations of Dirichlet tables through neutralities. *Publ. Math. Debrecen* **74** 321–339. MR2521378

[4] CHANG, W.-Y., GUPTA, R. D. and RICHARDS, D. ST. P. (2010). Structural properties of the generalized Dirichlet distributions. In *Algebraic Methods in Statistics and Probability II* (M. A. G. Viana and H. P. Wynn, eds.). *Contemp. Math.* **516** 109–124. Amer. Math. Soc., Providence, RI. MR2730744

[5] DARROCH, J. N. and RATCLIFF, D. (1971). A characterization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **66** 641–643. MR0292213

[6] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267

[7] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

MSC2010 subject classifications. 62H17, 62F15, 62E99.

Key words and phrases. Bayesian learning, directed strong hyper Markov, conjugate priors, hyper Dirichlet distribution, characterization, local and global independence.

[8] FABIUS, J. (1973). Two characterizations of the Dirichlet distribution. *Ann. Statist.* **1** 583–587. MR0353531

[9] GEIGER, D. and HECKERMAN, D. (1997). A characterization of the Dirichlet distribution through global and local parameter independence. *Ann. Statist.* **25** 1344–1369. MR1447755

[10] HECKERMAN, GEIGER, D. and CHICKERING, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Mach. Learn.* **20** 197–243.

[11] JAMES, I. R. and MOSIMANN, J. E. (1980). A new characterization of the Dirichlet distribution through neutrality. *Ann. Statist.* **8** 183–189. MR0557563

[12] JÁRAI, A. (1998). Regularity property of the functional equation of the Dirichlet distribution. *Aequationes Math.* **56** 37–46. MR1628311

[13] JÁRAI, A. (2005). *Regularity Properties of Functional Equations in Several Variables. Advances in Mathematics (Springer)* **8**. Springer, New York. MR2130441

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

[15] MASSAM, H. and WESOŁOWSKI, J. (2015). Supplement to "A new prior for discrete DAG models with a restricted set of directions." DOI:10.1214/15-AOS1396SUPP.

[16] NG, K. W., TIAN, G.-L. and TANG, M.-L. (2011). *Dirichlet and Related Distributions*. Wiley, Chichester. MR2830563

[17] RAJARATNAM, B., MASSAM, H. and CARVALHO, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.* **36** 2818–2849. MR2485014

[18] RAMAMOORTHI, R. V. and SANGALLI, L. M. (2007). On a characterization of the Dirichlet distribution. In *Bayesian Statistics and Its Applications* (S. K. Upadhyay, U. Singh and D. K. Dey, eds.) 385–397. Anamaya Publishers, New Delhi.

[19] SAKOWICZ, A. and WESOŁOWSKI, J. (2014). Dirichlet distribution through neutralities with respect to two partitions. *J. Multivariate Anal.* **129** 1–15. MR3215975

[20] STUDENÝ, M. and VOMLEL, J. (2009). A reconstruction algorithm for the essential graph. *Internat. J. Approx. Reason.* **50** 385–413. MR2514506

The Annals of Statistics
2016, Vol. 44, No. 3, 1038–1068
DOI: 10.1214/15-AOS1397

SLOPE IS ADAPTIVE TO UNKNOWN SPARSITY AND ASYMPTOTICALLY MINIMAX

BY WEIJIE SU[1] AND EMMANUEL CANDÈS[2]

Stanford University

We consider high-dimensional sparse regression problems in which we observe $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}$, where $\mathbf{X}$ is an $n \times p$ design matrix and $\mathbf{z}$ is an n-dimensional vector of independent Gaussian errors, each with variance σ^2. Our focus is on the recently introduced SLOPE estimator [*Ann. Appl. Stat.* **9** (2015) 1103–1140], which regularizes the least-squares estimates with the rank-dependent penalty $\sum_{1 \le i \le p} \lambda_i |\widehat{\beta}|_{(i)}$, where $|\widehat{\beta}|_{(i)}$ is the ith largest magnitude of the fitted coefficients. Under Gaussian designs, where the entries of $\mathbf{X}$ are i.i.d. $\mathcal{N}(0, 1/n)$, we show that SLOPE, with weights λ_i just about equal to $\sigma \cdot \Phi^{-1}(1 - iq/(2p))$ [$\Phi^{-1}(\alpha)$ is the αth quantile of a standard normal and q is a fixed number in $(0, 1)$] achieves a squared error of estimation obeying

$$\sup_{\|\boldsymbol{\beta}\|_0 \le k} \mathbb{P}\big(\|\widehat{\boldsymbol{\beta}}_{\text{SLOPE}} - \boldsymbol{\beta}\|^2 > (1 + \varepsilon) 2\sigma^2 k \log(p/k)\big) \longrightarrow 0$$

as the dimension p increases to ∞, and where $\varepsilon > 0$ is an arbitrary small constant. This holds under a weak assumption on the ℓ_0-sparsity level, namely, $k/p \to 0$ and $(k \log p)/n \to 0$, and is sharp in the sense that this is the best possible error *any* estimator can achieve. A remarkable feature is that SLOPE does not require any knowledge of the degree of sparsity, and yet automatically adapts to yield optimal total squared errors over a wide range of ℓ_0-sparsity classes. We are not aware of any other estimator with this property.

REFERENCES

[1] ABRAMOVICH, F. and BENJAMINI, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.* **22** 351–361. MR1411575

[2] ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879

[3] ADKE, S. R., WAIKAR, V. B. and SCHUURMANN, F. J. (1987). A two-stage shrinkage testimator for the mean of an exponential distribution. *Comm. Statist. Theory Methods* **16** 1821–1834. MR0903358

[4] BARAUD, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.* **6** 127–146 (electronic). MR1918295

[5] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876

MSC2010 subject classifications. Primary 62C20; secondary 62G05, 62G10, 62J15.

Key words and phrases. SLOPE, sparse regression, adaptivity, false discovery rate (FDR), Benjamini–Hochberg procedure, FDR thresholding.

[6] 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

[7] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inform. Theory* **58** 1997–2017. MR2951312

[8] 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

[9] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

[10] BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

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

[12] BIRGÉ, L. (2004). Model selection for Gaussian regression with random design. *Bernoulli* **10** 1039–1051. MR2108042

[13] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946

[14] BLAND, J. M. and ALTMAN, D. G. (1995). Multiple significance tests: The Bonferroni method. *The British Medical Journal* **310** 170.

[15] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717

[16] BOGDAN, M., VAN DEN BERG, E., SU, W. and CANDÈS, E. J. (2013). Statistical estimation and testing via the sorted ℓ_1 norm. Preprint. Available at arXiv:1310.1969.

[17] BONDELL, H. D. and REICH, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64** 115–123, 322–323. MR2422825

[18] BROWN, L. D., CAI, T. T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688–707. MR1922538

[19] BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. MR2351101

[20] CAI, T. T. and ZHOU, H. H. (2009). A data-driven block thresholding approach to wavelet estimation. *Ann. Statist.* **37** 569–595. MR2502643

[21] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

[22] CANDÈS, E. J. and PLAN, Y. (2009). Near-ideal model selection by ℓ_1 minimization. *Ann. Statist.* **37** 2145–2177. MR2543688

[23] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846

[24] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

[25] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886

[26] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464

[27] DONOHO, D. L., JOHNSTONE, I. M., MALEKI, A. and MONTANARI, A. (2011). Compressed sensing over ℓ_p-balls: Minimax mean square error. In *Proceedings of the IEEE International Symposium on Information Theory* 129–133. IEEE, New York.

[28] DONOHO, D. L. and MONTANARI, A. (2013). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. Preprint. Available at arXiv:1310.7320.
[29] DONOHO, D. L. and TANNER, J. (2009). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22** 1–53. MR2449053
[30] DONOHO, D. L. and TANNER, J. (2010). Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling. *IEEE Trans. Inform. Theory* **56** 2002–2016. MR2654490
[31] FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887
[32] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581
[33] FIGUEIREDO, M. and NOWAK, R. (2014). Sparse estimation with strongly correlated variables using ordered weighted ℓ_1 regularization. Preprint. Available at arXiv:1409.4005.
[34] FOSTER, D. P. and GEORGE, E. I. (1994). The risk inflation criterion for multiple regression. *Ann. Statist.* **22** 1947–1975. MR1329177
[35] FOSTER, D. P. and STINE, R. A. (1999). Local asymptotic coding and the minimum description length. *IEEE Trans. Inform. Theory* **45** 1289–1293. MR1686271
[36] G'SELL, M., WAGER, S., CHOULDECHOVA, A. and TIBSHIRANI, R. (2013). Sequential selection procedures and false discovery rate control. Preprint. Available at arXiv:1309.5352.
[37] GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988. MR2108039
[38] JI, P. and ZHAO, Z. (2014). Rate optimal multiple testing procedure in high-dimensional regression. Preprint. Available at arXiv:1404.2961.
[39] JIANG, W. and ZHANG, C.-H. (2013). A nonparametric empirical Bayes approach to adaptive minimax estimation. *J. Multivariate Anal.* **122** 82–95. MR3189309
[40] JOHNSTONE, I. M. (2013). Gaussian estimation: Sequence and wavelet models. Available at http://statweb.stanford.edu/~imj/GE06-11-13.pdf.
[41] KRUSKAL, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29** 115–129. MR0169713
[42] LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453
[43] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970
[44] MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2011). *Inequalities*: *Theory of Majorization and Its Applications*, 2nd ed. Springer, New York. MR2759813
[45] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. MR2488351
[46] PARIKH, N. and BOYD, S. (2013). Proximal algorithms. *Foundations and Trends in Optimization* **1** 123–231.
[47] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. MR2882274
[48] RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343
[49] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098
[50] SU, W. and CANDÈS, E. (2015). Supplement to "SLOPE is adaptive to unknown sparsity and asymptotically minimax." DOI:10.1214/15-AOS1397SUPP.

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

[52] TIBSHIRANI, R. and KNIGHT, K. (1999). The covariance inflation criterion for adaptive model selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 529–546. MR1707859

[53] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316

[54] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672

[55] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873

[56] WU, Z. and ZHOU, H. H. (2013). Model selection and sharp asymptotic minimaxity. *Probab. Theory Related Fields* **156** 165–191. MR3055256

[57] YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. MR2756192

[58] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

[59] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

The Annals of Statistics
2016, Vol. 44, No. 3, 1069–1102
DOI: 10.1214/15-AOS1398

SUPREMUM NORM POSTERIOR CONTRACTION AND CREDIBLE SETS FOR NONPARAMETRIC MULTIVARIATE REGRESSION

BY WILLIAM WEIMIN YOO AND SUBHASHIS GHOSAL

Université Paris Dauphine and North Carolina State University

In the setting of nonparametric multivariate regression with unknown error variance σ^2, we study asymptotic properties of a Bayesian method for estimating a regression function f and its mixed partial derivatives. We use a random series of tensor product of B-splines with normal basis coefficients as a prior for f, and σ is either estimated using the empirical Bayes approach or is endowed with a suitable prior in a hierarchical Bayes approach. We establish pointwise, L_2 and L_∞-posterior contraction rates for f and its mixed partial derivatives, and show that they coincide with the minimax rates. Our results cover even the anisotropic situation, where the true regression function may have different smoothness in different directions. Using the convergence bounds, we show that pointwise, L_2 and L_∞-credible sets for f and its mixed partial derivatives have guaranteed frequentist coverage with optimal size. New results on tensor products of B-splines are also obtained in the course.

REFERENCES

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

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

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

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

[5] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468

[6] CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392

[7] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525

[8] DE BOOR, C. (2001). *A Practical Guide to Splines*, Revised ed. Springer, New York. MR1900298

[9] DE JONGE, R. and VAN ZANTEN, J. H. (2012). Adaptive estimation of multivariate functions using conditionally Gaussian tensor-product spline priors. *Electron. J. Stat.* **6** 1984–2001. MR3020254

MSC2010 subject classifications. Primary 62G08; secondary 62G05, 62G15, 62G20.
Key words and phrases. Tensor product B-splines, sup-norm posterior contraction, nonparametric multivariate regression, mixed partial derivatives, anisotropic smoothness.

[10] DE JONGE, R. and VAN ZANTEN, J. H. (2013). Semiparametric Bernstein–von Mises for the error standard deviation. *Electron. J. Stat.* **7** 217–243. MR3020419

[11] DEMKO, S., MOSS, W. F. and SMITH, P. W. (1984). Decay rates for inverses of band matrices. *Math. Comp.* **43** 491–499. MR0758197

[12] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119

[13] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

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

[15] HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

[16] HOFFMANN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. MR1902892

[17] HOFFMANN, M., ROUSSEAU, J. and SCHMIDT-HIEBER, J. (2015). On adaptive posterior concentration rates. *Ann. Statist.* **43** 2259–2295. MR3396985

[18] 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

[19] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods* **42** 1294–1313. MR3031282

[20] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. MR2802048

[21] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin. MR1102015

[22] NEUMANN, M. H. and VON SACHS, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25** 38–76. MR1429917

[23] RAY, K. (2015). Adaptive Bernstein–von Mises theorems in Gaussian white noise. Preprint. Available at arXiv:1407.3397v2.

[24] SCHUMAKER, L. L. (2007). *Spline Functions: Basic Theory*, 3rd ed. Cambridge Univ. Press, Cambridge. MR2348176

[25] SCRICCIOLO, C. (2014). Adaptive Bayesian density estimation in L^p-metrics with Pitman–Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Anal.* **9** 475–520. MR3217004

[26] SEARLE, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley, Chichester. MR0670947

[27] SERRA, P. and KRIVOBOKOVA, T. (2014). Adaptive empirical Bayesian smoothing splines. Preprint. Available at arXiv:1411.6860.

[28] SHEN, W. and GHOSAL, S. (2015). Adaptive Bayesian procedures using random series priors. *Scand. J. Stat.* **42** 1194–1213.

[29] SHEN, W. and GHOSAL, S. (2016). Adaptive Bayesian density regression for high-dimensional data. *Bernoulli* **22** 396–420.

[30] SMIRNOV, N. V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR* (*N.S.*) **74** 189–191. MR0037494

[31] SNIEKERS, S. and VAN DER VAART, A. W. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166** 78–86. MR3390135

[32] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. MR0594650

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

[34] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

[35] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

[36] ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277

The Annals of Statistics
2016, Vol. 44, No. 3, 1103–1130
DOI: 10.1214/15-AOS1399

OPTIMAL DESIGNS FOR COMPARING CURVES[1]

BY HOLGER DETTE AND KIRSTEN SCHORNING

Ruhr-Universität Bochum

We consider the optimal design problem for a comparison of two regression curves, which is used to establish the similarity between the dose response relationships of two groups. An optimal pair of designs minimizes the width of the confidence band for the difference between the two regression functions. Optimal design theory (equivalence theorems, efficiency bounds) is developed for this non-standard design problem and for some commonly used dose response models optimal designs are found explicitly. The results are illustrated in several examples modeling dose response relationships. It is demonstrated that the optimal pair of designs for the comparison of the regression curves is *not* the pair of the optimal designs for the individual models. In particular, it is shown that the use of the optimal designs proposed in this paper instead of commonly used "non-optimal" designs yields a reduction of the width of the confidence band by more than 50%.

REFERENCES

ANTILLE, G., DETTE, H. and WEINBERG, A. (2003). A note on optimal designs in weighted polynomial regression for the classical efficiency functions. *J. Statist. Plann. Inference* **113** 285–292. MR1963047

BONYADI, M. R. and MICHALEWICZ, Z. (2014). A locally convergent rotationally invariant particle swarm optimization algorithm. *Swarm Intell.* **8** 159–198.

BRETZ, F., PINHEIRO, J. C. and BRANSON, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61** 738–748. MR2196162

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

CHANG, F.-C. (2005a). D-optimal designs for weighted polynomial regression—a functional-algebraic approach. *Statist. Sinica* **15** 153–163. MR2125725

CHANG, F.-C. (2005b). D-optimal designs for weighted polynomial regression—a functional approach. *Ann. Inst. Statist. Math.* **57** 833–844. MR2233552

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.* **24** 586–602. MR0058932

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

DETTE, H. (1990). A generalization of D- and D_1-optimal designs in polynomial regression. *Ann. Statist.* **18** 1784–1804. MR1074435

DETTE, H. (1993). A note on E-optimal designs for weighted polynomial regression. *Ann. Statist.* **21** 767–771. MR1232518

DETTE, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59** 97–110. MR1436556

MSC2010 subject classifications. 62K05.
Key words and phrases. Similarity of regression curves, confidence band, optimal design, robust design.

DETTE, H., BORNKAMP, B. and BRETZ, F. (2013). On the efficiency of two-stage response-adaptive designs. *Stat. Med.* **32** 1646–1660. MR3060632

DETTE, H. and MELAS, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39** 1266–1281. MR2816354

DETTE, H. and NEUMEYER, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29** 1361–1400. MR1873335

DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2008). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Comput. Statist. Data Anal.* **53** 312–320. MR2649087

DETTE, H. and TRAMPISCH, M. (2010). A general approach to D-optimal designs for weighted univariate polynomial regression models. *J. Korean Statist. Soc.* **39** 1–26. MR2655806

DETTE, H., BRETZ, F., PEPELYSHEV, A. and PINHEIRO, J. (2008). Optimal designs for dose-finding studies. *J. Amer. Statist. Assoc.* **103** 1225–1237. MR2462895

DRAGALIN, V., BORNKAMP, B., BRETZ, F., MILLER, F., PADMANABHAN, S. K., PATEL, N., PEREVOZSKAYA, I., PINHEIRO, J. and SMITH, J. R. (2010). A simulation study to compare new adaptive dose-ranging designs. *Stat. Biopharm. Res.* **2** 487–512.

GSTEIGER, S., BRETZ, F. and LIU, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *J. Biopharm. Statist.* **21** 708–725. MR2800311

HALL, P. and HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85** 1039–1049. MR1134500

HEILIGERS, B. (1994). E-optimal designs in weighted polynomial regression. *Ann. Statist.* **22** 917–929. MR1292548

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York. MR0204922

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879. MR0356386

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366. MR0117842

LAU, T.-S. and STUDDEN, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *Ann. Statist.* **13** 383–394. MR0773174

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379–398. MR0440812

LIU, W., JAMSHIDIAN, M. and ZHANG, Y. (2004). Multiple comparison of several linear regression models. *J. Amer. Statist. Assoc.* **99** 395–403. MR2062825

LIU, W., BRETZ, F., HAYTER, A. J. and WYNN, H. P. (2009). Assessing nonsuperiority, noninferiority, or equivalence when comparing two regression models over a restricted covariate region. *Biometrics* **65** 1279–1287. MR2756516

PSHENICHNYI, B. N. (1971). *Necessary Conditions for an Extremum*. Dekker, New York. MR0276845

PUKELSHEIM, F. (2006). *Optimal Design of Experiments*. *Classics in Applied Mathematics* **50**. SIAM, Philadelphia, PA. MR2224698

PUKELSHEIM, F. and RIEDER, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79** 763–770. MR1209476

VAN DEN BERGH, F. and ENGELBRECHT, A. P. (2010). A convergence proof for the particle swarm optimiser. *Fund. Inform.* **105** 341–374. MR2799460

WONG, W. K. and COOK, R. D. (1993). Heteroscedastic G-optimal designs. *J. Roy. Statist. Soc. Ser. B* **55** 871–880. MR1229885

YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 2499–2524. MR2676896

YANG, M., BIEDERMANN, S. and TANG, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *J. Amer. Statist. Assoc.* **108** 1411–1420. MR3174717

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *Ann. Statist.* **38** 1593–1606. MR2662353

Zen, M.-M. and Tsai, M.-H. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation in Fourier regression models. *J. Statist. Plann. Inference* **124** 475–487. MR2081498

The Annals of Statistics
2016, Vol. 44, No. 3, 1131–1164
DOI: 10.1214/15-AOS1400

RANDOMIZATION-BASED MODELS FOR MULTITIERED EXPERIMENTS: I. A CHAIN OF RANDOMIZATIONS

BY R. A. BAILEY[†] AND C. J. BRIEN[§]

University of St. Andrews[†], Queen Mary University of London[†], University of South Australia[§] and University of Adelaide[§]

We derive randomization-based models for experiments with a chain of randomizations. Estimation theory for these models leads to formulae for the estimators of treatment effects, their standard errors and expected mean squares in the analysis of variance. We discuss the practicalities in fitting these models and outline the difficulties that can occur, many of which do not arise in two-tiered experiments.

REFERENCES

[1] ALEJANDRO, P. P., BAILEY, R. A. and CAMERON, P. J. (2003). Association schemes and permutation groups. *Discrete Math*. **266** 47–67. MR1991706

[2] BAILEY, R. A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144** 214–223. MR0625801

[3] BAILEY, R. A. (1991). Strata for randomized experiments (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol*. **53** 27–78. MR1094275

[4] BAILEY, R. A. (2008). *Design of Comparative Experiments*. Cambridge Univ. Press, Cambridge. MR2422352

[5] BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalized wreath products of permutation groups. *Proc. Lond. Math. Soc*. (3) **47** 69–82. MR0698928

[6] BARDIN, A. and AZAÏS, J.-M. (1990). Une hypothèse minimale pour une théorie des plans d'expériences randomisés. *Rev. Stat. Appl*. **38** 21–41. MR1080504

[7] BATES, D., MAECHLER, M. and BOLKER, B. (2016). lme4: Linear mixed-effects models using S4 classes. Available at http://cran.at.r-project.org/web/packages/lme4/index.html (R package version 1.1-10, accessed January 8, 2016).

[8] BRIEN, C. J. (1983). Analysis of variance tables based on experimental structure. *Biometrics* **39** 51–59.

[9] BRIEN, C. J. (1992). Factorial linear model analysis. Ph.D. thesis, Department of Plant Science, The University of Adelaide, Adelaide, South Australia. Available at http://hdl.handle.net/2440/37701 (accessed January 8, 2016).

[10] BRIEN, C. J. (2015). dae: functions useful in the design and ANOVA of experiments. Available at http://cran.at.r-project.org/web/packages/dae/index.html (R package version 2.7-0, accessed April 26, 2015).

[11] BRIEN, C. J. and BAILEY, R. A. (2006). Multiple randomizations (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol*. **68** 571–609. MR2301010

MSC2010 subject classifications. Primary 62B15, 62J05; secondary 62J10.

Key words and phrases. Analysis of variance, expected mean square, mixed model, multiphase experiments, multitiered experiments, randomization-based model, REML, structure, tier.

[12] BRIEN, C. J. and BAILEY, R. A. (2009). Decomposition tables for experiments. I. A chain of randomizations. *Ann. Statist.* **37** 4184–4213. MR2572457

[13] BRIEN, C. J. and BAILEY, R. A. (2010). Decomposition tables for experiments. II. Two–one randomizations. *Ann. Statist.* **38** 3164–3190. MR2722467

[14] BRIEN, C. J. and DEMÉTRIO, C. G. B. (2009). Formulating mixed models for experiments, including longitudinal experiments. *J. Agric. Biol. Environ. Stat.* **14** 253–280. MR2750840

[15] BRIEN, C. J., HARCH, B. D., CORRELL, R. L. and BAILEY, R. A. (2011). Multiphase experiments with at least one later laboratory phase. I. Orthogonal designs. *J. Agric. Biol. Environ. Stat.* **16** 422–450. MR2843135

[16] BRIEN, C. J. and PAYNE, R. W. (1999). Tiers, structure formulae and the analysis of complicated experiments. *The Statistician* **48** 41–52.

[17] BRIEN, C. J. and PAYNE, R. W. (2012). AMTIER Procedure. In *GenStat Reference Manual Release* 15, *Part* 3 *Procedures* 106–108. VSN International, Hemel Hempstead, UK. Available at http://www.vsni.co.uk/resources/documentation/genstat-reference-procedure-library/ (accessed January 8, 2016).

[18] BUTLER, D., CULLIS, B. R., GILMOUR, A. R. and GOGEL, B. J. (2009). *Mixed Models for S Language Environments*: *ASReml-R Reference Manual*. *Version* 3. DPI Publications, Brisbane. Available at http://www.vsni.co.uk/resources/documentation/ (accessed January 8, 2016).

[19] COX, D. R. (1958). The interpretation of the effects of non-additivity in the Latin square. *Biometrika* **45** 69–73.

[20] CURNOW, R. N. (1959). The analysis of a two phase experiment. *Biometrics* **15** 60–73.

[21] GILMOUR, S. G. and GOOS, P. (2009). Analysis of data from non-orthogonal multistratum designs in industrial experiments. *J. R. Stat. Soc. Ser. C Appl. Stat.* **58** 467–484. MR2750088

[22] GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-Latin squares. *J. R. Stat. Soc. Ser. B* **12** 286–291.

[23] HOUTMAN, A. M. and SPEED, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11** 1069–1085. MR0720254

[24] JAMES, A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Stat.* **28** 993–1002. MR0097142

[25] LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D. and SCHABENBERGER, O. (2006). *SAS for Mixed Models*, 2nd ed. SAS Press, Cary, NC.

[26] MCINTYRE, G. A. (1955). Design and analysis of two phase experiments. *Biometrics* **11** 324–334.

[27] MONOD, H. and BAILEY, R. A. (1992). Pseudofactors: Normal use to improve design and facilitate analysis. *J. R. Stat. Soc. Ser. C Appl. Stat.* **41** 317–336.

[28] NELDER, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41** 544–548. MR0065083

[29] NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. Ser. A* **283** 147–162. MR0176576

[30] NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc. Ser. A* **283** 163–178. MR0174156

[31] NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **30** 303–311. MR0234582

[32] NELDER, J. A. (1977). A reformulation of linear models. *J. Roy. Statist. Soc. Ser. A* **140** 48–76. MR0458743

[33] OJIMA, Y. (1998). General formulae for expectations, variances and covariances of the mean squares for staggered nested designs. *J. Appl. Stat.* **25** 785–799. MR1652264

[34] PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554. MR0319325

[35] PATTERSON, H. D. and THOMPSON, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings of the Eighth International Biometric Conference* (L. C. A. Corsten and T. Postelnicu, eds.) 197–207. Editura Academiei Republicii Socialiste România, Bucureşti.

[36] PAYNE, R. W. and WILKINSON, G. N. (1977). A general algorithm for analysis of variance. *J. R. Stat. Soc. Ser. C Appl. Stat.* **26** 251–260.

[37] PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed Effects Models in S and S-PLUS*. Springer, New York.

[38] R DEVELOPMENT CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.r-project.org/ (accessed January 8, 2016).

[39] SAS INSTITUTE INC. (2010). *SAS/STAT*® 9.22 *User's Guide*. SAS Institute Inc., Cary, NC.

[40] SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York. MR1190470

[41] SMITH, A., CULLIS, B. and GILMOUR, A. (2001). The analysis of crop variety evaluation data in Australia. *Aust. N. Z. J. Stat.* **43** 129–145. MR1855705

[42] SPEED, T. P. (1987). What is an analysis of variance? (with discussion). *Ann. Statist.* **15** 885–941. MR0902237

[43] SPEED, T. P. and BAILEY, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (*University of Western Australia*, 1980) (P. Schultz, C. E. Praeger and R. P. Sullivan, eds.). *Lecture Notes in Pure and Applied Mathematics* **74** 55–74. Dekker, New York. MR0647166

[44] SPEED, T. P. and BAILEY, R. A. (1987). Factorial dispersion models. *Int. Stat. Rev.* **55** 261–277. MR0963143

[45] TJUR, T. (1984). Analysis of variance models in orthogonal designs (with discusion). *Int. Stat. Rev.* **52** 33–81. MR0967202

[46] VSN INTERNATIONAL (2015). *GenStat for Windows*, 18th ed. VSN International, Hemel Hempstead, UK. Available at GenStat.co.uk.

[47] WILKINSON, G. N. (1970). A general recursive procedure for analysis of variance. *Biometrika* **57** 19–46.

[48] WOLDE-TSADIK, G. and AFIFI, A. A. (1980). A comparison of the "sometimes pool", "sometimes switch" and "never pool" procedures in the two-way ANOVA random effects model. *Technometrics* **22** 367–373. MR0585634

[49] WOOD, J. T., WILLIAMS, E. R. and SPEED, T. P. (1988). Non-orthogonal block structure in two-phase designs. *Aust. J. Stat.* **30A** 225–237.

[50] YATES, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* **26** 424–455. Reprinted with additional author's note in Yates, F. (1970) *Experimental Design: Selected Papers*, pages 147–180. Griffin, London.

The Annals of Statistics
2016, Vol. 44, No. 3, 1165–1192
DOI: 10.1214/15-AOS1401

VECTOR QUANTILE REGRESSION: AN OPTIMAL TRANSPORT APPROACH

BY GUILLAUME CARLIER[1], VICTOR CHERNOZHUKOV[2] AND ALFRED GALICHON[3]

Université Paris IX Dauphine, Massachusetts Institute of Technology and New York University

We propose a notion of conditional vector quantile function and a vector quantile regression. A *conditional vector quantile function* (CVQF) of a random vector Y, taking values in $\mathbb{R}^d$ given covariates $Z = z$, taking values in $\mathbb{R}^k$, is a map $u \longmapsto Q_{Y|Z}(u, z)$, which is monotone, in the sense of being a gradient of a convex function, and such that given that vector U follows a reference non-atomic distribution F_U, for instance uniform distribution on a unit cube in $\mathbb{R}^d$, the random vector $Q_{Y|Z}(U, z)$ has the distribution of Y conditional on $Z = z$. Moreover, we have a strong representation, $Y = Q_{Y|Z}(U, Z)$ almost surely, for some version of U. The *vector quantile regression* (VQR) is a linear model for CVQF of Y given Z. Under correct specification, the notion produces strong representation, $Y = \beta(U)^\top f(Z)$, for $f(Z)$ denoting a known set of transformations of Z, where $u \longmapsto \beta(u)^\top f(Z)$ is a monotone map, the gradient of a convex function and the quantile regression coefficients $u \longmapsto \beta(u)$ have the interpretations analogous to that of the standard scalar quantile regression. As $f(Z)$ becomes a richer class of transformations of Z, the model becomes nonparametric, as in series modelling. A key property of VQR is the embedding of the classical Monge–Kantorovich's optimal transportation problem at its core as a special case. In the classical case, where Y is scalar, VQR reduces to a version of the classical QR, and CVQF reduces to the scalar conditional quantile function. An application to multiple Engel curve estimation is considered.

REFERENCES

[1] BELLONI, A. and WINKLER, R. L. (2011). On multivariate quantiles under partial orders. *Ann. Statist.* **39** 1125–1179. MR2816350

[2] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809

[3] CARLIER, G., CHERNOZHULOV, V. and GALICHON, A. (2015). Supplement to "Vector quantile regression: An optimal transport approach." DOI:10.1214/15-AOS1401SUPP.

[4] CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753

[5] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125. MR2667913

MSC2010 subject classifications. Primary 62J99; secondary 62H05, 62G05.

Key words and phrases. Vector quantile regression, vector conditional quantile function, Monge–Kantorovich–Brenier.

[6] CUNHA, F., HECKMAN, J. and SCHENNACH, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* **78** 883–931.

[7] DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277. MR0356350

[8] DUCPETIAUX, E. (1855). *Budgets économiques des Classes Ouvrières en Belgique*, Bruxelles.

[9] DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552. MR0690575

[10] EKELAND, I., GALICHON, A. and HENRY, M. (2012). Comonotonic measures of multivariate risks. *Math. Finance* **22** 109–132. MR2881882

[11] EKELAND, I. and TÉMAM, R. (1999). *Convex Analysis and Variational Problems*, English ed. *Classics in Applied Mathematics* **28**. SIAM, Philadelphia, PA. MR1727362

[12] ENGEL, E. (1857). Die Produktions und Konsumptionsverhältnisse des Königreichs Sachsen. *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Misisteriums des Innerm* **8** 1–54.

[13] HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670

[14] KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

[15] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[16] KOENKER, R. and BASSETT, G. JR. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–61. MR0640165

[17] KOLTCHINSKII, V. I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477. MR1439309

[18] KONG, L. and MIZERA, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statist. Sinica* **22** 1589–1610. MR3027100

[19] LEHMANN, E. L. (1974). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, CA. MR2279708

[20] LE PLAY, F. (1855). Les Ouvriers Européens. Etudes sur les travaux, la vie domestique et la condition morale des populations ouvrières de l'Europe. Paris.

[21] MATZKIN, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica* **71** 1339–1375. MR2000250

[22] MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. MR1369395

[23] SERFLING, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Statist. Plann. Inference* **123** 259–278. MR2062982

[24] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[25] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin. MR2459454

[26] WEI, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.* **103** 397–409. MR2420242

[27] YU, K. and JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

The Annals of Statistics
2016, Vol. 44, No. 3, 1193–1233
DOI: 10.1214/15-AOS1403

STRUCTURE IDENTIFICATION IN PANEL DATA ANALYSIS[1]

BY YUAN KE, JIALIANG LI AND WENYANG ZHANG

Princeton University, National University of Singapore and University of York

Panel data analysis is an important topic in statistics and econometrics. In such analysis, it is very common to assume the impact of a covariate on the response variable remains constant across all individuals. While the modelling based on this assumption is reasonable when only the global effect is of interest, in general, it may overlook some individual/subgroup attributes of the true covariate impact. In this paper, we propose a data driven approach to identify the groups in panel data with interactive effects induced by latent variables. It is assumed that the impact of a covariate is the same within each group, but different between the groups. An EM based algorithm is proposed to estimate the unknown parameters, and a binary segmentation based algorithm is proposed to detect the grouping. We then establish asymptotic theories to justify the proposed estimation, grouping method, and the modelling idea. Simulation studies are also conducted to compare the proposed method with the existing approaches, and the results obtained favour our method. Finally, the proposed method is applied to analyse a data set about income dynamics, which leads to some interesting findings.

REFERENCES

AHN, S. C. and SCHMIDT, P. (1995). Efficient estimation of models for dynamic panel data. *J. Econometrics* **68** 5–27. MR1345704

ARELLANO, M. (2003). *Panel Data Econometrics*. Oxford Univ. Press, Oxford. MR2060514

BAI, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. MR1455175

BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313

BAI, J. and LI, K. (2014). Theory and methods of panel data models with interactive effects. *Ann. Statist.* **42** 142–170. MR3178459

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BALTAGI, B. H. (2005). *Econometric Analysis of Panel Data*. Wiley, Chichester.

BONDELL, H. D. and REICH, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64** 115–123. MR2422825

FRED, A. and JAIN, A. K. (2003). Robust data clustering. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **3** 128–136. IEEE, Madison, WI.

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. MR2415737

MSC2010 subject classifications. Primary 62F08; secondary 62F12.
Key words and phrases. Binary segmentation algorithm, homogeneity, EM algorithm, interactive effects, panel data.

HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565

HSIAO, C. (2003). *Analysis of Panel Data*, 2nd ed. *Econometric Society Monographs* **34**. Cambridge Univ. Press, Cambridge. MR1962511

JIANG, Q., WANG, H., XIA, Y. and JIANG, G. (2013). On a principal varying coefficient model. *J. Amer. Statist. Assoc.* **108** 228–236. MR3174615

KARIYA, T. and KURATA, H. (2004). *Generalized Least Squares*. Wiley, Chichester. MR2120002

KE, Z. T., FAN, J. and WU, Y. (2015). Homogeneity pursuit. *J. Amer. Statist. Assoc.* **110** 175–194. MR3338495

KE, Y., LI, J. and ZHANG, W. (2015). Supplement to "Structure identification in panel data analysis." DOI:10.1214/15-AOS1403SUPP.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York. MR0702834

PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York.

SHEN, X. and HUANG, H.-C. (2010). Grouping pursuit through a regularization solution surface. *J. Amer. Statist. Assoc.* **105** 727–739. MR2724856

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. MR2136641

YANG, S., YUAN, L., LAI, Y.-C., SHEN, X., WONKA, P. and YE, J. (2012). Feature groupingand selection over an undirected graph. In *Proceedings of the* 18*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 922–930. ACM, New York.

ZHU, Y., SHEN, X. and PAN, W. (2013). Simultaneous grouping pursuit and feature selection over an undirected graph. *J. Amer. Statist. Assoc.* **108** 713–725. MR3174654

The Annals of Statistics
2016, Vol. 44, No. 3, 1234–1268
DOI: 10.1214/15-AOS1404

INFERENCE FOR SINGLE-INDEX QUANTILE REGRESSION MODELS WITH PROFILE OPTIMIZATION

BY SHUJIE MA AND XUMING HE

University of California, Riverside and University of Michigan

Single index models offer greater flexibility in data analysis than linear models but retain some of the desirable properties such as the interpretability of the coefficients. We consider a pseudo-profile likelihood approach to estimation and testing for single-index quantile regression models. We establish the asymptotic normality of the index coefficient estimator as well as the optimal convergence rate of the nonparametric function estimation. Moreover, we propose a score test for the index coefficient based on the gradient of the pseudo-profile likelihood, and employ a penalized procedure to perform consistent model selection and model estimation simultaneously. We also use Monte Carlo studies to support our asymptotic results, and use an empirical example to illustrate the proposed method.

REFERENCES

ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* **55** 1465–1471. MR0923471

BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841

BHATTACHARYA, P. K. and GANGOPADHYAY, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415. MR1062716

BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd ed. Springer, New York. MR1640691

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. MR1105843

CHAUDHURI, P., DOKSUM, K. and SAMAROV, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744. MR1439320

CUI, X., HÄRDLE, W. K. and ZHU, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39** 1658–1688. MR2850216

DE BOOR, C. (2001). *A Practical Guide to Splines*, Revised ed. *Applied Mathematical Sciences* **27**. Springer, New York. MR1900298

DE GOOIJER, J. G. and ZEROM, D. (2003). On additive conditional quantiles with high-dimensional covariates. *J. Amer. Statist. Assoc.* **98** 135–146. MR1965680

FAN, J., HU, T. C. and TRUONG, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Stat.* **21** 433–446. MR1310087

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

Key words and phrases. Quantile regression, single-index, profile principle, polynomial spline, model selection, score test.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179

HAMMER, S. M., SQUIRES, K. E., HUGHES, M. D., GRIMES, J. M., DEMETER, L. M., CURRIER, J. S., ERON, J. J., FEINBERG, J. E., BALFOUR, H. H., DEYTON, L. R., CHODAKEWITZ, J. A. and FISCHL, M. A. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS clinical trials group 320 study team. *N. Engl. J. Med.* **337** 725–733.

HE, X., NG, P. and PORTNOY, S. (1998). Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 537–550. MR1625950

HE, X. and SHI, P. D. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparametr. Stat.* **3** 299–308. MR1291551

HE, X. and SHI, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal.* **58** 162–181. MR1405586

HOROWITZ, J. L. and LEE, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100** 1238–1249. MR2236438

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981

KAI, B., LI, R. and ZOU, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Statist.* **39** 305–332. MR2797848

KIM, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35** 92–108. MR2332270

KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787

KOCHERGINSKY, M., HE, X. and MU, Y. (2005). Practical confidence intervals for regression quantiles. *J. Comput. Graph. Statist.* **14** 41–55. MR2137889

KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge. MR2268657

KOENKER, R. (2011). Additive models for quantile regression: Model selection and confidence bandaids. *Braz. J. Probab. Stat.* **25** 239–262. MR2832886

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. MR1326417

KONG, E. and XIA, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory* **28** 730–768. MR2959124

LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. MR2766869

MA, S. and HE, X. (2015). Supplement to "Inference for single-index quantile regression models with profile optimization." DOI:10.1214/15-AOS1404SUPP.

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168

POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199. MR1128411

PORTNOY, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Statist.* **25** 414–434. MR1429932

SEVERINI, T. A. and STANISWALIS, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89** 501–511. MR1294076

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

WANG, H., LI, R. and TSAI, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94** 553–568. MR2410008

WANG, J. and YANG, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. MR2487893

WANG, H. J., ZHU, Z. and ZHOU, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37** 3841–3866. MR2572445

WU, Y. and LIU, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19** 801–817. MR2514189

WU, T. Z., YU, K. and YU, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101** 1607–1621. MR2610735

XIA, Y. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.* **97** 1162–1184. MR2276153

XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. MR1741980

YU, K. and JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. MR1951258

ZHU, L., HUANG, M. and LI, R. (2012). Semiparametric quantile regression with high-dimensional covariates. *Statist. Sinica* **22** 1379–1401. MR3027092

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

ZOU, Q. and ZHU, Z. (2014). M-estimators for single-index model using B-spline. *Metrika* **77** 225–246. MR3157984

The Annals of Statistics
2016, Vol. 44, No. 3, 1269–1297
DOI: 10.1214/15-AOS1405

THEORETICAL ANALYSIS OF NONPARAMETRIC FILAMENT ESTIMATION[1]

BY WANLI QIAO AND WOLFGANG POLONIK

University of California, Davis

This paper provides a rigorous study of the nonparametric estimation of filaments or ridge lines of a probability density f. Points on the filament are considered as local extrema of the density when traversing the support of f along the integral curve driven by the vector field of second eigenvectors of the Hessian of f. We "parametrize" points on the filaments by such integral curves, and thus both the estimation of integral curves and of filaments will be considered via a plug-in method using kernel density estimation. We establish rates of convergence and asymptotic distribution results for the estimation of both the integral curves and the filaments. The main theoretical result establishes the asymptotic distribution of the uniform deviation of the estimated filament from its theoretical counterpart. This result utilizes the extreme value behavior of nonstationary Gaussian processes indexed by manifolds $M_h, h \in (0, 1]$ as $h \to 0$.

REFERENCES

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.* **34** 326–349. MR2275244

BARROW, J. D., SONODA, D. H. and BHAVSAR, S. P. (1985). Minimal spanning tree, filaments and galaxy clustering. *Mon. Not. R. Astron. Soc.* **216** 17–35.

BHARADWAJ, S., BHAVSAR, S. P. and SHETH, J. V. (2004). The size of the longest filaments in the universe. *Astrophys. J.* **606** 25–31.

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

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2013). Uncertainty measures and limiting distributions for filament estimation. Preprint. Available at arXiv:1312.2098v1.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2014). Generalized mode and ridge estimation. Preprint. Available at arXiv:1406.1803.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges *Ann. Statist.* **43** 1896–1928.

CHENG, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **17** 790–799.

COMANICIU, D. and MEER, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 603–619.

DIETRICH, J. P., WERNER, N., CLOWE, D., FINOGUENOV, A., KITCHING, T., MILLER, L. and SIMIONESCU, A. (2012). A filament of dark matter between two clusters of galaxies. *Nature* **487** 202–204.

MSC2010 subject classifications. Primary 62G20; secondary 62G05.

Key words and phrases. Extreme value distribution, nonparametric curve estimation, integral curves, kernel density estimation.

EBERLY, D. (1996). *Ridges in Image and Data Analysis*. Kluwer, Boston, MA.
EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403. MR2195639
FUKUNAGA, K. and HOSTETLER, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inform. Theory* **21** 32–40. MR0388638
GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2009). On the path density of a gradient field. *Ann. Statist.* **37** 3236–3271. MR2549559
GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012a). The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.* **107** 788–799. MR2980085
GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012b). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939
GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2014). Nonparametric ridge estimation. *Ann. Statist.* **42** 1511–1545. MR3262459
GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2016). Nonparametric inference for density modes. *J. R. Stat. Soc. Ser. B.* **78** 99–126.
GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344
GRONWALL, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math.* (2) **20** 292–296. MR1502565
HALL, P., QIAN, W. and TITTERINGTON, D. M. (1992). Ridge finding from noisy data. *J. Comput. Graph. Statist.* **1** 197–211. MR1270818
HASTIE, T. and STUETZLE, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84** 502–516. MR1010339
KOLTCHINSKII, V., SAKHANENKO, L. and CAI, S. (2007). Integral curves of noisy vector fields and statistical problems in diffusion tensor imaging: Nonparametric kernel estimation and hypotheses testing. *Ann. Statist.* **35** 1576–1607. MR2351098
MIKHALEVA, T. L. and PITERBARG, V. I. (1996). On the distribution of the maximum of a Gaussian field with constant variance on a smooth manifold. *Theory Probab. Appl.* **41** 367–379.
NOVIKOV, D., COLOMBI, S. and DORÉ, O. (2006). Skeleton as a probe of the cosmic web: Two-dimensional case. *Mon. Not. R. Astron. Soc.* **366** 1201–1216.
OZERTEM, U. and ERDOGMUS, D. (2011). Locally defined principal curves and surfaces. *J. Mach. Learn. Res.* **12** 1249–1286. MR2804600
PIMBBLET, K. A., DRINKWATER, M. J. and HAWKRIGG, M. C. (2004). Inter-cluster filaments of galaxies programme: Abundance and distribution of filaments in the 2dFGRS catalogue. *Mon. Not. R. Astron. Soc.* **354** L61–L65.
PITERBARG, V. and STAMATOVICH, S. (2001). On maximum of Gaussian non-centered fields indexed on smooth manifolds. In *Asymptotic Methods in Probability and Statistics with Applications* (*St. Petersburg*, 1998) (N. Balakrishnan, I. A. Ibragimov, and V. B. Nevzorov, eds.) 189–203. Birkhäuser, Boston, MA. MR1890326
QIAO, W. (2013). On estimation of filamentary structures. Ph.D. thesis, Univ. California, Davis.
QIAO, W. and POLONIK, W. (2015). Extrema of locally stationary Gaussian fields on growing manifolds. Preprint. Available at arXiv:1510.06833.
QIAO, W. and POLONIK, W. (2016). Supplement to "Theoretical analysis of nonparametric filament estimation." DOI:10.1214/15-AOS1405SUPP.
ROSENBLATT, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Probab.* **4** 1009–1015. MR0428580
VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

The Annals of Statistics
2016, Vol. 44, No. 3, 1298–1331
DOI: 10.1214/15-AOS1406

SEMIPARAMETRIC EFFICIENT ESTIMATION FOR SHARED-FRAILTY MODELS WITH DOUBLY-CENSORED CLUSTERED DATA

BY YU-RU SU[1] AND JANE-LING WANG[2]

Fred Hutchinson Cancer Research Center and University of California, Davis

In this paper, we investigate frailty models for clustered survival data that are subject to both left- and right-censoring, termed "doubly-censored data". This model extends current survival literature by broadening the application of frailty models from right-censoring to a more complicated situation with additional left-censoring.

Our approach is motivated by a recent Hepatitis B study where the sample consists of families. We adopt a likelihood approach that aims at the nonparametric maximum likelihood estimators (NPMLE). A new algorithm is proposed, which not only works well for clustered data but also improve over existing algorithm for independent and doubly-censored data, a special case when the frailty variable is a constant equal to one. This special case is well known to be a computational challenge due to the left-censoring feature of the data. The new algorithm not only resolves this challenge but also accommodates the additional frailty variable effectively.

Asymptotic properties of the NPMLE are established along with semiparametric efficiency of the NPMLE for the finite-dimensional parameters. The consistency of Bootstrap estimators for the standard errors of the NPMLE is also discussed. We conducted some simulations to illustrate the numerical performance and robustness of the proposed algorithm, which is also applied to the Hepatitis B data.

REFERENCES

BOOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Statist. Soc. B* **61** 265–285.

CAFFO, B. S., JANK, W. and JONES, G. L. (2005). Ascent-based Monte Carlo expectation-maximization. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 235–251. MR2137323

CAI, T. and CHENG, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika* **91** 277–290. MR2081301

CHAN, K. S. and LEDOLTER, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Amer. Statist. Assoc.* **90** 242–252. MR1325132

CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** 391–404. MR1041399

CHANG, M. N. and YANG, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547. MR0913572

MSC2010 subject classifications. Primary 62N02; secondary 62E20.

Key words and phrases. Frailty model, semi-parametric efficiency, EM algorithm, Monte Carlo integrations.

CHENG, G. (2015). Moment consistency of the exchangeably weighted bootstrap for semiparametric M-estimation. *Scand. J. Stat.* **42** 665–684.

CHENG, G. and HUANG, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38** 2884–2915. MR2722459

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. MR0400509

DE GRUTTOLA, V. and LAGAKOS, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45** 1–11. MR0999438

DUPUY, J.-F., GRAMA, I. and MESBAH, M. (2006). Asymptotic theory for the Cox model with missing time-dependent covariate. *Ann. Statist.* **34** 903–924. MR2283397

FORT, G. and MOULINES, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.* **31** 1220–1259. MR2001649

KIM, Y.-J. (2006). Regression analysis of doubly censored failure time data with frailty. *Biometrics* **62** 458–464. MR2227493

KIM, M. Y., DE GRUTTOLA, V. and LAGAKOS, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics* **49** 13–22.

KIM, Y., KIM, B. and JANG, W. (2010). Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data. *J. Multivariate Anal.* **101** 1339–1351. MR2609496

KIM, Y., KIM, J. and JANG, W. (2013). An EM algorithm for the proportional hazards model with doubly censored data. *Comput. Statist. Data Anal.* **57** 41–51. MR2981071

MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731. MR1292537

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23** 182–198. MR1331663

MURPHY, S. A., ROSSINI, A. J. and VAN DER VAART, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.* **92** 968–976. MR1482127

MYKLAND, P. A. and REN, J.-J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.* **24** 1740–1764. MR1416658

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. and SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Stat.* **19** 25–43. MR1172965

PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26** 183–214. MR1611788

RIPATTI, S. and PALMGREN, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56** 1016–1022. MR1806744

SU, Y.-R. (2011). Survival analysis for incomplete data. Ph.D. thesis, Univ. California, Davis.

SU, Y. and WANG, J. (2015). Supplement to "Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data." DOI:10.1214/15-AOS1406SUPP.

THERNEAU, T. M., GRAMBSCH, P. M. and PANKRATZ, V. S. (2003). Penalized survival models and frailty. *J. Comput. Graph. Statist.* **12** 156–175. MR1965213

TSENG, Y.-K., HSIEH, F. and WANG, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92** 587–603. MR2202648

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173. MR0381120

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

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

VAUPEL, J. W., MANTON, K. G. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.

WU, J. F., CHEN, C. C., HSIEH, R. P., SHIH, H. H., CHEN, Y. H., LI, C. R., CHIANG, C. Y., SHAU, W. Y., NI, Y. H., CHEN, H. L., HSU, H. Y. and CHANG, M. H. (2006). HLA typing associated with hepatitis B E antigen seroconversion in children with chronic hepatitis B virus infection: A long-term prospective sibling cohort study in Taiwan. *J. Pediatr.* **148** 647–651.

ZENG, D. and CAI, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Statist.* **33** 2132–2163. MR2211082

ZHANG, Y. and JAMSHIDIAN, M. (2004). On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *J. Comput. Graph. Statist.* **13** 123–140. MR2044874

The Annals of Statistics
2016, Vol. 44, No. 3, 1332–1359
DOI: 10.1214/15-AOS1408

APPROXIMATION AND ESTIMATION OF s-CONCAVE DENSITIES VIA RÉNYI DIVERGENCES

BY QIYANG HAN AND JON A. WELLNER[1]

University of Washington

In this paper, we study the approximation and estimation of s-concave densities via Rényi divergence. We first show that the approximation of a probability measure Q by an s-concave density exists and is unique via the procedure of minimizing a divergence functional proposed by [*Ann. Statist.* **38** (2010) 2998–3027] if and only if Q admits full-dimensional support and a first moment. We also show continuity of the divergence functional in Q: if $Q_n \to Q$ in the Wasserstein metric, then the projected densities converge in weighted L_1 metrics and uniformly on closed subsets of the continuity set of the limit. Moreover, directional derivatives of the projected densities also enjoy local uniform convergence. This contains both on-the-model and off-the-model situations, and entails strong consistency of the divergence estimator of an s-concave density under mild conditions. One interesting and important feature for the Rényi divergence estimator of an s-concave density is that the estimator is intrinsically related with the estimation of log-concave densities via maximum likelihood methods. In fact, we show that for $d = 1$ at least, the Rényi divergence estimators for s-concave densities converge to the maximum likelihood estimator of a log-concave density as $s \nearrow 0$. The Rényi divergence estimator shares similar characterizations as the MLE for log-concave distributions, which allows us to develop pointwise asymptotic distribution theory assuming that the underlying density is s-concave.

REFERENCES

AVRIEL, M. (1972). r-convex functions. *Math. Program.* **2** 309–323. MR0301151

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. MR2382657

BASU, A., HARRIS, I. R., HJORT, N. L. and JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559. MR1665873

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719

BORELL, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. MR0388475

BORELL, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6** 111–136. MR0404559

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480

MSC2010 subject classifications. Primary 62G07, 62H12; secondary 62G05, 62G20.

Key words and phrases. s-concavity, consistency, projection, asymptotic distribution, mode estimation, nonparametric estimation, shape constraints.

BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc*. *Sympos*., *Indiana Univ*., *Bloomington*, *Ind*., 1969) 177–197. Cambridge Univ. Press, London. MR0277070

CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron*. *J*. *Stat*. **4** 254–270. MR2645484

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J*. *R*. *Stat*. *Soc*. *Ser*. *B Stat*. *Methodol*. **72** 545–607. MR2758237

DAS GUPTA, S. (1976). *S*-unimodal function: Related inequalities and statistical applications. *Sankhyā Ser. B* **38** 301–314. MR0651042

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality*, *Convexity*, *and Applications*. Academic Press, Boston, MA. MR0954608

DOSS, C. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann*. *Statist*. **44** 954–981.

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann*. *Statist*. **39** 702–730. MR2816336

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol. II* (*Berkeley*, *Calif*., 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann*. *Statist*. **29** 1653–1698. MR1891742

GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621

HAN, Q. and WELLNER, J. A. (2015). Supplement to "Approximation and estimation of s-concave densities via Rényi divergences." DOI:10.1214/15-AOS1408SUPP.

JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist*. *Probab*. *Lett*. **50** 279–284. MR1792307

KIM, A. K. and SAMWORTH, R. J. (2015). Global rates of convergence in log-concave density estimation. Preprint. Available at arXiv:1404.2298v2.

KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann*. *Statist*. **38** 2998–3027. MR2722462

KOENKER, R. and MIZERA, I. (2014). Convex optimization in R. *J*. *Stat*. *Softw*. **60**(5) 1–23.

LANG, R. (1986). A note on the measurability of convex sets. *Arch*. *Math*. (*Basel*) **47** 90–92. MR0855142

MOSEK APS DENMARK (2011). The MOSEK Optimization Tools Manual, Version 6.0. Available at http://www.mosek.com.

PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function$_2$. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

RINOTT, Y. (1976). On convexity of measures. *Ann*. *Probab*. **4** 1020–1026. MR0428540

ROCKAFELLAR, R. T. (1971). Integrals which are convex functionals. II. *Pacific J*. *Math*. **39** 439–469. MR0310612

ROCKAFELLAR, R. T. (1997). *Convex Analysis*. *Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. MR1451876

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann*. *Statist*. **38** 3751–3781. MR2766867

UHRIN, B. (1984). Some remarks about the convolution of unimodal functions. *Ann. Probab.* **12** 640–645. MR0735860

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

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

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

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

The Annals of Statistics
2016, Vol. 44, No. 3, 1360–1361
DOI: 10.1214/15-AOS1415

CORRECTION NOTE

LIMIT THEOREMS FOR EMPIRICAL PROCESSES OF CLUSTER FUNCTIONALS

Ann. Statist. **38** (2010) 2145–2186

BY HOLGER DREES* AND HOLGER ROOTZÉN[†,‡]

*University of Hamburg**, *Chalmers University*[†] *and Gothenburg University*[‡]

We correct an error in a technical lemma of Drees and Rootzén [*Ann. Statist.* **38** (2010) 2145–2186] and discuss consequences for applications.

REFERENCES

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886

DREES, H., SEGERS, J. and WARCHOŁ, M. (2015). Statistics for tail processes of Markov chains. *Extremes* **18** 369–402. MR3377814

PRATT, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Stat.* **31** 74–77. MR0123673

MSC2010 subject classifications. Primary 60G70; secondary 60F17, 62G32.
Key words and phrases. Absolute regularity, core length, weak dependence.