BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Seven Number Four November 2021 ISSN: 1350-7265

CONTENTS

(continued)

The papers published in Bernoulli are indexed or abstracted in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts-Zentralblatt (STMA-Z)*, and *Zentralblatt für Mathematik* (also avalaible on the *MATH via STN* database and *Compact MATH CD-ROM*). A list of forthcoming papers can be found online at http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Seven Number Four November 2021 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 27 Number 4 November 2021 Pages 2159–2826 ISI/BS

Volume 27 Number 4   November 2021   ISSN 1350-7265

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Aims and Scope
BERNOULLI is the journal of the Bernoulli Society for Mathematical Statistics and Probability, issued four times per year. The journal provides a comprehensive account of important developments in the fields of statistics and probability, offering an international forum for both theoretical and applied work.

Bernoulli Society for Mathematical Statistics and Probability
The Bernoulli Society was founded in 1973. It is an autonomous Association of the International Statistical Institute, ISI. According to its statutes, the object of the Bernoulli Society is the advancement, through international contacts, of the sciences of probability (including the theory of stochastic processes) and mathematical statistics and of their applications to all those aspects of human endeavour which are directed towards the increase of natural knowledge and the welfare of mankind.

Meetings: http://www.bernoulli-society.org/index.php/meetings
The Society holds a World Congress every four years; more frequent meetings, coordinated by the Society's standing committees and often organised in collaboration with other organisations, are the European Meeting of Statisticians, the Conference on Stochastic Processes and their Applications, the CLAPEM meeting (Latin-American Congress on Probability and Mathematical Statistics), the European Young Statisticians Meeting, and various meetings on special topics – in the physical sciences in particular. The Society, as an association of the ISI, also collaborates with other ISI associations in the organization of the biennial ISI World Statistics Congresses (formerly ISI Sessions).

Executive Committee
Detailed information about the members of the Executive Committee can be found on http://www.bernoulli-society.org/who-is-who

---

The papers published in Bernoulli are indexed or abstracted in *Current Index to Statistics, Mathematical Reviews, Statistical Theory and Method Abstracts-Zentralblatt (STMA-Z), Thomson Scientific* and *Zentralblatt für Mathematik* (also available on the *MATH via STN* database and *Compact MATH CD-ROM*). A list of forthcoming papers can be found online at http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

In 2021 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **27**(4), 2021, 2159–2188
https://doi.org/10.3150/20-BEJ1275

On posterior contraction of parameters and interpretability in Bayesian mixture modeling

ARITRA GUHA[1], NHAT HO[2] and XUANLONG NGUYEN[3]

[1]*Department of Statistical Science, Duke University, Durham, NC, USA.*
E-mail: aritra.guha@duke.edu
[2]*Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX, USA.*
E-mail: minhnhat@utexas.edu
[3]*Department of Statistics, University of Michigan, Ann Arbor, MI, USA.*
E-mail: xuanlong@umich.edu

We study posterior contraction behaviors for parameters of interest in the context of Bayesian mixture modeling, where the number of mixing components is unknown while the model itself may or may not be correctly specified. Two representative types of prior specification will be considered: one requires explicitly a prior distribution on the number of mixture components, while the other places a nonparametric prior on the space of mixing distributions. The former is shown to yield an optimal rate of posterior contraction on the model parameters under minimal conditions, while the latter can be utilized to consistently recover the unknown number of mixture components, with the help of a fast probabilistic post-processing procedure. We then turn the study of these Bayesian procedures to the realistic settings of model misspecification. It will be shown that the modeling choice of kernel density functions plays perhaps the most impactful roles in determining the posterior contraction rates in the misspecified situations. Drawing on concrete posterior contraction rates established in this paper we wish to highlight some aspects about the interesting tradeoffs between model expressiveness and interpretability that a statistical modeler must negotiate in the rich world of mixture modeling.

Keywords: Mixture models; Wasserstein distance; Bayesian nonparametrics; Bayesian inference; misspecified models; post-processing algorithm

References

[1] Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969
[2] Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614
[3] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.
[4] Chambaz, A. and Rousseau, J. (2008). Bounds for Bayesian order identification with application to mixtures. *Ann. Statist.* **36** 938–962. MR2396820 https://doi.org/10.1214/009053607000000857
[5] Chen, J.H. (1995). Optimal rate of convergence for finite mixture models. *Ann. Statist.* **23** 221–233. MR1331665 https://doi.org/10.1214/aos/1176324464
[6] Dacunha-Castelle, D. and Gassiat, E. (1997). The estimation of the order of a mixture model. *Bernoulli* **3** 279–299. MR1468306 https://doi.org/10.2307/3318593
[7] Eltoft, T., Kim, T. and Lee, T. (2006). On the multivariate Laplace distribution. *IEEE Signal Process. Lett.* **13** 300–303.
[8] Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510
[9] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

1350-7265

[10] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

[11] Figueiredo, M. and Jain, A.K. (1993). Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**.

[12] Fúquene, J., Steel, M. and Rossell, D. (2019). On choosing mixture components via non-local priors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 809–837. MR4025398 https://doi.org/10.1111/rssb.12333

[13] Gao, F. and van der Vaart, A. (2016). Posterior contraction rates for deconvolution of Dirichlet–Laplace mixtures. *Electron. J. Stat.* **10** 608–627. MR3471990 https://doi.org/10.1214/16-EJS1119

[14] Ghosal, S., Ghosh, J.K. and Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158. MR1701105 https://doi.org/10.1214/aos/1018031105

[15] Ghosal, S., Ghosh, J.K. and van der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[16] Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35** 697–723. MR2336864 https://doi.org/10.1214/009053606000001271

[17] Ghosal, S. and van der Vaart, A.W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329 https://doi.org/10.1214/aos/1013203453

[18] Green, P.J. and Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scand. J. Stat.* **28** 355–375. MR1842255 https://doi.org/10.1111/1467-9469.00242

[19] Guha, A., Ho, N. and Nguyen, X. (2021). Supplement to “On posterior contraction of parameters and interpretability in Bayesian mixture modeling.” https://doi.org/10.3150/20-BEJ1275SUPP

[20] Heinrich, P. and Kahn, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870. MR3851757 https://doi.org/10.1214/17-AOS1641

[21] Hjort, N., Holmes, C., Mueller, P. and Walker, S. (2010). *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press.

[22] Ho, N. and Nguyen, X. (2016). On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electron. J. Stat.* **10** 271–307. MR3466183 https://doi.org/10.1214/16-EJS1105

[23] Ho, N., Nguyen, X. and Ritov, Y. (2020). Robust estimation of mixing measures in finite mixture models. *Bernoulli* **26** 828–857. MR4058353 https://doi.org/10.3150/18-BEJ1087

[24] Ishwaran, H., James, L.F. and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc.* **96** 1316–1332. MR1946579 https://doi.org/10.1198/016214501753382255

[25] Jain, S. and Neal, R.M. (2007). Splitting and merging components of a nonconjugate Dirichlet process mixture model. *Bayesian Anal.* **2** 445–472. MR2342168 https://doi.org/10.1214/07-BA219

[26] Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795. MR3363402 https://doi.org/10.1080/01621459.1995.10476572

[27] Kleijn, B.J.K. and van der Vaart, A.W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.* **34** 837–877. MR2283395 https://doi.org/10.1214/009053606000000029

[28] Kleijn, B.J.K. and van der Vaart, A.W. (2012). The Bernstein–Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675

[29] Leroux, B.G. (1992). Consistent estimation of a mixing distribution. *Ann. Statist.* **20** 1350–1360. MR1186253 https://doi.org/10.1214/aos/1176348772

[30] Lindsay, B. (1995). Mixture models: Theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*. Hayward, CA: IMS.

[31] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. MR0733519 https://doi.org/10.1214/aos/1176346412

[32] MacEachern, S. and Mueller, P. (1998). Estimating mixture of Dirichlet process models. *J. Comput. Graph. Statist.* **7** 223–238.

[33] McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. *Statistics: Textbooks and Monographs* **84**. New York: Dekker. MR0926484

[34] Mengersen, K.L., Robert, C.P. and Titterington, D.M., eds. (2011). *Mixtures: Estimation and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2867716 https://doi.org/10.1002/9781119995678

[35] Miller, J.W. and Harrison, M.T. (2014). Inconsistency of Pitman–Yor process mixtures for the number of components. *J. Mach. Learn. Res.* **15** 3333–3370. MR3277163

[36] Miller, J.W. and Harrison, M.T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

[37] Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

[38] Nguyen, X. (2015). Posterior contraction of the population polytope in finite admixture models. *Bernoulli* **21** 618–646. MR3322333 https://doi.org/10.3150/13-BEJ582

[39] Nobile, A. (1994). Bayesian analysis of finite mixture distributions. Ph.D. thesis, Dept. Statistics, Carnegie Mellon University, Pittsburgh, PA.

[40] Nobile, A. and Fearnside, A.T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Stat. Comput.* **17** 147–162. MR2380643 https://doi.org/10.1007/s11222-006-9014-7

[41] Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

[42] Rodríguez, A., Dunson, D.B. and Gelfand, A.E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103** 1131–1144. MR2528831 https://doi.org/10.1198/016214508000000553

[43] Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 689–710. MR2867454 https://doi.org/10.1111/j.1467-9868.2011.00781.x

[44] 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 https://doi.org/10.1214/14-BA863

[45] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

[46] Shen, W., Tokdar, S.T. and Ghosal, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* **100** 623–640. MR3094441 https://doi.org/10.1093/biomet/ast015

[47] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714. MR1865337 https://doi.org/10.1214/aos/1009210686

[48] Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Ann. Statist.* **28** 40–74. MR1762903 https://doi.org/10.1214/aos/1016120364

[49] Tang, J., Meng, Z., Nguyen, X., Mei, Q. and Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. In *Proceedings of the International Conference on Machine Learning*.

[50] Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

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

[52] Walker, S.G., Lijoi, A. and Prünster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Statist.* **35** 738–746. MR2336866 https://doi.org/10.1214/009053606000001361

[53] Xie, F. and Xu, Y. (2020). Bayesian repulsive Gaussian mixture model. *J. Amer. Statist. Assoc.* **115** 187–203. MR4078456 https://doi.org/10.1080/01621459.2018.1537918

[54] Yurochkin, M., Guha, A. and Nguyen, X. (2017). Conic scan and cover algorithms for nonparametric topic modeling. In *NIPS* **31**.

[55] Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831. MR1056338 https://doi.org/10.1214/aos/1176347627

Bernoulli **27**(4), 2021, 2189–2216
https://doi.org/10.3150/20-BEJ1305

Optimal tests for elliptical symmetry: Specified and unspecified location

SLAĐANA BABIĆ[1,2,*], LAETITIA GELBGRAS[3], MARC HALLIN[4] and CHRISTOPHE LEY[1,†]

[1]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium. E-mail:* [*]*sladana.babic@ugent.be;* [†]*christophe.ley@ugent.be*
[2]*Vlerick Business School, Gent, Belgium*
[3]*Département de Mathématique, Université libre de Bruxelles, Bruxelles, Belgium. E-mail: laetitia.gelbgras@yahoo.fr*
[4]*ECARES and Département de Mathématique, Université libre de Bruxelles, Bruxelles, Belgium. E-mail: mhallin@ulb.ac.be*

Although the assumption of elliptical symmetry is quite common in multivariate analysis and widespread in a number of applications, the problem of testing the null hypothesis of ellipticity so far has not been addressed in a fully satisfactory way. Most papers in the literature indeed are dealing with the null hypothesis of elliptical symmetry with specified location and actually address location rather than non-elliptical alternatives. In this paper, we are proposing new classes of testing procedures, both for specified and unspecified location. The backbone of our construction is Le Cam's asymptotic theory of statistical experiments, and optimality is to be understood locally and asymptotically within the family of generalized skew-elliptical distributions. The tests we are proposing are enjoying all the desirable properties of a "good" test of elliptical symmetry: they have simple asymptotic distributions under the entire null hypothesis of elliptical symmetry with unspecified radial density and shape parameter; they are affine-invariant, computationally fast, intuitively understandable, and not too demanding in terms of moments. While achieving optimality against generalized skew-elliptical alternatives, they remain quite powerful under a much broader class of non-elliptical distributions and significantly outperform the available competitors.

Keywords: Elliptical symmetry; local asymptotic normality; maximin tests; multivariate skewness; semiparametric inference; skew-elliptical densities

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12** 171–178. MR0808153
[2] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 579–602. MR1707862 https://doi.org/10.1111/1467-9868.00194
[3] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391
[4] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83** 715–726. MR1440039 https://doi.org/10.1093/biomet/83.4.715
[5] Babić, S., Gelbgras, L., Hallin, M. and Ley, C. (2021). Supplement to "Optimal tests for elliptical symmetry: Specified and unspecified location." https://doi.org/10.3150/20-BEJ1305SUPP
[6] Baringhaus, L. (1991). Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.* **19** 899–917. MR1105851 https://doi.org/10.1214/aos/1176348127
[7] Beran, R. (1979). Testing for ellipsoidal symmetry of a multivariate density. *Ann. Statist.* **7** 150–162. MR0515690
[8] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79** 99–113. MR1867257 https://doi.org/10.1006/jmva.2000.1960

[9] Cassart, D. (2007). Optimal tests for symmetry. Ph.D. thesis, Univ. libre de Bruxelles, Brussels.
[10] Cassart, D., Hallin, M. and Paindaveine, D. (2008). Optimal detection of Fechner-asymmetry. *J. Statist. Plann. Inference* **138** 2499–2525. MR2414260 https://doi.org/10.1016/j.jspi.2007.10.011
[11] Diks, C. and Tong, H. (1999). A test for symmetries of multivariate probability distributions. *Biometrika* **86** 605–614. MR1723781 https://doi.org/10.1093/biomet/86.3.605
[12] Dominicy, Y., Ilmonen, P. and Veredas, D. (2017). Multivariate Hill estimators. *Int. Stat. Rev.* **85** 108–142. MR3637738 https://doi.org/10.1111/insr.12120
[13] Dyckerhoff, R., Ley, C. and Paindaveine, D. (2015). Depth-based runs tests for bivariate central symmetry. *Ann. Inst. Statist. Math.* **67** 917–941. MR3390173 https://doi.org/10.1007/s10463-014-0480-y
[14] Genton, M.G., ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: CRC Press/CRC. MR2156754 https://doi.org/10.1201/9780203492000
[15] Genton, M.G. and Loperfido, N.M.R. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Statist. Math.* **57** 389–401. MR2160656 https://doi.org/10.1007/BF02507031
[16] Hallin, M. and Ley, C. (2012). Skew-symmetric distributions and Fisher information – a tale of two densities. *Bernoulli* **18** 747–763. MR2948899 https://doi.org/10.3150/12-BEJ346
[17] Hallin, M. and Ley, C. (2014). Skew-symmetric distributions and Fisher information: The double sin of the skew-normal. *Bernoulli* **20** 1432–1453. MR3217449 https://doi.org/10.3150/13-BEJ528
[18] Hallin, M., Oja, H. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann. Statist.* **34** 2757–2789. MR2329466 https://doi.org/10.1214/009053606000000948
[19] Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 https://doi.org/10.1214/aos/1031689019
[20] Hallin, M. and Paindaveine, D. (2002). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8** 787–815. MR1963662
[21] Hallin, M. and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678. MR2153998 https://doi.org/10.1214/009053604000000724
[22] Hallin, M. and Paindaveine, D. (2004). Multivariate signed-rank tests in vector autoregressive order identification. *Statist. Sci.* **19** 697–711. MR2185591 https://doi.org/10.1214/088342304000000602
[23] Hallin, M. and Paindaveine, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163. MR2119768 https://doi.org/10.1016/j.jmva.2004.01.005
[24] Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465 https://doi.org/10.1214/009053606000000731
[25] Hallin, M. and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **36** 1261–1298. MR2418657 https://doi.org/10.1214/07-AOS508
[26] Hallin, M. and Paindaveine, D. (2009). Optimal tests for homogeneity of covariance, scale, and shape. *J. Multivariate Anal.* **100** 422–444. MR2483429 https://doi.org/10.1016/j.jmva.2008.05.010
[27] Hallin, M., Paindaveine, D. and Verdebout, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 https://doi.org/10.1214/10-AOS810
[28] Hallin, M., Paindaveine, D. and Verdebout, T. (2010). Testing for common principal components under heterokurticity. *J. Nonparametr. Stat.* **22** 879–895. MR2738873 https://doi.org/10.1080/10485250903548737
[29] Hallin, M., Paindaveine, D. and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19** 2524–2556. MR3160563 https://doi.org/10.3150/12-BEJ461
[30] Hallin, M., Paindaveine, D. and Verdebout, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083. MR3265681 https://doi.org/10.1080/01621459.2014.880057
[31] Hodgson, D.J., Linton, O. and Vorkink, K. (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *J. Appl. Econometrics* **17** 617–639.
[32] Huffer, F.W. and Park, C. (2007). A test for elliptical symmetry. *J. Multivariate Anal.* **98** 256–281. MR2301752 https://doi.org/10.1016/j.jmva.2005.09.011

[33] Jones, M.C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika* **96** 761–780. MR2564489 https://doi.org/10.1093/biomet/asp053

[34] Koltchinskii, V. and Sakhanenko, L. (2000). Testing for ellipsoidal symmetry of a multivariate distribution. In *High Dimensional Probability*, *II* (*Seattle*, *WA*, 1999). *Progress in Probability* **47** 493–510. Boston, MA: Birkhäuser. MR1857342

[35] Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann*. *Statist*. **15** 112–133. MR0885727 https://doi.org/10.1214/aos/1176350256

[36] Le Cam, L. and Yang, G.L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*. New York, USA: Springer.

[37] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. MR2135927

[38] Ley, C. (2020). When the score function is the identity function – a tale of characterizations of the normal distribution. *Econom*. *Stat*. To appear.

[39] Ley, C. and Paindaveine, D. (2010). On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal*. **101** 1434–1444. MR2609504 https://doi.org/10.1016/j.jmva.2009.10.008

[40] Li, R.-Z., Fang, K.-T. and Zhu, L.-X. (1997). Some QQ probability plots to test spherical and elliptical symmetry. *J*. *Comput*. *Graph*. *Statist*. **6** 435–450.

[41] Liebscher, E. (2005). A semiparametric density estimator based on elliptical distributions. *J*. *Multivariate Anal*. **92** 205–225. MR2102252 https://doi.org/10.1016/j.jmva.2003.09.007

[42] Lombardi, M.J. and Veredas, D. (2009). Indirect estimation of elliptical stable distributions. *Comput*. *Statist*. *Data Anal*. **53** 2309–2324. MR2665830 https://doi.org/10.1016/j.csda.2008.04.035

[43] Manzotti, A., Pérez, F.J. and Quiroz, A.J. (2002). A statistic for testing the null hypothesis of elliptical symmetry. *J*. *Multivariate Anal*. **81** 274–285. MR1906381 https://doi.org/10.1006/jmva.2001.2007

[44] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J*. *Nonparametr. Stat*. **5** 201–213. MR1346895 https://doi.org/10.1080/10485259508832643

[45] Paindaveine, D. (2014). Elliptical symmetry. In *Encyclopedia of Environmetrics*, 2nd ed. (A.H. El-Shaarawi and W. Piegorsch, eds.) 802–807. Chichester, UK: John Wiley & Sons.

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

[47] Rousseeuw, P.J. and Driessen, K.V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41** 212–223.

[48] Sakhanenko, L. (2008). Testing for ellipsoidal symmetry: A comparison study. *Comput*. *Statist*. *Data Anal*. **53** 565–581. MR2649110 https://doi.org/10.1016/j.csda.2008.08.029

[49] Schott, J.R. (2002). Testing for elliptical symmetry in covariance-matrix-based analyses. *Statist*. *Probab*. *Lett*. **60** 395–404. MR1947179 https://doi.org/10.1016/S0167-7152(02)00306-1

[50] Serfling, R.J. (2006). Multivariate symmetry and asymmetry. In *Encyclopedia of Statistical Sciences*, 2nd ed. (S. Kotz, N. Balakrishnan, C.B. Read and B. Vidakovic, eds.) **8** 5338–5345. Wiley Online Library.

[51] Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann*. *Statist*. **15** 234–251. MR0885734 https://doi.org/10.1214/aos/1176350263

[52] Um, Y. and Randles, R.H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist*. *Sinica* **8** 801–812. MR1651509

[53] Vogel, D. and Fried, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. MR2860334 https://doi.org/10.1093/biomet/asr037

[54] Wraith, D. and Forbes, F. (2015). Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering. *Comput*. *Statist*. *Data Anal*. **90** 61–73. MR3354829 https://doi.org/10.1016/j.csda.2015.04.008

[55] Zhu, L.-X. and Neuhaus, G. (2000). Nonparametric Monte Carlo tests for multivariate distributions. *Biometrika* **87** 919–928. MR1813984 https://doi.org/10.1093/biomet/87.4.919

[56] Zhu, L.-X. and Neuhaus, G. (2003). Conditional tests for elliptical symmetry. *J*. *Multivariate Anal*. **84** 284–298. MR1965223 https://doi.org/10.1016/S0047-259X(02)00036-2

Bernoulli **27**(4), 2021, 2217–2240
https://doi.org/10.3150/20-BEJ1306

On weak conditional convergence of bivariate Archimedean and Extreme Value copulas, and consequences to nonparametric estimation

THIMO M. KASPER*, SEBASTIAN FUCHS† and WOLFGANG TRUTSCHNIG‡

Department for Mathematics, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria.
E-mail: *thimo.kasper@sbg.ac.at; †sebastian.fuchs@sbg.ac.at; ‡wolfgang.trutschnig@sbg.ac.at

Looking at bivariate copulas from the perspective of conditional distributions and considering weak convergence of almost all conditional distributions yields the notion of weak conditional convergence. At first glance, this notion of convergence for copulas might seem far too restrictive to be of any practical importance – in fact, given samples of a copula C the corresponding empirical copulas do not converge weakly conditional to C with probability one in general. Within the class of Archimedean copulas and the class of Extreme Value copulas, however, standard pointwise convergence and weak conditional convergence can even be proved to be equivalent. Moreover, it can be shown that every copula C is the weak conditional limit of a sequence of checkerboard copulas. After proving these three main results and pointing out some consequences, we sketch some implications for two recently introduced dependence measures and for the nonparametric estimation of Archimedean and Extreme Value copulas.

Keywords: Archimedean copula; extreme value copula; checkerboard copula; weak convergence; estimation; dependence measure

References

[1] Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B. (1996). On Kendall's process. *J. Multivariate Anal.* **58** 197–229. MR1405589 https://doi.org/10.1006/jmva.1996.0048

[2] Billingsley, P. (2013). *Convergence of Probability Measures*. *Wiley Series in Probability and Statistics*. New York: John Wiley & Sons Inc.

[3] Capéraà, P., Fougères, A.-L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985 https://doi.org/10.1093/biomet/84.3.567

[4] Charpentier, A. and Segers, J. (2008). Convergence of Archimedean copulas. *Statist. Probab. Lett.* **78** 412–419. MR2396414 https://doi.org/10.1016/j.spl.2007.07.014

[5] Chatterjee, S. (2020). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **0** 1–21.

[6] Chatterjee, S. (2020). XICOR: Association measurement through cross rank increments. R. package version 0.3.3.

[7] Cherubini, U., Durante, F. and Mulinacci, S. (2013). *Marshall–Olkin Distributions – Advances in Theory and Applications*. *Springer Proceedings in Mathematics & Statistics* Cham: Springer.

[8] de Haan, L. and Resnick, S.I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290 https://doi.org/10.1007/BF00533086

[9] Dette, H., Siburg, K.F. and Stoimenov, P.A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030 https://doi.org/10.1111/j.1467-9469.2011.00767.x

[10] Durante, F. and Sempi, C. (2016). *Principles of Copula Theory*. Boca Raton, FL: CRC Press. MR3443023

[11] Fernández Sánchez, J. and Trutschnig, W. (2015). Conditioning-based metrics on the space of multivariate copulas and their interrelation with uniform and levelwise convergence and iterated function systems. *J. Theoret. Probab.* **28** 1311–1336. MR3422932 https://doi.org/10.1007/s10959-014-0541-4

[12] Fernández Sánchez, J. and Trutschnig, W. (2015). Singularity aspects of Archimedean copulas. *J. Math. Anal. Appl.* **432** 103–113. MR3371224 https://doi.org/10.1016/j.jmaa.2015.06.036

1350-7265

[13] Genest, C., Nešlehová, J. and Ziegel, J. (2011). Inference in multivariate Archimedean copula models. *TEST* **20** 223–256. MR2834039 https://doi.org/10.1007/s11749-011-0250-6
[14] Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* **88** 1034–1043. MR1242947
[15] Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453 https://doi.org/10.1214/08-AOS672
[16] Ghoudi, K., Khoudraji, A. and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.* **26** 187–197. MR1624413 https://doi.org/10.2307/3315683
[17] Griessenberger, F., Junker, R.R. and Trutschnig, W. (2020). qad: Quantification of asymmetric dependence. R. package version 0.1.2.
[18] Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications. Lect. Notes Stat. Proc.* **198** 127–145. Heidelberg: Springer. MR3051266 https://doi.org/10.1007/978-3-642-12465-5_6
[19] Gudendorf, G. and Segers, J. (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *J. Multivariate Anal.* **102** 37–47. MR2729418 https://doi.org/10.1016/j.jmva.2010.07.011
[20] Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2020). copula: Multivariate dependence with copulas. R. package version 0.999-20.
[21] Junker, R.R., Griessenberger, F. and Trutschnig, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.* **153** 107058, 22. MR4141460 https://doi.org/10.1016/j.csda.2020.107058
[22] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8
[23] Kannan, R. and Krueger, C.K. (2012). *Advanced Analysis on the Real Line*. *Universitext*. New York: Springer. MR1390758
[24] Klement, E.P., Mesiar, R. and Pap, E. (2000). *Triangular Norms*. *Trends in Logic—Studia Logica Library* **8**. Dordrecht: Kluwer Academic. MR1790096 https://doi.org/10.1007/978-94-015-9540-7
[25] Klenke, A. (2008). *Wahrscheinlichkeitstheorie*. *Lehrbuch Masterclass Series*. Berlin: Springer.
[26] Lancaster, H.O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Stat.* **34** 1315–1321. MR0154376 https://doi.org/10.1214/aoms/1177703867
[27] Li, X., Mikusiński, P. and Taylor, M.D. (1998). Strong approximation of copulas. *J. Math. Anal. Appl.* **225** 608–623. MR1644300 https://doi.org/10.1006/jmaa.1998.6056
[28] Marshall, A.W. (1970). Discussion of Barlow and van Zwet's papers. In *Nonparametric Techniques in Statistical Inference* 175–176. London: Cambridge Univ. Press.
[29] Mikusiński, P. and Taylor, M.D. (2010). Some approximations of n-copulas. *Metrika* **72** 385–414. MR2746583 https://doi.org/10.1007/s00184-009-0259-y
[30] Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2197664 https://doi.org/10.1007/0-387-28678-0
[31] Pickands, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the* 43*rd Session of the International Statistical Institute, Vol.* 2 (*Buenos Aires*, 1981) **49** 859–878. MR0820979
[32] Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics* **8**. Cambridge: Cambridge Univ. Press. MR1873379
[33] Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Landmarks in Mathematics*. Princeton, NJ: Princeton Univ. Press. MR1451876
[34] Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill. MR0924157
[35] Sempi, C. (2004). Convergence of copulas: Critical remarks. *Rad. Mat.* **12** 241–249. MR2065562
[36] Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. MR2825218 https://doi.org/10.1016/j.jmaa.2011.06.013
[37] Trutschnig, W. (2012). Some results on the convergence of (quasi-) copulas. *Fuzzy Sets and Systems* **191** 113–121. MR2874827 https://doi.org/10.1016/j.fss.2011.06.013
[38] Trutschnig, W., Schreyer, M. and Fernández-Sánchez, J. (2016). Mass distributions of two-dimensional extreme-value copulas and related results. *Extremes* **19** 405–427. MR3535960 https://doi.org/10.1007/s10687-016-0249-1

Bernoulli **27**(4), 2021, 2241–2266
https://doi.org/10.3150/20-BEJ1307

The Goldenshluger–Lepski method for constrained least-squares estimators over RKHSs

STEPHEN PAGE[1] and STEFFEN GRÜNEWÄLDER[2]

[1]*STOR-i, Lancaster University, Lancaster, LA1 4YF, United Kingdom.*
E-mail: s.page@lancaster.ac.uk
[2]*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, United Kingdom.*
E-mail: s.grunewalder@lancaster.ac.uk

We study an adaptive estimation procedure called the Goldenshluger–Lepski method in the context of reproducing kernel Hilbert space (RKHS) regression. Adaptive estimation provides a way of selecting tuning parameters for statistical estimators using only the available data. This allows us to perform estimation without making strong assumptions about the estimand. In contrast to procedures such as training and validation, the Goldenshluger–Lepski method uses all of the data to produce non-adaptive estimators for a range of values of the tuning parameters. An adaptive estimator is selected by performing pairwise comparisons between these non-adaptive estimators. Applying the Goldenshluger–Lepski method is non-trivial as it requires a simultaneous high-probability bound on all of the pairwise comparisons. In the RKHS regression context, we choose our non-adaptive estimators to be clipped least-squares estimators constrained to lie in a ball in an RKHS. Applying the Goldenshluger–Lepski method in this context is made more complicated by the fact that we cannot use the L^2 norm for performing the pairwise comparisons as it is unknown. We use the method to address two regression problems. In the first problem the RKHS is fixed, while in the second problem we adapt over a collection of RKHSs.

Keywords: Adaptive estimation; Goldenshluger–Lepski method; RKHS regression

References

[1] Bach, F.R., Lanckriet, G.R.G. and Jordan, M.I. (2004). Multiple kernel learning, conic duality, and the SMO algorithm. In *Proceedings of the Twenty-First International Conference on Machine Learning* 6.
[2] Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Stat.* **6** 127–146. MR1918295 https://doi.org/10.1051/ps:2002007
[3] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210
[4] Baumeister, J. (1987). *Stable Solution of Inverse Problems*. *Advanced Lectures in Mathematics*. Braunschweig: Friedr. Vieweg & Sohn. MR0889048 https://doi.org/10.1007/978-3-322-83967-1
[5] Bergh, J. and Löfström, J. (1976). *Interpolation Spaces*. *An Introduction*. Berlin: Springer. MR0482275
[6] Birgé, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **36** 113–133. Beachwood, OH: IMS. MR1836557 https://doi.org/10.1214/lnms/1215090065
[7] Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam* 55–87. New York: Springer. MR1462939
[8] Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031
[9] Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138** 33–73. MR2288064 https://doi.org/10.1007/s00440-006-0011-8

1350-7265

[10] Blanchard, G., Mathé, P. and Mücke, N. (2019). Lepskii principle in supervised learning. Preprint. Available at arXiv:1905.10764.

[11] Caponnetto, A. and Yao, Y. (2010). Cross-validation based adaptation for regularization operators in learning theory. *Anal. Appl.* (*Singap.*) **8** 161–183. MR2646909 https://doi.org/10.1142/S0219530510001564

[12] Chapelle, O., Vapnik, V., Bousquet, O. and Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Mach. Learn.* **46**.

[13] De Vito, E., Pereverzyev, S. and Rosasco, L. (2010). Adaptive kernel methods using the balancing principle. *Found. Comput. Math.* **10** 455–479. MR2657949 https://doi.org/10.1007/s10208-010-9064-2

[14] Eberts, M. and Steinwart, I. (2013). Optimal regression rates for SVMs using Gaussian kernels. *Electron. J. Stat.* **7** 1–42. MR3020412 https://doi.org/10.1214/12-EJS760

[15] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. New York: Cambridge Univ. Press. MR3588285 https://doi.org/10.1017/CBO9781107337862

[16] Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590 https://doi.org/10.3150/08-BEJ144

[17] Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. MR2449122 https://doi.org/10.1007/s00440-007-0119-5

[18] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 https://doi.org/10.1214/11-AOS883

[19] Goldenshluger, A. and Pereverzev, S.V. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli* **9** 783–807. MR2047686 https://doi.org/10.3150/bj/1066418878

[20] Goldenshluger, A.V. and Lepski, O.V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl.* **57** 209–226. MR3201652 https://doi.org/10.1137/S0040585X97985923

[21] Gönen, M. and Alpaydın, E. (2011). Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12** 2211–2268. MR2825425 https://doi.org/10.1016/j.patcog.2012.09.002

[22] Hofmann, B. (2005). Approximate source conditions in regularization and an application to multiplication operators. In *Proceedings*: 10*th International Conference "Mathematical Modelling and Analysis* 2005*" and* 2*nd International Conference "Computational Methods in Applied Mathematics"* 29–34. Vilnius: Technika. MR2194416

[23] Kreĭn, S.G. and Petunin, J.I. (1966). Scales of Banach spaces. *Uspekhi Mat. Nauk* **21** 89–168. MR0193499

[24] Lepski, O. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

[25] Lepskiĭ, O.V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167 https://doi.org/10.1137/1136085

[26] Lepskiĭ, O.V. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatn. Primen.* **37** 468–481. MR1214353 https://doi.org/10.1137/1137095

[27] Lu, S., Mathé, P. and Pereverzev, S.V. (2020). Balancing principle in supervised learning for a general regularization scheme. *Appl. Comput. Harmon. Anal.* **48** 123–148. MR4016987 https://doi.org/10.1016/j.acha.2018.03.001

[28] Lu, S., Mathé, P. and Pereverzyev, S. Jr. (2019). Analysis of regularized Nyström subsampling for regression functions of low smoothness. *Anal. Appl.* (*Singap.*) **17** 931–946. MR4009938 https://doi.org/10.1142/S0219530519500039

[29] Mair, B.A. and Ruymgaart, F.H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56** 1424–1444. MR1409127 https://doi.org/10.1137/S0036139994264476

[30] Mathé, P. and Pereverzev, S.V. (2001). Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.* **38** 1999–2021. MR1856240 https://doi.org/10.1137/S003614299936175X

[31] Oneto, L., Ridella, S. and Anguita, D. (2016). Tikhonov, Ivanov and Morozov regularization for support vector machine learning. *Mach. Learn.* **103** 103–136. MR3472725 https://doi.org/10.1007/s10994-015-5540-x

[32] Page, S. and Grünewälder, S. (2019). Ivanov-regularised least-squares estimators over large RKHSs and their interpolation spaces. *J. Mach. Learn. Res.* **20** 120. MR4002874

[33] Page, S. and Grünewälder, S. (2021). Supplement to "The Goldenshluger–Lepski method for constrained least-squares estimators over RKHSs." https://doi.org/10.3150/20-BEJ1307SUPP

[34] Pereverzyev, S. and Tkachenko, P. (2017). Regularization by the linear functional strategy with multiple kernels. *Front. Appl. Math. Stat.* **3**.

[35] Pricop-Jeckstadt, M. (2019). Nonlinear Tikhonov regularization in Hilbert scales with balancing principle tuning parameter in statistical inverse problems. *Inverse Probl. Sci. Eng.* **27** 205–236. MR3883351 https://doi.org/10.1080/17415977.2018.1454918

[36] Smale, S. and Zhou, D.-X. (2003). Estimating the approximation error in learning theory. *Anal. Appl.* (*Singap.*) **1** 17–41. MR1959283 https://doi.org/10.1142/S0219530503000089

[37] Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. New York: Springer. MR2450103

[38] Steinwart, I., Hush, D.R. and Scovel, C. (2009). Optimal rates for regularized least squares regression. In *The* 22*nd Conference on Learning Theory*.

[39] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[40] Williams, D. (1991). *Probability with Martingales*. Cambridge: Cambridge Univ. Press. MR1155402 https://doi.org/10.1017/CBO9780511813658

Bernoulli **27**(4), 2021, 2267–2299
https://doi.org/10.3150/20-BEJ1308

The linear conditional expectation in Hilbert space

ILJA KLEBANOV[1,*], BJÖRN SPRUNGK[2] and T.J. SULLIVAN[1,3,†]

[1]*Zuse Institute Berlin, Takustraße 7, 14195 Berlin, Germany.*
E-mail: [*]*klebanov@zib.de;* [†]*sullivan@zib.de*
[2]*Technische Universität Bergakademie Freiberg, 09596 Freiberg, Germany.*
E-mail: bjoern.sprungk@math.tu-freiberg.de
[3]*Mathematics Institute and School of Engineering, The University of Warwick, Coventry, CV4 7AL, United Kingdom.*
E-mail: t.j.sullivan@warwick.ac.uk

The *linear conditional expectation* (LCE) provides a best linear (or rather, affine) estimate of the conditional expectation and hence plays an important rôle in approximate Bayesian inference, especially the *Bayes linear* approach. This article establishes the analytical properties of the LCE in an infinite-dimensional Hilbert space context. In addition, working in the space of affine Hilbert–Schmidt operators, we establish a regularisation procedure for this LCE. As an important application, we obtain a simple alternative derivation and intuitive justification of the *conditional mean embedding* formula, a concept widely used in machine learning to perform the conditioning of random variables by embedding them into reproducing kernel Hilbert spaces.

Keywords: Bayes linear analysis; conditional mean embedding; reproducing kernel Hilbert space; linear conditional expectation

References

[1] Arias, M.L., Corach, G. and Gonzalez, M.C. (2008). Generalized inverses and Douglas equations. *Proc. Amer. Math. Soc.* **136** 3177–3183. MR2407082 https://doi.org/10.1090/S0002-9939-08-09298-8

[2] Aubin, J.-P. (2000). *Applied Functional Analysis*, 2nd ed. *Pure and Applied Mathematics* (*New York*). New York: Wiley Interscience. MR1782330 https://doi.org/10.1002/9781118032725

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

[4] Brémaud, P. (2017). *Discrete Probability Models and Methods*: *Probability on Graphs and Trees*, *Markov Chains and Random Fields*, *Entropy and Coding*. *Probability Theory and Stochastic Modelling* **78**. Cham: Springer. MR3616988 https://doi.org/10.1007/978-3-319-43476-6

[5] Chatterji, S.D. (1960). Martingales of Banach-valued random variables. *Bull. Amer. Math. Soc.* **66** 395–398. MR0119242 https://doi.org/10.1090/S0002-9904-1960-10471-5

[6] Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics*: *Modeling Spatial Uncertainty*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2850475 https://doi.org/10.1002/9781118136188

[7] Corach, G., Maestripieri, A. and Stojanoff, D. (2001). Oblique projections and Schur complements. *Acta Sci. Math.* (*Szeged*) **67** 337–356. MR1830147

[8] Diestel, J. and Uhl, J.J. Jr. (1977). *Vector Measures*. Providence, RI: Amer. Math. Soc. MR0453964

[9] Douglas, R.G. (1966). On majorization, factorization, and range inclusion of operators on Hilbert space. *Proc. Amer. Math. Soc.* **17** 413–415. MR0203464 https://doi.org/10.2307/2035178

[10] Dudley, R.M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. MR1932358 https://doi.org/10.1017/CBO9780511755347

[11] Engl, H.W., Hanke, M. and Neubauer, A. (1996). *Regularization of Inverse Problems*. *Mathematics and Its Applications* **375**. Dordrecht: Kluwer Academic. MR1408680

[12] Ernst, O.G., Sprungk, B. and Starkloff, H.-J. (2015). Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantificat.* **3** 823–851. MR3400030 https://doi.org/10.1137/140981319
[13] Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. Berlin: Springer. MR2555209 https://doi.org/10.1007/978-3-642-03711-5
[14] Fillmore, P.A. and Williams, J.P. (1971). On operator ranges. *Adv. Math.* **7** 254–281. MR0293441 https://doi.org/10.1016/S0001-8708(71)80006-3
[15] Folland, G.B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics* (*New York*). New York: Wiley. MR1681462
[16] Fukumizu, K., Song, L. and Gretton, A. (2013). Kernel Bayes' rule: Bayesian inference with positive definite kernels. *J. Mach. Learn. Res.* **14** 3753–3783. MR3159407
[17] Goldstein, M. (1999). Bayes linear analysis. In *Encyclopaedia of Statistical Sciences* (S. Kotz, B.C. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson, eds.) 29–34. Chichester: Wiley.
[18] Goldstein, M. and Wooff, D. (2007). *Bayes Linear Statistics: Theory and Methods*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2335584 https://doi.org/10.1002/9780470065662
[19] Hairer, M., Stuart, A.M., Voss, J. and Wiberg, P. (2005). Analysis of SPDEs arising in path sampling. I. The Gaussian case. *Commun. Math. Sci.* **3** 587–603. MR2188686
[20] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2722294 https://doi.org/10.1007/978-0-387-84858-7
[21] Kallenberg, O. (2006). *Foundations of Modern Probability*. *Probability and Its Applications* (*New York*). New York: Springer. MR1464694
[22] Klebanov, I., Schuster, I. and Sullivan, T.J. (2020). A rigorous theory of conditional mean embeddings. *SIAM J. Math. Data Sci.* **2** 583–606. MR4121886 https://doi.org/10.1137/19M1305069
[23] Klenke, A. (2013). *Wahrscheinlichkeitstheorie*, 3rd ed. Berlin: Springer. https://doi.org/10.1007/978-3-642-36018-3
[24] Klus, S., Husic, B.E., Mollenhauer, M. and Noé, F. (2019). Kernel methods for detecting coherent structures in dynamical data. *Chaos* **29** 123112, 15. MR4041090 https://doi.org/10.1063/1.5100267
[25] Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C. and Noé, F. (2018). Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.* **28** 985–1010. MR3800253 https://doi.org/10.1007/s00332-017-9437-7
[26] Klus, S., Schuster, I. and Muandet, K. (2020). Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces. *J. Nonlinear Sci.* **30** 283–315. MR4054854 https://doi.org/10.1007/s00332-019-09574-z
[27] Mandelbaum, A. (1984). Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrsch. Verw. Gebiete* **65** 385–397. MR0731228 https://doi.org/10.1007/BF00533743
[28] Meise, R. and Vogt, D. (1997). *Introduction to Functional Analysis*. *Oxford Graduate Texts in Mathematics* **2**. New York: The Clarendon Press, Oxford Univ. Press. MR1483073
[29] Owhadi, H. and Scovel, C. (2018). Conditioning Gaussian measure on Hilbert space. *J. Math. Stat. Anal.* **1** 1–15.
[30] Sazonov, V. (1958). On characteristic functionals. *Teor. Veroyatn. Primen.* **3** 201–205. MR0098423
[31] Schillings, C. and Stuart, A.M. (2017). Analysis of the ensemble Kalman filter for inverse problems. *SIAM J. Numer. Anal.* **55** 1264–1290. MR3654885 https://doi.org/10.1137/16M105959X
[32] Schwantes, C.R. and Pande, V.S. (2015). Modeling molecular kinetics with tICA and the kernel trick. *J. Chem. Theory Comput.* **11** 600–608. https://doi.org/10.1021/ct5007357
[33] Song, L., Huang, J., Smola, A. and Fukumizu, K. (2009). Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 961–968. https://doi.org/10.1145/1553374.1553497
[34] Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. *Springer Series in Statistics*. New York: Springer. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6
[35] Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. *Information Science and Statistics*. New York: Springer. MR2450103
[36] Tarieladze, V. and Vakhania, N. (2007). Disintegration of Gaussian measures and average-case optimal algorithms. *J. Complexity* **23** 851–866. MR2371996 https://doi.org/10.1016/j.jco.2007.04.005

Bernoulli **27**(4), 2021, 2300–2336
https://doi.org/10.3150/20-BEJ1309

Universal sieve-based strategies for efficient estimation using machine learning tools

HONGXIANG QIU[1,*], ALEX LUEDTKE[2] and MARCO CARONE[1,†]

[1] *Department of Biostatistics, University of Washington, Seattle, WA, USA.*
E-mail: [*] *qiuhx@uw.edu;* [†] *mcarone@uw.edu*
[2] *Department of Statistics, University of Washington, Seattle, WA, USA.*
E-mail: aluedtke@uw.edu

Suppose that we wish to estimate a finite-dimensional summary of one or more function-valued features of an underlying data-generating mechanism under a nonparametric model. One approach to estimation is by plugging in flexible estimates of these features. Unfortunately, in general, such estimators may not be asymptotically efficient, which often makes these estimators difficult to use as a basis for inference. Though there are several existing methods to construct asymptotically efficient plug-in estimators, each such method either can only be derived using knowledge of efficiency theory or is only valid under stringent smoothness assumptions. Among existing methods, sieve estimators stand out as particularly convenient because efficiency theory is not required in their construction, their tuning parameters can be selected data adaptively, and they are universal in the sense that the same fits lead to efficient plug-in estimators for a rich class of estimands. Inspired by these desirable properties, we propose two novel universal approaches for estimating function-valued features that can be analyzed using sieve estimation theory. Compared to traditional sieve estimators, these approaches are valid under more general conditions on the smoothness of the function-valued features by utilizing flexible estimates that can be obtained, for example, using machine learning.

Keywords: Nonparametric inference; asymptotic efficiency; sieve estimation

References

[1] Benkeser, D. and van Der Laan, M. (2016). The highly adaptive lasso estimator. In *Data Science and Advanced Analytics* (*DSAA*), 2016 *IEEE International Conference on* 689–696. IEEE.

[2] Bickel, P.J., Götze, F. and van Zwet, W.R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist*. *Sinica* **7** 1–31. MR1441142

[3] Bickel, P.J. and Ritov, Y. (2003). Nonparametric estimators which can be "plugged-in". *Ann*. *Statist*. **31** 1033–1053. MR2001641 https://doi.org/10.1214/aos/1059655904

[4] Chen, X. (2007). Chapter 76 Large sample sieve estimation of semi-nonparametric models. *Handb*. *Econom*. **6** 5549–5632.

[5] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C. and Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *Am*. *Econ*. *Rev*. **107** 261–265. https://doi.org/10.1257/aer.p20171038

[6] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom*. *J*. **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[7] Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Ann*. *Statist*. **29** 1189–1232. MR1873328 https://doi.org/10.1214/aos/1013203451

[8] Friedman, J.H. (2002). Stochastic gradient boosting. *Comput*. *Statist*. *Data Anal*. **38** 367–378. MR1884869 https://doi.org/10.1016/S0167-9473(01)00065-2

[9] Gill, R.D., van der Laan, M.J. and Wellner, J.A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **31** 545–597. MR1338452

[10] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. New York: Springer. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7
[11] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. MR1134488
[12] Levy, J., van der Laan, M., Hubbard, A. and Pirracchio, R. (2018). A fundamental measure of treatment effect heterogeneity. arxiv preprint arXiv:1811.03745v3.
[13] Mallat, S. (2009). *A Wavelet Tour of Signal Processing*: *The Sparse Way*, 3rd ed. Amsterdam: Elsevier/Academic Press. With contributions from Gabriel Peyré. MR2479996
[14] Marron, J.S. (1994). Visual understanding of higher-order kernels. *J. Comput. Graph. Statist.* **3** 447–458. https://doi.org/10.1080/10618600.1994.10474657
[15] Mason, L., Baxter, J., Bartlett, P. and Frean, M. (1999). Boosting algorithms as gradient descent in function space. Technical Report.
[16] Mason, L., Baxter, J., Bartlett, P.L. and Frean, M. (2000). Boosting algorithms as gradient descent. Technical Report.
[17] Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.
[18] Newey, W., Hsieh, F. and Robins, J. (1998). Undersmoothing and bias corrected functional estimation. Working Papers.
[19] Newey, W.K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. MR1457700 https://doi.org/10.1016/S0304-4076(97)00011-0
[20] Newey, W.K., Hsieh, F. and Robins, J.M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* **72** 947–962. MR2051442 https://doi.org/10.1111/j.1468-0262.2004.00518.x
[21] Owen, A.B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*. *Ser. Biostat.* **2** 49–74. Hackensack, NJ: World Sci. Publ. MR2271076
[22] Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **13**. New York–Berlin: Springer. With the assistance of W. Wefelmeyer. MR0675954
[23] Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701. https://doi.org/10.1037/h0037350
[24] Shen, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591. MR1604416 https://doi.org/10.1214/aos/1030741085
[25] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
[26] van der Laan, M. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. *Int. J. Biostat.* **13** 20150097, 35. MR3724476 https://doi.org/10.1515/ijb-2015-0097
[27] van der Laan, M.J., Benkeser, D. and Cai, W. (2019). Efficient estimation of pathwise differentiable target parameters with the undersmoothed highly adaptive lasso. arXiv preprint arXiv:1908.05607v1.
[28] van der Laan, M.J. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples. U.C. Berkeley Division of Biostatistics Working Paper.
[29] van der Laan, M.J. and Robins, J.M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. *Springer Series in Statistics*. New York: Springer. https://doi.org/10.1007/978-0-387-21700-0
[30] van der Laan, M.J. and Rose, S. (2018). Targeted learning in data science.
[31] van der Laan, M.J. and Rubin, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500 https://doi.org/10.2202/1557-4679.1043
[32] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[33] Williamson, B.D., Gilbert, P.B., Carone, M. and Simon, N. (2020). Nonparametric variable importance assessment using machine learning techniques. *Biometrics*. https://doi.org/10.1111/biom.13392.
[34] Yang, S. and Ding, P. (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika* **105** 487–493. MR3804418 https://doi.org/10.1093/biomet/asy008
[35] Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Bernoulli **27**(4), 2021, 2337–2352
https://doi.org/10.3150/20-BEJ1310

Partial generalized four moment theorem revisited[1]

DANDAN JIANG[1] and ZHIDONG BAI[2]

[1]*School of Mathematics and Statistics, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an 710049, China.*
E-mail: jiangdd@xjtu.edu.cn
[2]*KLASMOE and School of Mathematics and Statistics, Northeast Normal University, No. 5268 People's Street, Changchun 130024, China.*
E-mail: baizd@nenu.edu.cn

This is a complementary proof of partial generalized 4 moment theorem (PG4MT) mentioned and described in "Generalized Four Moment Theorem (G4MT) and its Application to CLT for Spiked Eigenvalues of High-dimensional Covariance Matrices". Since the G4MT proposed in that paper requires both the matrices **X** and **Y** satisfying the assumption $\max_{t,s} |u_{ts}|^2 \mathrm{E}\{|x_{11}|^4 I(|x_{11}| < \sqrt{n}) - \mu\} \to 0$ with the same μ which maybe restrictive in real applications, we proposed a new G4MT, called PG4MT, without proof. After the manuscript posed in ArXiv, the authors received high interests in the proof of PG4MT through private communications and find the PG4MT more general than G4MT, it is necessary to give a detailed proof of it. Moreover, it is found that the PG4MT derives a CLT of spiked eigenvalues of sample covariance matrices which covers the work in Bai and Yao (*J. Multivariate Anal.* **106** (2012) 167–177) as a special case.

Keywords: High-dimensional covariance matrix; random matrix theory; spiked model; partial generalized four moment theorem; central limit theorem

References

[1] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273
[2] Bai, Z. and Yao, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118
[3] Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009
[4] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233
[5] Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003
[6] Cai, T.T., Han, X. and Pan, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Statist.* **48** 1255–1280. MR4124322 https://doi.org/10.1214/18-AOS1798
[7] Hoyle, D.C. and Rattray, M. (2004). Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. *Phys. Rev. E* **69** 026124.
[8] Jiang, D. and Bai, Z. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli* **27** 274–294. MR4177370 https://doi.org/10.3150/20-BEJ1237
[9] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[10] Onatski, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070 https://doi.org/10.3982/ECTA6964
[11] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist*. *Sinica* **17** 1617–1642. MR2399865
[12] Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann*. *Statist*. **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

Bernoulli **27**(4), 2021, 2353–2361
https://doi.org/10.3150/20-BEJ1311

Finite-energy infinite clusters without anchored expansion

GÁBOR PETE[1] and ÁDÁM TIMÁR[2]

[1]*Alfréd Rényi Institute of Mathematics, Budapest, Hungary.*
E-mail: robagetep@gmail.com
[2]*Budapest University of Technology and Economics, Budapest, Hungary.*
E-mail: madaramit@gmail.com

Hermon and Hutchcroft have recently proved the long-standing conjecture that in Bernoulli(p) bond percolation on any nonamenable transitive graph G, at any $p > p_c(G)$, the probability that the cluster of the origin is finite but has a large volume n decays exponentially in n. A corollary is that all infinite clusters have anchored expansion almost surely. They have asked if these results could hold more generally, for any finite energy ergodic invariant percolation. We give a counterexample, an invariant percolation on the 4-regular tree.

Keywords: Anchored expansion; invariant percolation

References

[1] Abächerli, A. and Černý, J. (2020). Level-set percolation of the Gaussian free field on regular graphs I: Regular trees. *Electron. J. Probab.* **25** 65. MR4115734 https://doi.org/10.1214/20-ejp468
[2] Aizenman, M., Barsky, D.J. and Fernández, R. (1987). The phase transition in a general class of Ising-type models is sharp. *J. Stat. Phys.* **47** 343–374. MR0894398 https://doi.org/10.1007/BF01007515
[3] Aldous, D. and Lyons, R. (2007). Processes on unimodular random networks. *Electron. J. Probab.* **12** 1454–1508. MR2354165 https://doi.org/10.1214/EJP.v12-463
[4] Angel, O. and Hutchcroft, T. (2018). Counterexamples for percolation on unimodular random graphs. In *Unimodularity in Randomly Generated Graphs. Contemp. Math.* **719** 11–28. Providence, RI: Amer. Math. Soc. MR3880008 https://doi.org/10.1090/conm/719/14465
[5] Athreya, K.B. and Ney, P. (1970). The local limit theorem and some related aspects of super-critical branching processes. *Trans. Amer. Math. Soc.* **152** 233–251. MR0268971 https://doi.org/10.2307/1995649
[6] Backhausz, Á. and Virág, B. (2017). Spectral measures of factor of i.i.d. processes on vertex-transitive graphs. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2260–2278. MR3729654 https://doi.org/10.1214/16-AIHP790
[7] Bandyopadhyay, A., Steif, J. and Timár, Á. (2010). On the cluster size distribution for percolation on some general graphs. *Rev. Mat. Iberoam.* **26** 529–550. MR2677006 https://doi.org/10.4171/RMI/608
[8] Benjamini, I. and Curien, N. (2012). Ergodic theory on stationary random graphs. *Electron. J. Probab.* **17** 93. MR2994841 https://doi.org/10.1214/EJP.v17-2401
[9] Benjamini, I., Lyons, R. and Schramm, O. (1999). Percolation perturbations in potential theory and random walks. In *Random Walks and Discrete Potential Theory* (*Cortona*, 1997). *Sympos. Math.*, *XXXIX* 56–84. Cambridge: Cambridge Univ. Press. MR1802426
[10] Benjamini, I., Lyons, R. and Schramm, O. (2015). Unimodular random trees. *Ergodic Theory Dynam. Systems* **35** 359–373. MR3316916 https://doi.org/10.1017/etds.2013.56
[11] Benjamini, I. and Schramm, O. (1996). Percolation beyond $\mathbb{Z}^d$, many questions and a few answers. *Electron. Commun. Probab.* **1** 71–82. MR1423907 https://doi.org/10.1214/ECP.v1-978
[12] Beringer, D., Pete, G. and Timár, Á. (2017). On percolation critical probabilities and unimodular random graphs. *Electron. J. Probab.* **22** 106. MR3742403 https://doi.org/10.1214/17-EJP124

[13] Carmesin, J., Federici, B. and Georgakopoulos, A. (2017). A Liouville hyperbolic souvlaki. Appendix: A unimodular Liouville hyperbolic souvlaki. *Electron*. *J*. *Probab*. **22** 36. MR3646062 https://doi.org/10.1214/17-EJP44
[14] Chen, D. and Peres, Y. (2004). Anchored expansion, percolation and speed. *Ann*. *Probab*. **32** 2978–2995. MR2094436 https://doi.org/10.1214/009117904000000586
[15] Duminil-Copin, H., Goswami, S., Rodriguez, P.-F. and Severo, F. Equality of critical parameters for percolation of Gaussian free field level-sets. Available at arXiv:2002.07735 [math.PR].
[16] Duminil-Copin, H., Raoufi, A. and Tassion, V. (2019). Sharp phase transition for the random-cluster and Potts models via decision trees. *Ann*. *of Math*. (2) **189** 75–99. MR3898174 https://doi.org/10.4007/annals.2019.189.1.2
[17] Duminil-Copin, H. and Tassion, V. (2016). A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model. *Comm*. *Math*. *Phys*. **343** 725–745. MR3477351 https://doi.org/10.1007/s00220-015-2480-z
[18] Hermon, J. and Hutchcroft, T. Supercritical percolation on nonamenable graphs: Isoperimetry, analyticity, and exponential decay of the cluster size distribution. *Invent*. *Math*. To appear. Available at arXiv:1904.10448v3 [math.PR].
[19] Hutchcroft, T. Continuity of the Ising phase transition on nonamenable groups. Available at arXiv:2007.15625 [math.PR].
[20] Kesten, H. and Zhang, Y. (1990). The probability of a large finite cluster in supercritical Bernoulli percolation. *Ann*. *Probab*. **18** 537–555. MR1055419
[21] Lyons, R. (2000). Phase transitions on nonamenable graphs. *J*. *Math*. *Phys*. **41** 1099–1126. MR1757952 https://doi.org/10.1063/1.533179
[22] Lyons, R. (2017). Factors of IID on trees. *Combin*. *Probab*. *Comput*. **26** 285–300. MR3603969 https://doi.org/10.1017/S096354831600033X
[23] Lyons, R. and Peres, Y. (2016). *Probability on Trees and Networks*. *Cambridge Series in Statistical and Probabilistic Mathematics* **42**. New York: Cambridge Univ. Press. MR3616205 https://doi.org/10.1017/9781316672815
[24] Pete, G. Probability and Geometry on Groups. http://www.math.bme.hu/~Gabor/PGG.pdf.
[25] Pete, G. (2008). A note on percolation on $\mathbb{Z}^d$: Isoperimetric profile via exponential cluster repulsion. *Electron*. *Commun*. *Probab*. **13** 377–392. MR2415145 https://doi.org/10.1214/ECP.v13-1390
[26] Teixeira, A. and Tykesson, J. (2013). Random interlacements and amenability. *Ann*. *Appl*. *Probab*. **23** 923–956. MR3076674 https://doi.org/10.1214/12-aap860
[27] Thomassen, C. (1992). Isoperimetric inequalities and transient random walks on graphs. *Ann*. *Probab*. **20** 1592–1600. MR1175279
[28] Timár, Á. A nonamenable "factor" of a Euclidean space. *Ann*. *Probab*. To appear.
[29] Virág, B. (2000). Anchored expansion and random walk. *Geom*. *Funct*. *Anal*. **10** 1588–1605. MR1810755 https://doi.org/10.1007/PL00001663

Bernoulli **27**(4), 2021, 2362–2388
https://doi.org/10.3150/20-BEJ1312

Nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes

NATHALIE KRELL[1] and ÉMELINE SCHMISSER[2]

[1] *Univ Rennes, CNRS, IRMAR, UMR 6625, F-35000 Rennes, France.*
E-mail: nathalie.krell@univ-rennes1.fr
[2] *Laboratoire Paul Painelvé, Université de Lille, CNRS UMR 8524, Bâtiment M2, Cité Scientifique, 59655 Villeneuve d'Ascq, France.*
E-mail: emeline.schmisser@math.univ-lille1.fr

In this paper, we consider a unidimensional piecewise deterministic Markov process (PDMP), with homogeneous jump rate $\lambda(x)$. This process is observed continuously, so the flow ϕ is known. To estimate nonparametrically the jump rate, we first construct an adaptive estimator of the stationary density, then we derive a quotient estimator $\hat{\lambda}_n$ of λ. Under some ergodicity conditions, we bound the risk of these estimators (and give a uniform bound on a small class of functions), and prove that the estimator of the jump rate is nearly minimax (up to a $\ln^2(n)$ factor). The simulations illustrate our theoretical results.

Keywords: Piecewise deterministic Markov processes; model selection; nonparametric estimation

References

[1] Azaïs, R., Bardet, J.-B., Génadot, A., Krell, N. and Zitt, P.-A. (2014). Piecewise deterministic Markov process – recent results. In *Journées MAS* 2012. *ESAIM Proc*. **44** 276–290. Les Ulis: EDP Sci. MR3178622 https://doi.org/10.1051/proc/201444017

[2] Azaïs, R., Dufour, F. and Gégout-Petit, A. (2014). Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes. *Scand. J. Stat*. **41** 950–969. MR3277032 https://doi.org/10.1111/sjos.12076

[3] Azaïs, R. and Genadot, A. (2018). A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions. *Comm. Statist. Theory Methods* **47** 1812–1829. MR3757712 https://doi.org/10.1080/03610926.2017.1327072

[4] Azaïs, R. and Muller-Gueudin, A. (2016). Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes. *Electron. J. Stat*. **10** 3648–3692. MR3579198 https://doi.org/10.1214/16-EJS1207

[5] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210

[6] Blanchard, Ph. and Jadczyk, A. (1995). Events and piecewise deterministic dynamics in event-enhanced quantum theory. *Phys. Lett. A* **203** 260–266. MR1346698 https://doi.org/10.1016/0375-9601(95)00432-3

[7] Cloez, B., Dessalles, R., Genadot, A., Malrieu, F., Marguet, A. and Yvinec, R. (2017). Probabilistic and piecewise deterministic models in biology. In *Journées MAS* 2016 *de la SMAI – Phénomènes Complexes et Hétérogènes*. *ESAIM Proc. Surveys* **60** 225–245. Les Ulis: EDP Sci. MR3772482 https://doi.org/10.1051/proc/201760225

[8] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262 https://doi.org/10.3150/07-BEJ5173

[9] Davis, M.H.A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. London: CRC Press. MR1283589 https://doi.org/10.1007/978-1-4899-4483-2

[10] de Saporta, B., Dufour, F. and Zhang, H. (2016). *Numerical Methods for Simulation and Optimization of Piecewise Deterministic Markov Processes*: *Application to Reliability*. *Mathematics and Statistics Series*. London: ISTE. MR3445895

[11] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Berlin: Springer. MR1261635 https://doi.org/10.1007/978-3-662-02888-9

[12] Doumic, M., Hoffmann, M., Krell, N. and Robert, L. (2015). Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli* **21** 1760–1799. MR3352060 https://doi.org/10.3150/14-BEJ623

[13] Dumas, V., Guillemin, F. and Robert, P. (2002). A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Adv. in Appl. Probab.* **34** 85–111. MR1895332 https://doi.org/10.1239/aap/1019160951

[14] Farid, M. and Davis, M.H.A. (1999). Optimal consumption and exploration: A case study in piecewise-deterministic Markov modelling. *Ann. Oper. Res.* **88** 121–137. Optimal control and differential games (Vienna, 1997). MR1704028 https://doi.org/10.1023/A:1018918027153

[15] Fujii, T. (2013). Nonparametric estimation for a class of piecewise-deterministic Markov processes. *J. Appl. Probab.* **50** 931–942. MR3161365 https://doi.org/10.1239/jap/1389370091

[16] Guillemin, F., Robert, P. and Zwart, B. (2004). AIMD algorithms and exponential functionals. *Ann. Appl. Probab.* **14** 90–117. MR2023017 https://doi.org/10.1214/aoap/1075828048

[17] Hodara, P., Krell, N. and Löcherbach, E. (2018). Non-parametric estimation of the spiking rate in systems of interacting neurons. *Stat. Inference Stoch. Process.* **21** 81–111. MR3769833 https://doi.org/10.1007/s11203-016-9150-4

[18] Höpfner, R. and Brodda, K. (2006). A stochastic model and a functional central limit theorem for information processing in large systems of neurons. *J. Math. Biol.* **52** 439–457. MR2235514 https://doi.org/10.1007/s00285-005-0361-3

[19] Krell, N. (2016). Statistical estimation of jump rates for a piecewise deterministic Markov processes with deterministic increasing motion and jump mechanism. *ESAIM Probab. Stat.* **20** 196–216. MR3528624 https://doi.org/10.1051/ps/2016013

[20] Krell, N. and Schmisser, É. (2021). Supplement (additional proofs) to nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes. https://doi.org/10.3150/20-BEJ1312SUPPA

[21] Krell, N. and Schmisser, É. (2021). Supplement (simulations) to nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes. https://doi.org/10.3150/20-BEJ1312SUPPB

[22] Laurençot, P. and Perthame, B. (2009). Exponential decay for the growth-fragmentation/cell-division equation. *Commun. Math. Sci.* **7** 503–510. MR2536450

[23] Meyer, Y. (1990). *Ondelettes et Opérateurs*. *I*. *Actualités Mathématiques*. [*Current Mathematical Topics*]. Paris: Hermann. Ondelettes. [Wavelets]. MR1085487

[24] Renault, V., Thieullen, M. and Trélat, E. (2017). Optimal control of infinite-dimensional piecewise deterministic Markov processes and application to the control of neuronal dynamics via optogenetics. *Netw. Heterog. Media* **12** 417–459. MR3714978 https://doi.org/10.3934/nhm.2017019

[25] Robert, L., Hoffmann, M., Krell, N., Aymerich, S., Robert, J. and Doumic, M. (2014). Division control in *Escherichia coli* is based on a size-sensing rather than timing mechanism. *BMC Biol.* **12** 17.

[26] Rudnicki, R. and Tyran-Kamińska, M. (2015). Piecewise deterministic Markov processes in biological models. In *Semigroups of Operators – Theory and Applications*. *Springer Proc. Math. Stat.* **113** 235–255. Cham: Springer. MR3295786 https://doi.org/10.1007/978-3-319-12145-1_15

[27] Tsybakov, A.B. (2004). *Introduction à L'estimation Non-paramétrique*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Berlin: Springer. MR2013911

Bernoulli **27**(4), 2021, 2389–2412
https://doi.org/10.3150/20-BEJ1313

Minimax semi-supervised set-valued approach to multi-class classification

EVGENII CHZHEN[1], CHRISTOPHE DENIS[2,*] and MOHAMED HEBIRI[2,†]

[1]*CNRS, Inria, Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, 91405 Orsay, France.*
E-mail: evgenii.chzhen@universite-paris-saclay.fr
[2]*Laboratoire d'Analyse et de Mathématiques Appliquées, Université Gustave Eiffel, 77454 Marne-la-Vallée cedex 2, France.*
E-mail: [*]*christophe.denis@univ-eiffel.fr;* [†]*mohamed.hebiri@univ-eiffel.fr*

We study supervised and semi-supervised algorithms in the set-valued classification framework with controlled expected size. While the former methods can use only n labeled samples, the latter are able to make use of N additional unlabeled data. We obtain semi-supervised minimax rates of convergence under the α-margin assumption and a β-Hölder condition on the conditional distribution of labels. Our analysis implies that if no further assumption is made, there is no supervised method that outperforms the semi-supervised estimator proposed in this work – the best achievable rate for any supervised method is $O(n^{-1/2})$, even if the margin assumption is extremely favorable; on the contrary, the developed semi-supervised estimator can achieve faster $O((n/\log n)^{-(1+\alpha)\beta/(2\beta+d)})$ rate of convergence provided that sufficiently many unlabeled samples are available. We also show that under additional smoothness assumption, supervised methods are able to achieve faster rates and the unlabeled sample cannot improve the rate of convergence. Finally, a numerical study supports our theory and emphasizes the relevance of the assumptions we required from an empirical perspective.

Keywords: Multi-class classification; set-valued classification; minimax optimality; semi-supervised classification

References

[1] Arlot, S. and Genuer, R. (2014). Analysis of purely random forests bias. Preprint. Available at arXiv:1407.3939.
[2] Audibert, J.-Y. and Tsybakov, A.B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217
[3] Bellec, P.C., Dalalyan, A.S., Grappin, E. and Paris, Q. (2018). On the prediction loss of the lasso in the partially labeled setting. *Electron. J. Stat.* **12** 3443–3472. MR3864589 https://doi.org/10.1214/18-EJS1457
[4] Birgé, L. (2005). A new lower bound for multiple hypothesis testing. *IEEE Trans. Inf. Theory* **51** 1611–1615. MR2241522 https://doi.org/10.1109/TIT.2005.844101
[5] Bobkov, S. and Ledoux, M. (2019). *One-Dimensional Empirical Measures, Order Statistics and Kantorovich Transport Distances. Mem. Amer. Math. Soc.* Providence: American Mathematical Society. https://doi.org/10.1090/memo/1259
[6] Breiman, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.
[7] Chzhen, E., Denis, C. and Hebiri, M. (2021). Supplement to "Minimax semi-supervised set-valued approach to multi-class classification." https://doi.org/10.3150/20-BEJ1313SUPP
[8] Denil, M., Matheson, D. and De Freitas, N. (2014). Narrowing the gap: Random forests in theory and in practice. In *International Conference on Machine Learning* 665–673.
[9] Denis, C. and Hebiri, M. (2017). Confidence sets with expected sizes for multiclass classification. *J. Mach. Learn. Res.* **18(102)** 1–28. MR3725441
[10] Gadat, S., Klein, T. and Marteau, C. (2016). Classification in general finite dimensional spaces with the k-nearest neighbor rule. *Ann. Statist.* **44** 982–1009. MR3485951 https://doi.org/10.1214/15-AOS1395
[11] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics.* New York: Springer. MR1920390 https://doi.org/10.1007/b97848

[12] Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270. MR0883354
[13] Herbei, R. and Wegkamp, M.H. (2006). Classification with reject option. *Canad. J. Statist.* **34** 709–721. MR2347054 https://doi.org/10.1002/cjs.5550340410
[14] Kpotufe, S. and Martinet, G. (2018). Marginal singularity, and the benefits of labels in covariate-shift. In *Conference on Learning Theory* 1882–1886.
[15] Lapin, M., Hein, M. and Schiele, B. (2015). Top-k multiclass SVM. In *Advances in Neural Information Processing Systems* 325–333.
[16] LeCun, Y. and Cortes, C. (2010). MNIST handwritten digit database.
[17] Lei, J. (2014). Classification with confidence. *Biometrika* **101** 755–769. MR3286915 https://doi.org/10.1093/biomet/asu038
[18] Ma, C. and Robinson, J. (1998). Approximations to distributions of sample quantiles. In *Order Statistics: Theory & Methods. Handbook of Statist.* **16** 463–484. Amsterdam: North-Holland. MR1668756 https://doi.org/10.1016/S0169-7161(98)16019-4
[19] Mammen, E. and Tsybakov, A.B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618 https://doi.org/10.1214/aos/1017939240
[20] Massart, P. and Nédélec, É. (2006). Risk bounds for statistical learning. *Ann. Statist.* **34** 2326–2366. MR2291502 https://doi.org/10.1214/009053606000000786
[21] Mourtada, J., Gaïffas, S. and Scornet, E. (2020). Minimax optimal rates for Mondrian trees and forests. *Ann. Statist.* **48** 2253–2276. MR4134794 https://doi.org/10.1214/19-AOS1886
[22] Oh, S. (2017). Top-k hierarchical classification. In *AAAI Conference on Artificial Intelligence*.
[23] Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348
[24] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters – An excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204 https://doi.org/10.1214/aos/1176324626
[25] Rigollet, P. (2007). Generalized error bounds in semi-supervised classification under the cluster assumption. *J. Mach. Learn. Res.* **8** 1369–1392. MR2332435 https://doi.org/10.1016/j.nonrwa.2006.08.003
[26] Rigollet, P. and Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. MR2597587 https://doi.org/10.3150/09-BEJ184
[27] Russakovsky, O., Deng, J., Su, H. et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115** 211–252. MR3422482 https://doi.org/10.1007/s11263-015-0816-y
[28] Sadinle, M., Lei, J. and Wasserman, L. (2019). Least ambiguous set-valued classifiers with bounded error levels. *J. Amer. Statist. Assoc.* **114** 223–234. MR3941250 https://doi.org/10.1080/01621459.2017.1395341
[29] Singh, A., Nowak, R. and Zhu, J. (2009). Unlabeled data: Now it helps, now it doesn't. In *NIPS* 1513–1520.
[30] Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. MR0443204
[31] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642
[32] Tsybakov, A.B. (1986). Robust reconstruction of functions by a local approximation method. *Problemy Peredachi Informatsii* **22** 69–84. MR0855002
[33] Tsybakov, A.B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735 https://doi.org/10.1214/aos/1069362732
[34] Tsybakov, A.B. (2008). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. New York: Springer. MR2724359 https://doi.org/10.1007/b13794
[35] Vapnik, V.N. (1998). *Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing, Communications, and Control*. New York: Wiley. A Wiley-Interscience Publication. MR1641250
[36] Vovk, V. (2002). Asymptotic optimality of transductive confidence machine. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **2533** 336–350. Berlin: Springer. MR2071624 https://doi.org/10.1007/3-540-36169-3_27
[37] Vovk, V. (2002). On-line confidence machines are well-calibrated. In *Proceedings of the Forty-Third Annual Symposium on Foundations of Computer Science* 187–196. Los Alamitos, CA: IEEE Computer Society.
[38] Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World*. New York: Springer. MR2161220

[39] Yang, Y. (1999). Minimax nonparametric classification. I. Rates of convergence. *IEEE Trans. Inf. Theory* **45** 2271–2284. MR1725115 https://doi.org/10.1109/18.796368
[40] You, M., Zhao, R.-W., Li, G.-Z. and Hu, X. (2011). MAPLSC: A novel multi-class classifier for medical diagnosis. *Int. J. Data Min. Bioinform.* **5** 383–401.
[41] Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., Vongsangnak, W. and Shen, B. (2013). Biomedical text mining and its applications in cancer research. *J. Biomed. Inform.* **46** 200–211.

Bernoulli **27**(4), 2021, 2413–2436
https://doi.org/10.3150/20-BEJ1314

Recovering Brownian and jump parts from high-frequency observations of a Lévy process

JORGE GONZÁLEZ CÁZARES[1] and JEVGENIJS IVANOVS[2]

[1] *University of Warwick & The Alan Turing Institute, United Kingdom.*
E-mail: jorge.gonzalez-cazares@warwick.ac.uk
[2] *Aarhus University, Denmark.*
E-mail: jevgenijs.ivanovs@math.au.dk

We introduce two general non-parametric methods for recovering paths of the Brownian and jump components from high-frequency observations of a Lévy process. The first procedure relies on reordering of independently sampled normal increments and thus avoids tuning parameters. The functionality of this method is a consequence of the small time predominance of the Brownian component, the presence of exchangeable structures, and fast convergence of normal empirical quantile functions. The second procedure amounts to filtering the increments and compensating with the final value. It requires a carefully chosen threshold, in which case both methods yield the same rate of convergence. This rate depends on the small-jump activity and is given in terms of the Blumenthal–Getoor index. Finally, we discuss possible extensions, including the multidimensional case, and provide numerical illustrations.

Keywords: Brownian bridge; coupling; exchangeability; high-frequency statistics; reordering of increments

References

[1] Ait-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *J. Econ. Lit.* **50** 1007–50.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871 https://doi.org/10.1017/CBO9780511721434

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

[4] Boyat, A. and Joshi, B. (2015). A review paper: Noise models in digital image processing. *Sampl. Theory Signal Image Process.* **6** 05.

[5] Cont, R. and Mancini, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli* **17** 781–813. MR2787615 https://doi.org/10.3150/10-BEJ293

[6] Deheuvels, P. (1985). The limiting behaviour of the maximal spacing generated by an i.i.d. sequence of Gaussian random variables. *J. Appl. Probab.* **22** 816–827. MR0808861 https://doi.org/10.1017/s0021900200108058

[7] del Barrio, E., Giné, E. and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245

[8] Eliazar, I. and Klafter, J. (2009). A unified and universal explanation for Lévy laws and 1/f noises. *Proc. Natl. Acad. Sci. USA* **106** 12251–12254.

[9] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events. Applications of Mathematics (New York)* **33**. Berlin: Springer. For insurance and finance. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

[10] González Cázares, J.I., Mijatović, A. and Uribe Bravo, G. (2021). Geometrically convergent simulation of the extrema of Lévy processes. *Math. Oper. Res.* To appear.

[11] Jacod, J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM Probab. Stat.* **11** 173–196. MR2320815 https://doi.org/10.1051/ps:2007013

[12] Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.* **118** 517–559. MR2394762 https://doi.org/10.1016/j.spa.2007.05.005

[13] Jacod, J. and Protter, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

[14] Jacod, J. and Reiss, M. (2014). A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *Ann. Statist.* **42** 1131–1144. MR3224283 https://doi.org/10.1214/13-AOS1179

[15] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer. MR1943877 https://doi.org/10.1007/978-3-662-05265-5

[16] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[17] Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles. Probability and Its Applications* (*New York*). New York: Springer. MR2161313

[18] Kunita, H. (2004). Representation of martingales with jumps and applications to mathematical finance. In *Stochastic Analysis and Related Topics in Kyoto. Adv. Stud. Pure Math.* **41** 209–232. Tokyo: Math. Soc. Japan. MR2083711 https://doi.org/10.2969/aspm/04110209

[19] Lee, S. and Hannig, J. (2010). Detecting jumps from Levy jump diffusion processes. *J. Financ. Econ.* **96** 05.

[20] Lee, S.S. and Mykland, P.A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financ. Stud.* **21** 2535–2563.

[21] Mao, X., Marion, G. and Renshaw, E. (2002). Environmental Brownian noise suppresses explosions in population dynamics. *Stochastic Process. Appl.* **97** 95–110. MR1870962 https://doi.org/10.1016/S0304-4149(01)00126-0

[22] Masuda, H. (2015). Parametric estimation of Lévy processes. In *Lévy Matters. IV. Lecture Notes in Math.* **2128** 179–286. Cham: Springer. MR3364258 https://doi.org/10.1007/978-3-319-12373-8_3

[23] Milotti, E. (2019). *The Physics of Noise*. Morgan & Claypool Publishers.

[24] Milstein, J., Mormann, F., Fried, I. and Koch, C. (2009). Neuronal shot noise and Brownian 1/f2 behavior in the local field potential. *PLoS ONE* **4** 1–5.

[25] Ogihara, T. and Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.* **14** 189–229. MR2832818 https://doi.org/10.1007/s11203-011-9057-z

[26] Pruitt, W.E. (1981). The growth of random walks and Lévy processes. *Ann. Probab.* **9** 948–956. MR0632968

[27] Sato, K. (2013). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation. MR3185174

[28] Shimizu, Y. (2008). A practical inference for discretely observed jump-diffusions from finite samples. *J. Japan Statist. Soc.* **38** 391–413. MR2510946 https://doi.org/10.14490/jjss.38.391

Bernoulli **27**(4), 2021, 2437–2472
https://doi.org/10.3150/20-BEJ1316

Local continuity of log-concave projection, with applications to estimation under model misspecification

RINA FOYGEL BARBER[1] and RICHARD J. SAMWORTH[2]

[1] *The University of Chicago Department of Statistics, George Herbert Jones Laboratory, Suite 222, 5747 S. Ellis Avenue, Chicago, IL 60637, United States of America.*
E-mail: rina@uchicago.edu; url: https://www.stat.uchicago.edu/~rina/
[2] *Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom.*
E-mail: r.samworth@statslab.cam.ac.uk; url: http://www.statslab.cam.ac.uk/~rjs57

The log-concave projection is an operator that maps a d-dimensional distribution P to an approximating log-concave density. It is known that, with suitable metrics on the underlying spaces, this projection is continuous, but not uniformly continuous. In this work, we prove a local uniform continuity result for log-concave projection – in particular, establishing that this map is locally Hölder-(1/4) continuous. A matching lower bound verifies that this exponent cannot be improved. We also examine the implications of this continuity result for the empirical setting – given a sample drawn from a distribution P, we bound the squared Hellinger distance between the log-concave projection of the empirical distribution of the sample, and the log-concave projection of P. In particular, this yields interesting statistical results for the misspecified setting, where P is not itself log-concave.

Keywords: Hellinger distance; Hölder continuity; log-concavity; maximum likelihood estimation; Wasserstein distance

References

[1] Aliprantis, C.D. and Border, K.C. (2006). *Infinite Dimensional Analysis*: *A Hitchhiker's Guide*, 3rd ed. Berlin: Springer. MR2378491
[2] 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 Series in Probability and Mathematical Statistics*. London–Sydney: Wiley. MR0326887
[3] Bellec, P.C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann*. *Statist*. **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566
[4] Birgé, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann*. *Statist*. **17** 1532–1549. MR1026298 https://doi.org/10.1214/aos/1176347380
[5] Cai, T.T. and Low, M.G. (2015). A framework for estimation of convex functions. *Statist*. *Sinica* **25** 423–456. MR3379081
[6] Carpenter, T., Diakonikolas, I., Sidiropoulos, A. and Stewart, A. (2018). Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. In *COLT* 2018.
[7] Chatterjee, S., Guntuboyina, A. and Sen, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann*. *Statist*. **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324
[8] Chen, Y. and Samworth, R.J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist*. *Sinica* **23** 1373–1398. MR3114718
[9] Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. *Wiley Series in Telecommunications*. New York: Wiley. A Wiley-Interscience Publication. MR1122806 https://doi.org/10.1002/0471200611
[10] 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 https://doi.org/10.1214/09-EJS505

[11] 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 https://doi.org/10.1111/j.1467-9868.2010.00753.x
[12] Doss, C.R. and Wellner, J.A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950 https://doi.org/10.1214/15-AOS1394
[13] Dudley, R.M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. Revised reprint of the 1989 original. MR1932358 https://doi.org/10.1017/CBO9780511755347
[14] Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798 https://doi.org/10.3150/08-BEJ141
[15] Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336 https://doi.org/10.1214/10-AOS853
[16] Durot, C. and Lopuhaä, H.P. (2018). Limit theory in monotone function estimation. *Statist. Sci.* **33** 547–567. MR3881208 https://doi.org/10.1214/18-STS664
[17] Fang, B. and Guntuboyina, A. (2019). On the risk of convex-constrained least squares estimators under misspecification. *Bernoulli* **25** 2206–2244. MR3961246 https://doi.org/10.3150/18-BEJ1051
[18] Feng, O.Y., Guntuboyina, A., Kim, A.K.H. and Samworth, R.J. (2021). Adaptation in multivariate log-concave density estimation. *Ann. Statist.* **49** 129–153. MR4206672 https://doi.org/10.1214/20-AOS1950
[19] Fresen, D. (2013). A multivariate Gnedenko law of large numbers. *Ann. Probab.* **41** 3051–3080. MR3127874 https://doi.org/10.1214/12-AOP804
[20] Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415 https://doi.org/10.1080/03461238.1956.10414944
[21] 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. Belmont, CA: Wadsworth. MR0822052
[22] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators, Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. New York: Cambridge Univ. Press. MR3445293 https://doi.org/10.1017/CBO9781139020893
[23] Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621 https://doi.org/10.1007/s00440-014-0595-3
[24] Han, Q. (2019). Global empirical risk minimizers with "shape constraints" are rate optimal in general dimensions. arXiv preprint arXiv:1905.12823.
[25] Han, Q., Wang, T., Chatterjee, S. and Samworth, R.J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753
[26] Han, Q. and Wellner, J.A. (2016). Multivariate convex regression: Global risk bounds and adaptation. arXiv preprint arXiv:1601.06844.
[27] Han, Q. and Wellner, J.A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408
[28] Hildreth, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. MR0065093
[29] Jankowski, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. MR3210981 https://doi.org/10.1214/13-AOS1196
[30] Kim, A.K.H., Guntuboyina, A. and Samworth, R.J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.* **46** 2279–2306. MR3845018 https://doi.org/10.1214/17-AOS1619
[31] Kim, A.K.H. and Samworth, R.J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560 https://doi.org/10.1214/16-AOS1480
[32] Koenker, R. and Mizera, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462 https://doi.org/10.1214/10-AOS814
[33] Kur, G., Dagan, Y. and Rakhlin, A. (2019). The log-concave maximum likelihood estimator is optimal in high dimensions. arXiv preprint arXiv:1903.05315v3.
[34] Lei, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151

[35] Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621 https://doi.org/10.1002/rsa.20135
[36] 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. Beachwood, OH: IMS. MR2459192 https://doi.org/10.1214/074921707000000184
[37] Patilea, V. (2001). Convex models, MLE and misspecification. *Ann*. *Statist*. **29** 94–123. MR1833960 https://doi.org/10.1214/aos/996986503
[38] Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677
[39] Samworth, R.J. (2018). Recent progress in log-concave density estimation. *Statist*. *Sci*. **33** 493–509. MR3881205 https://doi.org/10.1214/18-STS666
[40] Samworth, R.J. and Sen, B. (2018). Editorial: Special issue on "Nonparametric inference under shape constraints". *Statist*. *Sci*. **33** 469–472. MR3881203 https://doi.org/10.1214/18-STS673
[41] Samworth, R.J. and Yuan, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann*. *Statist*. **40** 2973–3002. MR3097966 https://doi.org/10.1214/12-AOS1060
[42] Schneider, R. (2014). *Convex Bodies*: *The Brunn–Minkowski Theory*, expanded ed. *Encyclopedia of Mathematics and Its Applications* **151**. Cambridge: Cambridge Univ. Press. MR3155183
[43] Schuhmacher, D., Hüsler, A. and Dümbgen, L. (2011). Multivariate log-concave distributions as a nearly parametric model. *Stat*. *Risk Model*. **28** 277–295. MR2838319 https://doi.org/10.1524/stnd.2011.1073
[44] Seijo, E. and Sen, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann*. *Statist*. **39** 1633–1657. MR2850215 https://doi.org/10.1214/10-AOS852
[45] Seregin, A. and Wellner, J.A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann*. *Statist*. **38** 3751–3781. MR2766867 https://doi.org/10.1214/10-AOS840
[46] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165
[47] Shorack, G.R. and Wellner, J.A. (2009). *Empirical Processes with Applications to Statistics*. *Classics in Applied Mathematics* **59**. Philadelphia, PA: SIAM. Reprint of the 1986 original [MR0838963]. MR3396731 https://doi.org/10.1137/1.9780898719017.ch1
[48] van de Geer, S.A. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge Univ. Press.
[49] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[50] Villani, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer. MR2459454 https://doi.org/10.1007/978-3-540-71050-9
[51] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann*. *Math*. *Stat*. **18** 309–348. MR0022330 https://doi.org/10.1214/aoms/1177730385
[52] Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J*. *Amer. Statist*. *Assoc*. **97** 508–513. MR1941467 https://doi.org/10.1198/016214502760047032
[53] Walther, G. (2009). Inference and modeling with log-concave distributions. *Statist*. *Sci*. **24** 319–327. MR2757433 https://doi.org/10.1214/09-STS303
[54] Xu, M. and Samworth, R.J. (2021). High-dimensional nonparametric density estimation via symmetry and shape constraints. *Ann*. *Statist*. **49** 650–672. https://doi.org/10.1214/20-AOS1972
[55] Yang, F. and Barber, R.F. (2019). Contraction and uniform convergence of isotonic regression. *Electron*. *J*. *Stat*. **13** 646–677. MR3914177 https://doi.org/10.1214/18-ejs1520
[56] Zhang, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

Bernoulli **27**(4), 2021, 2473–2493
https://doi.org/10.3150/20-BEJ1318

Minimum spanning trees of random geometric graphs with location dependent weights

GHURUMURUHAN GANESAN

Institute of Mathematical Sciences, Chennai, India.
E-mail: gganesan82@gmail.com

Consider n nodes $\{X_i\}_{1\leq i\leq n}$ independently distributed in the unit square S, each according to a distribution f. Nodes X_i and X_j are joined by an edge if the Euclidean distance $d(X_i, X_j)$ is less than r_n, the adjacency distance and the resulting random graph G_n is called a random geometric graph (RGG). We now assign a location dependent weight to each edge of G_n and define MST_n to be the sum of the weights of the minimum spanning trees of all components of G_n. For values of r_n above the connectivity regime, we obtain upper and lower bound deviation estimates for MST_n and L^2-convergence of MST_n appropriately scaled and centred.

Keywords: Minimum spanning tree; random geometric graphs; location dependent edge weights

References

[1] Alon, N. and Spencer, J.H. (2008). *The Probabilistic Method*, 3rd ed. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Hoboken, NJ: Wiley. MR2437651 https://doi.org/10.1002/9780470277331
[2] Beardwood, J., Halton, J.H. and Hammersley, J.M. (1959). The shortest path through many points. *Proc. Camb. Philos. Soc.* **55** 299–327. MR0109316 https://doi.org/10.1017/s0305004100034095
[3] Djauhari, M.A. and Gan, S.L. (2015). Optimality problem of network topology in stocks market analysis. *Phys. A, Stat. Mech. Appl.* **419** 108–114.
[4] Gupta, P. and Kumar, P.R. (1999). Critical power for asymptotic connectivity in wireless networks. In *Stochastic Analysis, Control, Optimization and Applications. Systems Control Found. Appl.* 547–566. Boston, MA: Birkhäuser. MR1702981
[5] Jothi, R. and Raghavachari, B. (2005). Approximation algorithms for the capacitated minimum spanning tree problem and its variants in network design. *ACM Trans. Algorithms* **1** 265–282. MR2190124 https://doi.org/10.1145/1103963.1103967
[6] Kesten, H. and Lee, S. (1996). The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6** 495–527. MR1398055 https://doi.org/10.1214/aoap/1034968141
[7] Penrose, M. (2003). *Random Geometric Graphs. Oxford Studies in Probability* **5**. Oxford: Oxford Univ. Press. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001
[8] Penrose, M.D. (2016). Connectivity of soft random geometric graphs. *Ann. Appl. Probab.* **26** 986–1028. MR3476631 https://doi.org/10.1214/15-AAP1110
[9] Penrose, M.D. and Yukich, J.E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303. MR1952000 https://doi.org/10.1214/aoap/1042765669
[10] Steele, J.M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16** 1767–1787. MR0958215
[11] Steele, J.M. (1993). Probability and problems in Euclidean combinatorial optimization. *Statist. Sci.* **8** 48–56.
[12] Yukich, J.E. (1998). *Probability Theory of Classical Euclidean Optimization Problems. Lecture Notes in Math.* **1675**. Berlin: Springer. MR1632875 https://doi.org/10.1007/BFb0093472
[13] Yukich, J.E. (2000). Asymptotics for weighted minimal spanning trees on random points. *Stochastic Process. Appl.* **85** 123–138. MR1730615 https://doi.org/10.1016/S0304-4149(99)00068-X

Bernoulli **27**(4), 2021, 2494–2518
https://doi.org/10.3150/20-BEJ1319

Online drift estimation for jump-diffusion processes

THEERAWAT BHUDISAKSANG* and ÁLVARO CARTEA†

Mathematical Institute, University of Oxford, Woodstock Rd, Oxford, OX2 6GG, UK.
E-mail: *theerawat.bhudisaksang@maths.ox.ac.uk; †alvaro.cartea@maths.ox.ac.uk

We show the convergence of an online stochastic gradient descent estimator to obtain the drift parameter of a continuous-time jump-diffusion process. The stochastic gradient descent follows a stochastic path in the gradient direction of a function to find a minimum, which in our case determines the estimate of the unknown drift parameter. We decompose the deviation of the stochastic descent direction from the deterministic descent direction into four terms: the weak solution of the non-local Poisson equation, a Riemann integral, a stochastic integral, and a covariation term. This decomposition is employed to prove the convergence of the online estimator and we use simulations to illustrate the performance of the online estimator.

Keywords: SGDCT; online estimation; jump-diffusion; extended Itô lemma; non-local Poisson equation; Lévy process

References

[1] Bertsekas, D.P. and Tsitsiklis, J.N. (2000). Gradient convergence in gradient methods with errors. *SIAM J. Optim*. **10** 627–642. MR1741189 https://doi.org/10.1137/S1052623497331063
[2] Bhudisaksang, T. and Cartea, Á. (2021). Supplement to "Online drift estimation for jump-diffusion processes." https://doi.org/10.3150/20-BEJ1319SUPP
[3] Bielecki, T., Chen, T. and Cialenco, I. (2017). Recursive construction of confidence regions. *Electron. J. Stat*. **11** 4674–4700. MR3724972 https://doi.org/10.1214/17-EJS1362
[4] Cartea, Á., Jaimungal, S. and Penalva, J. (2015). *Algorithmic and High-Frequency Trading*. Cambridge: Cambridge Univ. Press. MR0838085
[5] Delyon, B. (1996). General results on the convergence of stochastic algorithms. *IEEE Trans. Automat. Control* **41** 1245–1255. MR1409470 https://doi.org/10.1109/9.536495
[6] Errami, M., Russo, F. and Vallois, P. (2002). Itô's formula for $C^{1,\lambda}$-functions of a càdlàg process and related calculus. *Probab. Theory Related Fields* **19** 535–552. MR1894067
[7] Föllmer, H., Protter, P. and Shiryayev, A.N. (1995). Quadratic covariation and an extension of Itô's formula. *Bernoulli* **1** 149–169. MR1354459 https://doi.org/10.2307/3318684
[8] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat*. **24** 211–229. MR1455868 https://doi.org/10.1111/1467-9469.00059
[9] Kunita, H. (2004). Representation of martingales with jumps and applications to mathematical finance. In *Stochastic Analysis and Related Topics in Kyoto*. *Adv. Stud. Pure Math*. **41** 209–232. Tokyo: Math. Soc. Japan. MR2083711 https://doi.org/10.2969/aspm/04110209
[10] Kutoyants, Y.A. (2013). *Statistical Inference for Ergodic Diffusion Processes*. Berlin: Springer.
[11] Levanony, D., Shwartz, A. and Zeitouni, O. (1994). Recursive identification in continuous-time stochastic processes. *Stochastic Process. Appl*. **49** 245–275. MR1260193 https://doi.org/10.1016/0304-4149(94)90137-6
[12] Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl*. **117** 35–56. MR2287102 https://doi.org/10.1016/j.spa.2006.04.010
[13] Métivier, M. and Priouret, P. (1984). Applications of a Kushner and Clark lemma to general classes of stochastic algorithms. *IEEE Trans. Inf. Theory* **30** 140–151. MR0807052 https://doi.org/10.1109/TIT.1984.1056894

[14] Pardoux, E. and Veretennikov, A.Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

[15] Pardoux, È. and Veretennikov, A.Y. (2003). On Poisson equation and diffusion approximation. II. *Ann. Probab.* **31** 1166–1192. MR1988467 https://doi.org/10.1214/aop/1055425774

[16] Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. *Applications of Mathematics (New York)* **21**. Berlin: Springer. MR1037262 https://doi.org/10.1007/978-3-662-02619-9

[17] Sirignano, J. and Spiliopoulos, K. (2017). Stochastic gradient descent in continuous time. *SIAM J. Financial Math.* **8** 933–961. MR3732944 https://doi.org/10.1137/17M1126825

[18] Sirignano, J. and Spiliopoulos, K. (2020). Stochastic gradient descent in continuous time: A central limit theorem. *Stoch. Syst.* **10** 124–151. MR4119247 https://doi.org/10.1287/stsy.2019.0050

[19] Sørensen, M. (1991). Likelihood methods for diffusions with jumps. In *Statistical Inference in Stochastic Processes*. *Probab. Pure Appl.* **6** 67–105. New York: Dekker. MR1138259

[20] Surace, S.C. and Pfister, J.-P. (2019). Online maximum-likelihood estimation of the parameters of partially observed diffusion processes. *IEEE Trans. Automat. Control* **64** 2814–2829. MR3978276

[21] Uehara, Y. (2019). Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. *Stochastic Process. Appl.* **129** 4051–4081. MR3997671 https://doi.org/10.1016/j.spa.2018.11.007

[22] Veretennikov, A. and Kulik, A. (2012). Extended Poisson equation for weakly ergodic Markov processes. *Theory Probab. Math. Statist.* **85** 23–39. MR2933700

[23] Wang, J. (2010). Regularity of semigroups generated by Lévy type operators via coupling. *Stochastic Process. Appl.* **120** 1680–1700. MR2673970 https://doi.org/10.1016/j.spa.2010.04.007

[24] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41** 220–242. MR1172898 https://doi.org/10.1016/0047-259X(92)90068-Q

Bernoulli **27**(4), 2021, 2519–2531
https://doi.org/10.3150/20-BEJ1320

Universality and least singular values of random matrix products: A simplified approach

ROHIT CHAUDHURI[1,*], VISHESH JAIN[2] and NATESH S. PILLAI[1,†]

[1]*Department of Statistics, Harvard University, Cambridge, USA.*
E-mail: [*]*rohitchaudhuri@g.harvard.edu;* [†]*pillai@fas.harvard.edu*
[2]*Department of Statistics, Stanford University, Stanford, USA.*
E-mail: visheshj@stanford.edu

In this note, we show how to provide sharp control on the least singular value of a certain translated linearization matrix arising in the study of the local universality of products of independent random matrices. This problem was first considered in a recent work of Kopel, O'Rourke, and Vu, and compared to their work, our proof is substantially simpler and established in much greater generality. In particular, we only assume that the entries of the ensemble are centered, and have second and fourth moments uniformly bounded away from 0 and infinity, whereas previous work assumed a uniform subgaussian decay condition and that the entries within each factor of the product are identically distributed.

A consequence of our least singular value bound is that the four moment matching universality results for the products of independent random matrices, recently obtained by Kopel, O'Rourke, and Vu, hold under much weaker hypotheses. Our proof technique is also of independent interest in the study of structured sparse matrices.

Keywords: Universality; least singular value; product random matrices

References

[1] Burda, Z., Janik, R.A. and Waclaw, B. (2010). Spectrum of the product of independent random Gaussian matrices. *Phys. Rev. E* (3) **81** 041132, 12. MR2736204 https://doi.org/10.1103/PhysRevE.81.041132

[2] Cook, N.A., Guionnet, A. and Husson, J. (2020). Spectrum and pseudospectrum for quadratic polynomials in Ginibre matrices. Preprint. Available at arXiv:2008.08833.

[3] Jain, V., Jana, I., Luh, K. and O'Rourke, S. (2020). Circular law for random block band matrices with genuinely sublinear bandwidth. Preprint. Available at arXiv:2008.03850.

[4] Jain, V. and Silwal, S. (2020). A note on the universality of ESDs of inhomogeneous random matrices. *ALEA Lat. Am. J. Probab. Math. Stat.* To appear. Available at arXiv:2006.05418.

[5] Kopel, P., O'Rourke, S. and Vu, V. (2020). Random matrix products: Universality and least singular values. *Ann. Probab.* **48** 1372–1410. MR4112718 https://doi.org/10.1214/19-AOP1396

[6] Latała, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** 1273–1282. MR2111932 https://doi.org/10.1090/S0002-9939-04-07800-1

[7] Livshyts, G.V., Tikhomirov, K. and Vershynin, R. (2019). The smallest singular value of inhomogeneous square random matrices. *Ann. Probab.* To appear. Available at arXiv:1909.04219.

[8] Nemish, Y. (2017). Local law for the product of independent non-Hermitian random matrices with independent entries. *Electron. J. Probab.* **22** Paper No. 22, 35. MR3622892 https://doi.org/10.1214/17-EJP38

[9] Rudelson, M. and Vershynin, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633. MR2407948 https://doi.org/10.1016/j.aim.2008.01.010

[10] Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075 https://doi.org/10.1002/cpa.20294

Bernoulli **27**(4), 2021, 2532–2555
https://doi.org/10.3150/20-BEJ1321

Scalable Monte Carlo inference and rescaled local asymptotic normality

NING NING[*], EDWARD L. IONIDES[†] and YA'ACOV RITOV[‡]

Department of Statistics, University of Michigan, Ann Arbor.
E-mail: [*]*patning@umich.edu;* [†]*ionides@umich.edu;* [‡]*yritov@umich.edu*

In this paper, we generalize the property of local asymptotic normality (LAN) to an enlarged neighborhood, under the name of rescaled local asymptotic normality (RLAN). We obtain sufficient conditions for a regular parametric model to satisfy RLAN. We show that RLAN supports the construction of a statistically efficient estimator which maximizes a cubic approximation to the log-likelihood on this enlarged neighborhood. In the context of Monte Carlo inference, we find that this maximum cubic likelihood estimator can maintain its statistical efficiency in the presence of asymptotically increasing Monte Carlo error in likelihood evaluation.

Keywords: Monte Carlo; local asymptotic normality; big data; scalability

References

[1] Bickel, P., Choi, D., Chang, X. and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124
[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Baltimore, MD: Johns Hopkins Univ. Press. MR1245941
[3] Bickel, P.J. and Ritov, Y. (1996). Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli* **2** 199–228. MR1416863 https://doi.org/10.2307/3318520
[4] Bretó, C., Ionides, E.L. and King, A.A. (2020). Panel data analysis via mechanistic models. *J. Amer. Statist. Assoc.* **115** 1178–1188. MR4143457 https://doi.org/10.1080/01621459.2019.1604367
[5] Diggle, P.J. and Gratton, R.J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. Ser. B* **46** 193–227. MR0781880
[6] Hallin, M., Taniguchi, M., Serroukh, A. and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27** 2054–2080. MR1765628 https://doi.org/10.1214/aos/1017939250
[7] Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86** 105–129. MR1061951 https://doi.org/10.1007/BF01207516
[8] Ionides, E.L. (2005). Maximum smoothed likelihood estimation. *Statist. Sinica* **15** 1003–1014. MR2234410
[9] Ionides, E.L., Breto, C., Park, J., Smith, R.A. and King, A.A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface* **14** 1–10.
[10] Keener, R.W. (2010). *Theoretical Statistics*: *Topics for a Core Course*. *Springer Texts in Statistics*. New York: Springer. MR2683126 https://doi.org/10.1007/978-0-387-93839-4
[11] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. New York: Springer. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7
[12] Le Cam, L. and Yang, G.L. (2012). *Asymptotics in Statistics*: *Some Basic Concepts*. New York: Springer.
[13] Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168 https://doi.org/10.2307/2669386

1350-7265

[14] Ning, N., Ionides, E.L. and Ritov, Y. (2021). Supplement to "Scalable Monte Carlo inference and rescaled local asymptotic normality." https://doi.org/10.3150/20-BEJ1321SUPP
[15] Pons-Salort, M. and Grassly, N.C. (2018). Serotype-specific immunity explains the incidence of diseases caused by human enteroviruses. *Science* **361** 800–803. https://doi.org/10.1126/science.aat6777
[16] Ranjeva, S., Subramanian, R., Fang, V.J., Leung, G.M., Ip, D.K., Perera, R.A., Peiris, J.S.M., Cowling, B.J. and Cobey, S. (2019). Age-specific differences in the dynamics of protective immunity to influenza. *Nat*. *Commun*. **10** 1660.
[17] Ranjeva, S.L., Baskerville, E.B., Dukic, V., Villa, L.L., Lazcano-Ponce, E., Giuliano, A.R., Dwyer, G. and Cobey, S. (2017). Recurring infection with ecologically distinct HPV types can explain high prevalence and diversity. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **114** 13573–13578.
[18] Rao, C.R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear Models and Generalizations*. *Least Squares and Alternatives*, 3rd ed. *Springer Series in Statistics*. Berlin: Springer. MR2370506
[19] Smith, R.A., Ionides, E.L. and King, A.A. (2017). Infectious disease dynamics inferred from genetic data via sequential Monte Carlo. *Mol*. *Biol*. *Evol*. **34** 2065–2084. https://doi.org/10.1093/molbev/msx124
[20] Uhlenbeck, G.E. and Ornstein, L.S. (1930). On the theory of the Brownian motion. *Phys*. *Rev*. **36** 823.

Bernoulli **27**(4), 2021, 2556–2563
https://doi.org/10.3150/21-BEJ1322

Testing against uniform stochastic ordering with paired observations

DEWEI WANG[1] and CHUAN-FA TANG[2]

[1]*Department of Statistics, University of South Carolina, 1523 Greene Street, Columbia, SC 29208, USA.*
E-mail: deweiwang@stat.sc.edu
[2]*Department of Mathematical Sciences, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA.*
E-mail: chuan-fa.tang@utdallas.edu

This article develops a two-sample nonparametric goodness-of-fit (GOF) test for uniform stochastic ordering (USO) when observations are taken in pairs. We propose a data-driven critical value that controls the type I error and yields a consistent test. A simulation study illustrates the finite-sample performance of our test. All the proofs are included in the supplemental file.

Keywords: Bootstrap; copula; hazard rate ordering; least favorable configuration; order-restricted inference; ordinal dominance curve

References

[1] Arcones, M.A. and Samaniego, F.J. (2000). On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. *Ann. Statist.* **28** 116–150. MR1762906 https://doi.org/10.1214/aos/1016120367

[2] Balakrishnan, N., Zhang, Y. and Zhao, P. (2018). Ordering the largest claim amounts and ranges from two sets of heterogeneous portfolios. *Scand. Actuar. J.* **1** 23–41. MR3765136 https://doi.org/10.1080/03461238.2017.1278717

[3] Da, G. and Ding, W. (2016). Component level versus system level k-out-of-n assembly systems. *IEEE Trans. Reliab.* **65** 425–433.

[4] Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *J. Amer. Statist. Assoc.* **93** 1112–1123. MR1649205 https://doi.org/10.2307/2669854

[5] Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 https://doi.org/10.1007/BF01197342

[6] Dykstra, R., Kochar, S. and Rojo, J. (1997). Stochastic comparisons of parallel systems of heterogeneous exponential components. *J. Statist. Plann. Inference* **65** 203–211. MR1622774 https://doi.org/10.1016/S0378-3758(97)00058-X

[7] Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. MR3936869 https://doi.org/10.1093/restud/rdy049

[8] Hsieh, F. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24** 25–40. MR1389878 https://doi.org/10.1214/aos/1033066197

[9] Mukerjee, H. (1996). Estimation of survival functions under uniform stochastic ordering. *J. Amer. Statist. Assoc.* **91** 1684–1689. MR1439110 https://doi.org/10.2307/2291596

[10] Navarro, J. and Shaked, M. (2006). Hazard rate ordering of order statistics and systems. *J. Appl. Probab.* **43** 391–408. MR2248572 https://doi.org/10.1239/jap/1152413730

[11] Park, C.G., Lee, C.-I.C. and Robertson, T. (1998). Goodness-of-fit test for uniform stochastic ordering among several distributions. *Canad. J. Statist.* **26** 69–81. MR1624377 https://doi.org/10.2307/3315674

[12] Rojo, J. and Samaniego, F.J. (1991). On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statist. Probab. Lett.* **11** 267–271. MR1097984 https://doi.org/10.1016/0167-7152(91)90154-J

[13] Rojo, J. and Samaniego, F.J. (1993). On estimating a survival curve subject to a uniform stochastic ordering constraint. *J. Amer. Statist. Assoc.* **88** 566–572. MR1224382

[14] Tang, C.-F., Wang, D. and Tebbs, J.M. (2017). Nonparametric goodness-of-fit tests for uniform stochastic ordering. *Ann. Statist.* **45** 2565–2589. MR3737902 https://doi.org/10.1214/16-AOS1535

[15] Wang, D. and Tang, C.-F. (2021). Supplement to "Testing against uniform stochastic ordering with paired observations." https://doi.org/10.3150/21-BEJ1322SUPP

Bernoulli **27**(4), 2021, 2564–2597
https://doi.org/10.3150/21-BEJ1323

Over-parametrized deep neural networks minimizing the empirical risk do not generalize well

MICHAEL KOHLER[1] and ADAM KRZYŻAK[2]

[1]*Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany. E-mail: kohler@mathematik.tu-darmstadt.de*
[2]*Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, QC, Canada H3G 1M8.*
E-mail: krzyzak@cs.concordia.ca

Recently it was shown in several papers that backpropagation is able to find the global minimum of the empirical risk on the training data using over-parametrized deep neural networks. In this paper, a similar result is shown for deep neural networks with the sigmoidal squasher activation function in a regression setting, and a lower bound is presented which proves that these networks do not generalize well on a new data in the sense that networks which minimize the empirical risk do not achieve the optimal minimax rate of convergence for estimation of smooth regression functions.

Keywords: Neural networks; nonparametric regression; over-parametrization; rate of convergence

References

[1] Allen-Zhu, Z., Li, Y. and Liang, Y. (2019). Learning and generalization in overparameterized neural networks, going beyond two layers. In 33*rd Conference on Neural Information Processing Systems* (*NeurIPS* 2019). **32** Vancouver, Canada. 6155–6166.

[2] Allen-Zhu, Z., Li, Y. and Song, Z. (2019). A convergence theory for deep learning via over-parameterization. In *Proceedings of the* 36*th International Conference on Machine Learning* (*PMLR* 2019). **97** Long Beach, California. 242–252.

[3] Arora, S., Cohen, N., Golowich, N. and Hu, W. (2019). A convergence analysis of gradient descent for deep linear neural networks. In *International Conference on Learning Representations* (*ICLR* 2019). New Orleans, Louisiana.

[4] Bartlett, P.L., Long, P.M. and Lugosi, G. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070.

[5] Bauer, B. and Kohler, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Annals of Statistics* **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747

[6] Belkin, M., Hsu, D., Ma, S. and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

[7] Belkin, M., Rakhlin, A. and Tsybakov, A.B. (2020). Does data interpolation contradict statistical optimality? In *Proceedings of the* 22*nd International Conference on Artificial Intelligence and Statistics* (*AISTATS*). *Proceeding of Machine Learning Research* **119** 1611–1619.

[8] Braun, A., Kohler, M. and Walk, H. (2019). On the rate of convergence of a neural network regression estimate learned by gradient descent. Submitted for publication.

[9] Bubeck, S., Eldan, R. Lee, T. and Mikulincer, D. (2020). Network size and weights size for memorization with two-layers neural networks. Available at arXiv:2006.02855.

1350-7265

[10] Cao, Y. and Gu, Q. (2020). Generalization error bounds of gradient descent for learning over-parameterized deep relu networks. In *Proceedings of the* 34*th AAAI Conference on Artificial Intelligence*. **34** 3349–3356.

[11] Chizat, L., Oyallon, E. and Bach, E. (2020). On lazy training in differentiable programming. Available at arXiv:1812.07956.

[12] Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B. and LeCun, Y. (2015). The loss surface of multilayer networks. In *International Conference on Articial Intelligence and Statistics* (*AISTATS*) 2015. *Proceeding of Machine Learning Research* **38** San Diego, CA, USA. 192–204.

[13] Daniely, A. (2019). Neural networks learning and memorization with (almost) no over-parameterization. Available at arXiv:1911.09873.

[14] Daniely, A. (2020). Memorizing gaussians with no over-parameterization via gradient decent on neural networks. Available at arXiv:2003.12895.

[15] Devroye, L. Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. New York, USA: Springer.

[16] Devroye, L. and Wagner, J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics* **8** 231–239.

[17] Du, S. and Lee, J. (2018). On the power of over-parametrization in neural networks with quadratic activation. In *Proceedings of the* 35*th International Conference on Machine Learning* (*PMLR* 2018). **80** Stockholm, Sweden. 1329–1338.

[18] Du, S., Lee, J., Tian, Y., Poczos, B. and Singh, A. (2018). Gradient descent learns one-hidden-layer CNN: Don't be afraid of spurious local minima. In *Proceedings of the* 35*th International Conference on Machine Learning* (*PMLR* 2018) **80** Stockholm, Sweden. 1339–1348.

[19] Eckle, K. and Schmidt-Hieber, J. (2019). A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Networks* **110** 232–242.

[20] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution – Free Theory of Nonparametric Regression*. Berlin: Springer.

[21] Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R.J. (2019). Surprises in high-dimensional ridgeless least squares interpolation. Available at arXiv:1903.08560v4.

[22] Huang, J. and Yau, H.-T. (2020). Dynamics of deep neural networks and neural tangent hierarchy. In *Proceedings of the* 37*th International Conference on Machine Learning* (*PMLR* 2020). **119**. Available at arXiv:1909.08156v1.

[23] Huang, W., Du, W. and Xu, R.Y.D. (2020). On the neural tangent kernel of deep networks with orthogonal initialization. Available at arXiv:2004.05867v2.

[24] Imaizumi, M. and Fukamizu, K. (2019). Deep neural networks learn non-smooth functions effectively. In *Proceedings of the* 22*nd International Conference on Artificial Intelligence and Statistics* (*AISTATS* 2019). Naha, Okinawa, Japan. 869–878.

[25] Jacot, A. Gabriel, F. and Hongler, C. (2018). Neural tangent kernel: convergence and generalization in neural networks. In 32*nd Conference on Neural Information Processing Systems* (*NeurIPS* 2018). **31** Montreal, Canada. 8580–8589.

[26] Kawaguchi, K. (2016). Deep learning without poor local minima. In 30*th Conference on Neural Information Processing Systems* (*NIPS* 2016). Barcelona, Spain. 586–594.

[27] Kawaguchi, K. and Huang, J. (2019). Gradient descent finds global minima for generalizable deep neural networks of practical sizes. In 57*th Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*). 92–99.

[28] Kohler, M. and Krzyżak, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Trans*. *Inf. Theory* **63** 1620–1630. MR3625984 https://doi.org/10.1109/TIT.2016.2634401

[29] Kohler, M., Krzyżak, A. and Langer, S. (2019). Estimation of a function of low local dimensionality by deep neural networks. Submitted for publication. Available at arXiv:1908.11140.

[30] Kohler, M. and Langer, S. (2021). On the rate of convergence of fully connected very deep neural network regression estimates. *Annals of Statistics*. Available at arXiv:1908.11133. To appear.

[31] Liang, S., Sun, R., Lee, J. and Srikant, R. (2018). Adding one neuron can eliminate all bad local minima. In 32*nd Conference on Neural Information Processing Systems* (*NeurIPS* 2018) Montreal, Canada. 4355–4365.

[32] Montanari, A. and Zhong, Y. (2020). The interpolation phase transition in neural networks: Memorization and generalization under lazy training. Available at arXiv:2007.12826.

[33] Mücke, N. and Steinwart, I. (2019). Global minima of DNNs: The plenty pantry. Available at arXiv:1905.10686.

[34] Neyshabur, B., Li, Z., Bhojanapalli, S. LeCun, Y. and Srebro, N. (2019). The role of over-parametrization in generalization of neural networks. In 7*th International Conference on Learning Representations*, *ICLR* 2019.

[35] Nitanda, A. and Suzuki, T. (2020). Optimal rates for averaged stochastic gradient descent under neural tangent kernel regime. Available at arXiv:2006.12297v1.

[36] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks* **61** 85–117.

[37] Schmidt-Hieber, J. (2020a). Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics* **48** 1875–1897.

[38] Schmidt-Hieber, J. (2020b). Rejoinder to discussions of Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics* **48** 1916–1921.

[39] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* **10** 1040–1053. MR0673642

[40] Stone, C.J. (1985). Additive regression and other nonparametric models. *Annals of Statistics* **13** 689–705.

[41] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Annals of Statistics* **22** 118–184.

[42] Woodworth, B., Gunasekar, S., Lee, J., Moroshko, E., Savarese, P., Golan, I., Soudry, D. and Srebro, N. (2020). Kernel and rich regimes in overparametrized models. In 33*rd Annual Conference on Learning Theory*. *Proceeding of Machine Learning Research*. **125** 1–39.

Bernoulli **27**(4), 2021, 2598–2623
https://doi.org/10.3150/21-BEJ1324

Equidistribution of random walks on compact groups II. The Wasserstein metric

BENCE BORDA[1,2]

[1]*Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria*
[2]*Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13–15, 1053 Budapest, Hungary.*
E-mail: borda@math.tugraz.at

We consider a random walk S_k with i.i.d. steps on a compact group equipped with a bi-invariant metric. We prove quantitative ergodic theorems for the sum $\sum_{k=1}^{N} f(S_k)$ with Hölder continuous test functions f, including the central limit theorem, the law of the iterated logarithm and an almost sure approximation by a Wiener process, provided that the distribution of S_k converges to the Haar measure in the p-Wasserstein metric fast enough. As an example, we construct discrete random walks on an irrational lattice on the torus $\mathbb{R}^d/\mathbb{Z}^d$, and find their precise rate of convergence to uniformity in the p-Wasserstein metric. The proof uses a new Berry–Esseen type inequality for the p-Wasserstein metric on the torus, and the simultaneous Diophantine approximation properties of the lattice. These results complement the first part of this paper on random walks with an absolutely continuous component and quantitative ergodic theorems for Borel measurable test functions.

Keywords: Ergodic theorem; empirical distribution; central limit theorem; law of the iterated logarithm; Berry–Esseen inequality; simultaneous Diophantine approximation

References

[1] Basu, K. and Owen, A.B. (2016). Transformations and Hardy-Krause variation. *SIAM J. Numer. Anal.* **54** 1946–1966. MR3514716 https://doi.org/10.1137/15M1052184
[2] Berger, A. and Evans, S.N. (2013). A limit theorem for occupation measures of Lévy processes in compact groups. *Stoch. Dyn.* **13** 1250008, 16. MR3007246 https://doi.org/10.1142/S0219493712500086
[3] Berkes, I. and Borda, B. (2018). Berry-Esseen bounds and Diophantine approximation. *Anal. Math.* **44** 149–161. MR3814060 https://doi.org/10.1007/s10476-018-0203-3
[4] Berkes, I. and Borda, B. (2019). On the law of the iterated logarithm for random exponential sums. *Trans. Amer. Math. Soc.* **371** 3259–3280. MR3896111 https://doi.org/10.1090/tran/7415
[5] Berkes, I. and Borda, B. (2019). On the discrepancy of random subsequences of $\{n\alpha\}$ II.. *Acta Arith.* To appear.
[6] Bobkov, S. and Ledoux, M. A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.* To appear.
[7] Bobkov, S.G. (2016). Closeness of probability distributions in terms of Fourier-Stieltjes transforms. *Uspekhi Mat. Nauk* **71** 37–98. MR3588939 https://doi.org/10.4213/rm9749
[8] Bobkov, S.G. (2018). Berry-Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probab. Theory Related Fields* **170** 229–262. MR3748324 https://doi.org/10.1007/s00440-017-0756-2
[9] Bobkov, S.G. (2018). Central limit theorem and Diophantine approximations. *J. Theoret. Probab.* **31** 2390–2411. MR3866618 https://doi.org/10.1007/s10959-017-0770-4
[10] Bobkov, S.G. and Ledoux, M. (2020). Transport inequalities on Euclidean spaces for non-Euclidean metrics. *J. Fourier Anal. Appl.* **26** Paper No. 60, 27. MR4120057 https://doi.org/10.1007/s00041-020-09766-2
[11] Borda, B. Equidistribution of random walks on compact groups. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 54–72.

[12] Bourgain, J., Furman, A., Lindenstrauss, E. and Mozes, S. (2011). Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. *J. Amer. Math. Soc.* **24** 231–280. MR2726604 https://doi.org/10.1090/S0894-0347-2010-00674-1

[13] Boyer, J.-B. (2019). Almost sure functional central limit theorem for the linear random walk on the torus. *Probab. Theory Related Fields* **173** 651–696. MR3916116 https://doi.org/10.1007/s00440-018-0871-8

[14] Cassels, J.W.S. (1957). *An Introduction to Diophantine Approximation*. *Cambridge Tracts in Mathematics and Mathematical Physics*, *No*. 45. New York: Cambridge Univ. Press. MR0087708

[15] Cuny, C., Dedecker, J. and Jan, C. (2017). Limit theorems for the left random walk on $\mathrm{GL}_d(\mathbb{R})$. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1839–1865. MR3729637 https://doi.org/10.1214/16-AIHP773

[16] Diaconis, P. (1988). *Group Representations in Probability and Statistics*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. Hayward, CA: IMS. MR0964069

[17] Furman, A. and Shalom, Y. (1999). Sharp ergodic theorems for group actions and strong ergodicity. *Ergodic Theory Dynam. Systems* **19** 1037–1061. MR1709429 https://doi.org/10.1017/S0143385799133881

[18] Hensley, D. and Su, F.E. (2004). Random walks with badly approximable numbers. In *Unusual Applications of Number Theory*. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **64** 95–101. Providence, RI: Amer. Math. Soc. MR2063205 https://doi.org/10.1090/dimacs/064/10

[19] Hewitt, E. and Ross, K.A. (1979). *Abstract Harmonic Analysis*. *Vol. i*: *Structure of Topological Groups*, *Integration Theory*, *Group Representations*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **115**. Berlin: Springer. MR0551496

[20] Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*: *Pure and Applied Mathematics*. New York–London–Sydney: Wiley. MR0419394

[21] Loève, M. (1977). *Probability Theory. I*, 4th ed. *Graduate Texts in Mathematics* **45**. New York: Springer. MR0651017

[22] Móricz, F. (1976). Moment inequalities and the strong laws of large numbers. *Z. Wahrsch. Verw. Gebiete* **35** 299–314. MR0407950 https://doi.org/10.1007/BF00532956

[23] Niederreiter, H. and Philipp, W. (1973). Berry-Esseen bounds and a theorem of Erdős and Turán on uniform distribution mod 1. *Duke Math. J.* **40** 633–649. MR0337873

[24] Petrov, V.V. (1995). *Limit Theorems of Probability Theory*: *Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. Oxford University Press, New York: The Clarendon Press. MR1353441

[25] Robbins, H. (1953). On the equidistribution of sums of independent random variables. *Proc. Amer. Math. Soc.* **4** 786–799. MR0056869 https://doi.org/10.2307/2032412

[26] Schatte, P. (1988). On a law of the iterated logarithm for sums mod 1 with application to Benford's law. *Probab. Theory Related Fields* **77** 167–178. MR0927235 https://doi.org/10.1007/BF00334035

[27] Schatte, P. (1991). On a uniform law of the iterated logarithm for sums mod 1 and Benford's law. *Litovsk. Mat. Sb.* **31** 205–217. MR1161363 https://doi.org/10.1007/BF00972327

[28] Schmidt, W.M. (1980). *Diophantine Approximation*. *Lecture Notes in Math.* **785**. Berlin: Springer. MR0568710

[29] Sprindžuk, V.G. (1979). *Metric Theory of Diophantine Approximations*. Washington, D.C.: V. H. Winston & Sons; New York: Wiley. MR0548467

[30] Strassen, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley*, *Calif.*, 1965/66) 315–343. Berkeley, Calif.: Univ. California Press. MR0214118

[31] Stromberg, K. (1960). Probabilities on a compact group. *Trans. Amer. Math. Soc.* **94** 295–309. MR0114874 https://doi.org/10.2307/1993313

[32] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: Amer. Math. Soc. MR1964483 https://doi.org/10.1090/gsm/058

[33] von Bahr, B. (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88. MR0228044 https://doi.org/10.1007/BF02591678

Bernoulli **27**(4), 2021, 2624–2648
https://doi.org/10.3150/21-BEJ1325

Pure-jump semimartingales

ALEŠ ČERNÝ[1] and JOHANNES RUF[2]

[1]*Business School (formerly Cass), City, University of London, London, UK.*
E-mail: ales.cerny.1@city.ac.uk
[2]*Department of Mathematics, London School of Economics and Political Science, London, UK.*
E-mail: j.ruf@lse.ac.uk

A new integral with respect to an integer-valued random measure is introduced. In contrast to the finite variation integral ubiquitous in semimartingale theory, the new integral is closed under stochastic integration, composition, and smooth transformations. The new integral gives rise to a previously unstudied class of pure-jump processes – the sigma-locally finite variation pure-jump processes. As an application, it is shown that every semimartingale X has a unique decomposition

$$X = X_0 + X^{\mathrm{qc}} + X^{\mathrm{dp}},$$

where X^{qc} is quasi-left-continuous and X^{dp} is a sigma-locally finite variation pure-jump process that jumps only at predictable times, both starting at zero. The decomposition mirrors the classical result for local martingales and gives a rigorous meaning to the notions of continuous-time and discrete-time components of a semimartingale. Against this backdrop, the paper investigates a wider class of processes that are equal to the sum of their jumps in the semimartingale topology and constructs a taxonomic hierarchy of pure-jump semimartingales.

Keywords: Jump measure; Lévy process; predictable compensator; semimartingale topology; stochastic calculus

References

[1] Černý, A. and Ruf, J. (2020). Simplified stochastic calculus via semimartingale representations. Available at: https://arxiv.org/abs/2006.11914.

[2] Černý, A. and Ruf, J. (2021). Simplified stochastic calculus with applications in Economics and Finance. *European J. Oper. Res.* **293** 547–560. https://doi.org/10.1016/j.ejor.2020.12.037

[3] Dellacherie, C. and Meyer, P.-A. (1978). *Probabilities and Potential. North-Holland Mathematics Studies* **29**. Amsterdam: North-Holland. MR0521810

[4] Doléans-Dade, C. and Meyer, P.-A. (1970). Intégrales stochastiques par rapport aux martingales locales. In *Séminaire de Probabilités, IV* (*Univ. Strasbourg*, 1968/69). *Lecture Notes in Mathematics* **124** 77–107. Berlin: Springer. MR0270425

[5] Émery, M. (1979). Une topologie sur l'espace des semimartingales. In *Séminaire de Probabilités, XIII* (*Univ. Strasbourg, Strasbourg*, 1977/78). *Lecture Notes in Math.* **721** 260–280. Berlin: Springer. MR0544800

[6] Galtchouk, L.I. (1980). On the predictable jumps of martingales. In *Stochastic Differential Systems* (*Proc. IFIP-WG* 7/1 *Working Conf., Vilnius*, 1978). *Lecture Notes in Control and Information Sci.* **25** 50–57. Berlin: Springer. MR0609171

[7] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer. MR1943877 https://doi.org/10.1007/978-3-662-05265-5

[8] Kallsen, J. (2004). σ-Localization and σ-martingales. *Theory Probab. Appl.* **48** 152–163. MR2013413 https://doi.org/10.1137/S0040585X980312

[9] Kunita, H. and Watanabe, S. (1967). On square integrable martingales. *Nagoya Math. J.* **30** 209–245. MR0217856

[10] Larsson, M. and Ruf, J. (2020). Convergence of local supermartingales. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2774–2791. MR4164856 https://doi.org/10.1214/20-AIHP1058

[11] Perkowski, N. and Ruf, J. (2015). Supermartingales as Radon–Nikodym densities and related measure extensions. *Ann. Probab.* **43** 3133–3176. MR3433578 https://doi.org/10.1214/14-AOP956

[12] Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. *Applications of Mathematics (New York)* **21**. Berlin: Springer. MR1037262 https://doi.org/10.1007/978-3-662-02619-9

[13] Yoeurp, Ch. (1976). Décompositions des martingales locales et formules exponentielles. In *Séminaire de Probabilités, X (Seconde Partie: Théorie des Intégrales Stochastiques, Univ. Strasbourg, Strasbourg, Année Universitaire* 1974/1975). *Lecture Notes in Math*. **511** 432–480. MR0451395

Bernoulli **27**(4), 2021, 2649–2674
https://doi.org/10.3150/21-BEJ1326

Spectral-free estimation of Lévy densities in high-frequency regime

CÉLINE DUVAL[1] and ESTER MARIUCCI[2]

[1] *Université de Paris, MAP5, UMR CNRS 8145, Paris, France.*
E-mail: celine.duval@u-paris.fr
[2] *Otto von Guericke Universität Magdeburg, Germany.*
E-mail: ester.mariucci@uvsq.fr

We construct an estimator of the Lévy density of a pure jump Lévy process, possibly of infinite variation, from the discrete observation of one trajectory at high frequency. The novelty of our procedure is that we directly estimate the Lévy density relying on a pathwise strategy, whereas existing procedures rely on spectral techniques. By taking advantage of a compound Poisson approximation, we circumvent the use of spectral techniques and in particular of the Lévy–Khintchine formula. A linear wavelet estimator is built and its performance is studied in terms of L_p loss functions, $p \geq 1$, over Besov balls. We recover classical nonparametric rates for finite variation Lévy processes and for a nonparametric class of symmetric infinite variation Lévy processes. We show that the procedure is robust when the estimation set gets close to the critical value 0 and also discuss its robustness to the presence of a Brownian part.

Keywords: Lévy density estimation; infinite variation; Lévy processes; nonparametric estimation

References

[1] Adams, R.A. and Fournier, J.J.F. (2003). *Sobolev Spaces*, 2nd ed. *Pure and Applied Mathematics* (*Amsterdam*) **140**. Amsterdam: Elsevier/Academic Press. MR2424078

[2] Barndorff-Nielsen, O.E. and Hubalek, F. (2008). Probability measures, Lévy measures and analyticity in time. *Bernoulli* **14** 764–790. MR2537811 https://doi.org/10.3150/07-BEJ6114

[3] Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I. (2012). *Lévy Processes*: *Theory and Applications*. Springer.

[4] Bec, M. and Lacour, C. (2015). Adaptive pointwise estimation for pure jump Lévy processes. *Stat*. *Inference Stoch*. *Process*. **18** 229–256. MR3395606 https://doi.org/10.1007/s11203-014-9113-6

[5] Belomestny, D., Comte, F., Genon-Catalot, V., Masuda, H. and Reiß, M. (2015). *Lévy Matters*. *IV*: *Estimation for Discretely Observed Lévy Processes*, *Lévy Matters*. *Lecture Notes in Math*. **2128**. Cham: Springer. MR3364253

[6] Belomestny, D., Gugushvili, S., Schauer, M. and Spreij, P. (2019). Nonparametric Bayesian inference for gamma-type Lévy subordinators. *Commun*. *Math*. *Sci*. **17** 781–816. MR4001482 https://doi.org/10.4310/CMS.2019.v17.n3.a8

[7] Biagini, F., Bregman, J. and Meyer-Brandis, T. (2015). Electricity futures price modeling with Lévy term structure models. *Int*. *J*. *Theor*. *Appl*. *Finance* **18** 1550003, 21. MR3315829 https://doi.org/10.1142/S021902491550003X

[8] Boxma, O., Ivanovs, J., Kosiński, K. and Mandjes, M. (2011). Lévy-driven polling systems and continuous-state branching processes. *Stoch*. *Syst*. **1** 411–436. MR2949545 https://doi.org/10.1214/10-SSY008

[9] Bretagnolle, J. and Huber, C. (1979). Estimation des densités: Risque minimax. *Z*. *Wahrsch*. *Verw*. *Gebiete* **47** 119–137. MR0523165 https://doi.org/10.1007/BF00535278

[10] Buchmann, B. and Grübel, R. (2003). Decompounding: An estimation problem for Poisson random sums. *Ann*. *Statist*. **31** 1054–1074. MR2001642 https://doi.org/10.1214/aos/1059655905

1350-7265

[11] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus.* **75** 305–333.

[12] Coca, A. (2018). Adaptive nonparametric estimation for compound poisson processes robust to the discrete-observation scheme. arXiv:1803.09849.

[13] Coca, A.J. (2018). Efficient nonparametric inference for discretely observed compound Poisson processes. *Probab. Theory Related Fields* **170** 475–523. MR3748329 https://doi.org/10.1007/s00440-017-0761-5

[14] Comte, F. and Genon-Catalot, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.* **119** 4088–4123. MR2565560 https://doi.org/10.1016/j.spa.2009.09.013

[15] Comte, F. and Genon-Catalot, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 595–617. MR2682259 https://doi.org/10.1214/09-AIHP323

[16] Comte, F. and Genon-Catalot, V. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.* **39** 803–837. MR2816339 https://doi.org/10.1214/10-AOS856

[17] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. *Chapman & Hall/CRC Financial Mathematics Series*. Boca Raton, FL: CRC Press/CRC. MR2042661

[18] Duval, C. (2013). Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.* **123** 3963–3986. MR3091096 https://doi.org/10.1016/j.spa.2013.06.006

[19] Duval, C. and Mariucci, E. Non-asymptotic control of the cumulative distribution function of Lévy processes. arXiv:2003.09281.

[20] Duval, C. and Mariucci, E. (2021). Supplement to "Spectral-free estimation of Lévy densities in high-frequency regime." https://doi.org/10.3150/21-BEJ1326SUPP

[21] Figueroa-López, J.E. (2009). Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **57** 117–146. Beachwood, OH: IMS. MR2681661 https://doi.org/10.1214/09-LNMS5709

[22] Figueroa-López, J.E. (2011). Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli* **17** 643–670. MR2787609 https://doi.org/10.3150/10-BEJ286

[23] Figueroa-López, J.E. and Houdré, C. (2009). Small-time expansions for the transition distributions of Lévy processes. *Stochastic Process. Appl.* **119** 3862–3889. MR2552308 https://doi.org/10.1016/j.spa.2009.09.002

[24] Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590 https://doi.org/10.3150/08-BEJ144

[25] Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21** 321–343. MR2530929 https://doi.org/10.1080/10485250802645824

[26] Gugushvili, S. (2012). Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 282–307. MR2919207 https://doi.org/10.1214/11-AIHP433

[27] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (2012). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statistics* **129**. New York: Springer. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

[28] Hoffmann, M. and Vetter, M. (2017). Weak convergence of the empirical truncated distribution function of the Lévy measure of an Itô semimartingale. *Stochastic Process. Appl.* **127** 1517–1543. MR3630234 https://doi.org/10.1016/j.spa.2016.08.009

[29] Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.* **124** 730–758. MR3131312 https://doi.org/10.1016/j.spa.2013.08.010

[30] Kappus, J. (2018). Nonparametric estimation for irregularly sampled Lévy processes. *Stat. Inference Stoch. Process.* **21** 141–167. MR3769835 https://doi.org/10.1007/s11203-016-9144-2

[31] Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24. MR1147634 https://doi.org/10.1016/0167-7152(92)90231-S

[32] Marchal, P. (2009). Small time expansions for transition probabilities of some Lévy processes. *Electron. Commun. Probab.* **14** 132–142. MR2481673 https://doi.org/10.1214/ECP.v14-1452

[33] Mariucci, E. (2016). Asymptotic equivalence for pure jump Lévy processes with unknown Lévy density and Gaussian white noise. *Stochastic Process. Appl.* **126** 503–541. MR3434992 https://doi.org/10.1016/j.spa.2015.09.009

[34] Neumann, M.H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805 https://doi.org/10.3150/08-BEJ148

[35] Nickl, R. and Reiß, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.* **263** 3306–3332. MR2973342 https://doi.org/10.1016/j.jfa.2012.08.012

[36] Nickl, R., Reiß, M., Söhl, J. and Trabs, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields* **164** 61–108. MR3449386 https://doi.org/10.1007/s00440-014-0607-3

[37] Noven, R.C., Veraart, A.E.D. and Gandy, A. (2015). A Lévy-driven rainfall model with applications to futures pricing. *AStA Adv. Stat. Anal.* **99** 403–432. MR3397093 https://doi.org/10.1007/s10182-015-0246-8

[38] Rüschendorf, L. and Woerner, J.H.C. (2002). Expansion of transition distributions of Lévy processes in small time. *Bernoulli* **8** 81–96. MR1884159

[39] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. Translated from the 1990 Japanese original, Revised by the author. MR1739520

[40] Trabs, M. (2015). Quantile estimation for Lévy measures. *Stochastic Process. Appl.* **125** 3484–3521. MR3357617 https://doi.org/10.1016/j.spa.2015.04.004

[41] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli* **13** 672–694. MR2348746 https://doi.org/10.3150/07-BEJ6091

Bernoulli **27**(4), 2021, 2675–2713
https://doi.org/10.3150/21-BEJ1327

A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation

CHRISTOPHE DENIS[1,3,*], CHARLOTTE DION-BLANC[2] and MIGUEL MARTINEZ[1]

[1] *LAMA, Université Gustave Eiffel, France.*
E-mail: [*] *christophe.denis@u-pem.fr*
[2] *LPSM, UMR CNRS 8001, Sorbonne Université, France*
[3] *MIA-PARIS, AgroParisTech, INRAE, Université Paris-Saclay, France*

This work focuses on the nonparametric estimation of a drift function from N discrete repeated independent observations of a diffusion process over a *fixed* time interval $[0, T]$. We study a ridge estimator obtained by the minimization of a constrained least squares contrast. The resulting projection estimator is based on the B-spline basis. Under mild assumptions, this estimator is universally consistent with respect to an integrate norm. We establish that, up to a logarithmic factor and when the estimation is performed on a compact interval, our estimation procedure reaches the best possible rate of convergence. Furthermore, we build an adaptive estimator that achieves this rate. Finally, we illustrate our procedure through an intensive simulation study which highlights the good performance of the proposed estimator in various models.

Keywords: Stochastic differential equation; nonparametric estimation; ridge estimator

References

[1] Abraham, K. (2019). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli* **25** 2696–2728. MR4003562 https://doi.org/10.3150/18-BEJ1067
[2] Baraud, Y., Comte, F. and Viennet, G. (2001). Model selection for (auto-)regression with dependent data. *ESAIM Probab. Stat.* **5** 33–49. MR1845321 https://doi.org/10.1051/ps:2001101
[3] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39. MR1354454 https://doi.org/10.2307/3318679
[4] Comte, F. and Genon-Catalot, V. (2020). Nonparametric drift estimation for I.I.d. paths of stochastic differential equations. *Ann. Statist.* **48** 3336–3365. MR4185811 https://doi.org/10.1214/19-AOS1933
[5] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262 https://doi.org/10.3150/07-BEJ5173
[6] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507–2528. MR2253093 https://doi.org/10.1214/009053605000000615
[7] de Boor, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences* **27**. New York-Berlin: Springer. MR0507062
[8] Denis, C., Dion, C. and Martinez, M. (2020). Consistent procedures for multiclass classification of discrete diffusion paths. *Scand. J. Stat.* **47** 516–554. MR4157149 https://doi.org/10.1111/sjos.12415
[9] Fournier, N. and Printems, J. (2010). Absolute continuity for some one-dimensional processes. *Bernoulli* **16** 343–360. MR2668905 https://doi.org/10.3150/09-BEJ215
[10] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 711–737. MR1931584 https://doi.org/10.1016/S0246-0203(02)01107-X
[11] Gobet, E., Hoffmann, M. and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 https://doi.org/10.1214/009053604000000797

1350-7265

[12] Graham, C. and Talay, D. (2013). *Stochastic Simulation and Monte Carlo Methods*: *Mathematical Foundations of Stochastic Simulation*. *Stochastic Modelling and Applied Probability* **68**. Heidelberg: Springer. MR3097957 https://doi.org/10.1007/978-3-642-39363-1
[13] Gugushvili, S. and Spreij, P. (2014). Consistent non-parametric Bayesian estimation for a time-inhomogeneous Brownian motion. *ESAIM Probab. Stat.* **18** 332–341. MR3333993 https://doi.org/10.1051/ps/2013039
[14] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. New York: Springer. MR1920390 https://doi.org/10.1007/b97848
[15] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2722294 https://doi.org/10.1007/978-0-387-84858-7
[16] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process*. *Appl*. **79** 135–163. MR1670522 https://doi.org/10.1016/S0304-4149(98)00074-X
[17] Iacus, S.M. (2008). *Simulation and Inference for Stochastic Differential Equations*: *With* ßf *R Examples*. *Springer Series in Statistics*. New York: Springer. MR2410254 https://doi.org/10.1007/978-0-387-75839-8
[18] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314. MR1681700 https://doi.org/10.2307/3318437
[19] Konakov, V. and Menozzi, S. (2017). Weak error for the Euler scheme approximation of diffusions with non-smooth coefficients. *Electron*. *J*. *Probab*. **22** Paper No. 46, 47. MR3661660 https://doi.org/10.1214/17-EJP53
[20] Koskela, J., Spanò, D. and Jenkins, P.A. (2019). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. *Bernoulli* **25** 2183–2205. MR3961245 https://doi.org/10.3150/18-BEJ1050
[21] Kutoyants, Y.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. London: Springer London, Ltd. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2
[22] Lorentz, G.G., Golitschek, M.V. and Makovoz, Y. (1996). *Constructive Approximation*: *Advanced Problems*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **304**. Berlin: Springer. MR1393437 https://doi.org/10.1007/978-3-642-60932-9
[23] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993
[24] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Berlin: Springer. MR1725357 https://doi.org/10.1007/978-3-662-06400-9
[25] Schmisser, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47** 61–84. MR3023015 https://doi.org/10.1080/02331888.2011.591931
[26] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359 https://doi.org/10.1007/b13794
[27] van der Meulen, F. and van Zanten, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485 https://doi.org/10.3150/11-BEJ385
[28] Wang, J.L., Chiou, J.M. and Mueller, H.G. (2015). Review of functional data analysis. Preprint. Available at arXiv:1507.05135.
[29] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal*. **41** 220–242. MR1172898 https://doi.org/10.1016/0047-259X(92)90068-Q

Bernoulli **27**(4), 2021, 2714–2739
https://doi.org/10.3150/21-BEJ1328

Approximation of occupation time functionals

RANDOLF ALTMEYER

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
E-mail: altmeyrx@math.hu-berlin.de

The strong L^2-approximation of occupation time functionals is studied with respect to discrete observations of a d-dimensional càdlàg process. Upper bounds on the error are obtained under weak assumptions, generalizing previous results in the literature considerably. The approach relies on regularity for the marginals of the process and applies also to non-Markovian processes, such as fractional Brownian motion. The results are used to approximate occupation times and local times. For Brownian motion, the upper bounds are shown to be sharp up to a log-factor.

Keywords: Occupation time; local time; stochastic process; integral functional; heat kernel bounds; fractional Brownian motion; lower bound

References

[1] Altmeyer, R. (2019). Central limit theorems for discretized occupation time functionals. Preprint. Available at arXiv:1909.00474.
[2] Altmeyer, R. and Chorowski, J. (2018). Estimation error for occupation time functionals of stationary Markov processes. *Stochastic Process. Appl.* **128** 1830–1848. MR3797645 https://doi.org/10.1016/j.spa.2017.08.013
[3] Berman, S.M. (1969). Local times and sample function properties of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **137** 277–299. MR0239652 https://doi.org/10.2307/1994804
[4] Berman, S.M. (1973/74). Local nondeterminism and local times of Gaussian processes. *Indiana Univ. Math. J.* **23** 69–94. MR0317397 https://doi.org/10.1512/iumj.1973.23.23006
[5] Chorowski, J. (2018). Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies. *Bernoulli* **24** 2934–2990. MR3779707 https://doi.org/10.3150/17-BEJ950
[6] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. *Chapman & Hall/CRC Financial Mathematics Series*. Boca Raton, FL: CRC Press/CRC. MR2042661
[7] Cruz-Uribe, D. and Neugebauer, C.J. (2002). Sharp error bounds for the trapezoidal rule and Simpson's rule. *JIPAM. J. Inequal. Pure Appl. Math.* **3** Article 49, 22. MR1923348
[8] Diaconis, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics*, *IV*, *Vol.* 1 (*West Lafayette*, *Ind.*, 1986) 163–175. New York: Springer. MR0927099
[9] Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804. MR1242012 https://doi.org/10.2307/3214513
[10] Friedman, A. (1964). *Partial Differential Equations of Parabolic Type*. Englewood Cliffs, NJ: Prentice-Hall, Inc. MR0181836
[11] Ganychenko, I. (2015). Fast L_2-approximation of integral-type functionals of Markov processes. *Mod. Stoch. Theory Appl.* **2** 165–171. MR3389588 https://doi.org/10.15559/15-vmsta29
[12] Ganychenko, I., Knopova, V. and Kulik, A. (2015). Accuracy of discrete approximation for integral functionals of Markov processes. *Mod. Stoch. Theory Appl.* **2** 401–420. MR3456146 https://doi.org/10.15559/15-VMSTA46
[13] Ganychenko, I. and Kulik, A. (2014). Rates of approximation of nonsmooth integral-type functionals of Markov processes. *Mod. Stoch. Theory Appl.* **1** 117–126. MR3316480 https://doi.org/10.15559/vmsta-2014.12
[14] Geman, D. and Horowitz, J. (1980). Occupation densities. *Ann. Probab.* **8** 1–67. MR0556414
[15] Gobet, E. and Labart, C. (2008). Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electron. Commun. Probab.* **13** 352–363. MR2415143 https://doi.org/10.1214/ECP.v13-1393

[16] Ivanovs, J. and Podolskij, M. (2020). Optimal estimation of some random quantities of a Lévy process. Preprint. Available at arXiv:2001.02517.
[17] Jacod, J. (1998). Rates of convergence to the local time of a diffusion. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 505–544. MR1632849 https://doi.org/10.1016/S0246-0203(98)80026-5
[18] Jacod, J. and Protter, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7
[19] Jaramillo, A., Nourdin, I. and Peccati, G. (2021). Approximation of fractional local times: Zero energy and weak derivatives. *Ann. Appl. Probab.* To appear.
[20] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. New York: Springer. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2
[21] Knopova, V. and Kulik, A. (2018). Parametrix construction of the transition probability density of the solution to an SDE driven by α-stable noise. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 100–140. MR3765882 https://doi.org/10.1214/16-AIHP796
[22] Kohatsu-Higa, A., Makhlouf, A. and Ngo, H.L. (2014). Approximations of non-smooth integral type functionals of one dimensional diffusion processes. *Stochastic Process. Appl.* **124** 1881–1909. MR3170228 https://doi.org/10.1016/j.spa.2014.01.003
[23] Kühn, F. (2019). Transition probabilities of Lévy-type processes: Parametrix construction. *Math. Nachr.* **292** 358–376. MR3912204 https://doi.org/10.1002/mana.201700441
[24] Marcus, M.B. and Rosen, J. (2008). L^p moduli of continuity of Gaussian processes and local times of symmetric Lévy processes. *Ann. Probab.* **36** 594–622. MR2393991 https://doi.org/10.1214/009117907000000277
[25] Neuenkirch, A. and Szölgyenyi, M. (2019). The Euler–Maruyama scheme for SDEs with irregular drift: Convergence rates via reduction to a quadrature problem. *IMA J. Numer. Anal.* https://doi.org/10.1093/imanum/draa007
[26] Ngo, H.-L. and Ogawa, S. (2011). On the discrete approximation of occupation time of diffusion processes. *Electron. J. Stat.* **5** 1374–1393. MR2842909 https://doi.org/10.1214/11-EJS645
[27] Nualart, D. (1995). *The Malliavin Calculus and Related Topics. Probability and Its Applications* (*New York*). New York: Springer. MR1344217 https://doi.org/10.1007/978-1-4757-2437-0
[28] Ritter, K. (2000). *Average-Case Analysis of Numerical Problems. Lecture Notes in Math.* **1733**. Berlin: Springer. MR1763973 https://doi.org/10.1007/BFb0103934
[29] Sacks, J. and Ylvisaker, N.D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Stat.* **37** 66–89. MR0192601 https://doi.org/10.1214/aoms/1177699599
[30] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR1739520
[31] Xiao, Y. (2006). Properties of local-nondeterminism of Gaussian and stable random fields and their applications. *Ann. Fac. Sci. Toulouse Math.* (6) **15** 157–193. MR2225751

Bernoulli **27**(4), 2021, 2740–2765
https://doi.org/10.3150/21-BEJ1329

Extremal eigenvalues of sample covariance matrices with general population

JINWOONG KWAK*, JI OON LEE† and JAEWHI PARK‡

Department of Mathematical Sciences, KAIST, Republic of Korea.
E-mail: *jw-kwak@kaist.ac.kr; †jioon.lee@kaist.edu; ‡jw-park@kaist.ac.kr

We consider the eigenvalues of sample covariance matrices of the form $\mathcal{Q} = (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*$. The sample X is an $M \times N$ rectangular random matrix with real independent entries and the population covariance matrix Σ is a positive definite diagonal matrix independent of X. Assuming that the limiting spectral density of Σ exhibits convex decay at the right edge of the spectrum, in the limit $M, N \to \infty$ with $N/M \to d \in (0, \infty)$, we find a certain threshold d_+ such that for $d > d_+$ the limiting spectral distribution of $\mathcal{Q}$ also exhibits convex decay at the right edge of the spectrum. In this case, the largest eigenvalues of $\mathcal{Q}$ are determined by the order statistics of the eigenvalues of Σ, and in particular, the limiting distribution of the largest eigenvalue of $\mathcal{Q}$ is given by a Weibull distribution. In case $d < d_+$, we also prove that the limiting distribution of the largest eigenvalue of $\mathcal{Q}$ is Gaussian if the entries of Σ are i.i.d. random variables. While Σ is considered to be random mostly, the results also hold for deterministic Σ with some additional assumptions.

Keywords: Sample covariance matrix; deformed Marchenko–Pastur distribution; largest eigenvalue

References

[1] Aparicio, L., Bordyuh, M., Blumberg, A.J. and Rabadan, R. (2020). A random matrix theory approach to denoise single-cell data. *Patterns* **1** 100035.
[2] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233
[3] Bao, Z., Pan, G. and Zhou, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281
[4] Bloemendal, A. and Virág, B. (2013). Limits of spiked random matrices I. *Probab. Theory Related Fields* **156** 795–825. MR3078286 https://doi.org/10.1007/s00440-012-0443-2
[5] Bloemendal, A. and Virág, B. (2016). Limits of spiked random matrices II. *Ann. Probab.* **44** 2726–2769. MR3531679 https://doi.org/10.1214/15-AOP1033
[6] El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592 https://doi.org/10.1214/009117906000000917
[7] Erdős, L. (2010). Universality of Wigner random matrices. In *XVIth International Congress on Mathematical Physics* 86–105. Hackensack, NJ: World Sci. Publ. MR2730790 https://doi.org/10.1142/9789814304634_0004
[8] Erdős, L., Knowles, A. and Yau, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y
[9] Erdős, L., Knowles, A., Yau, H.-T. and Yin, J. (2012). Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770 https://doi.org/10.1007/s00220-012-1527-7
[10] Erdős, L., Knowles, A., Yau, H.-T. and Yin, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** 59. MR3068390 https://doi.org/10.1214/EJP.v18-2473
[11] Erdős, L., Yau, H.-T. and Yin, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427 https://doi.org/10.1007/s00440-011-0390-3

1350-7265

[12] Erdős, L., Yau, H.-T. and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 https://doi.org/10.1016/j.aim.2011.12.010

[13] Forrester, P.J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Phys. B* **402** 709–728. MR1236195 https://doi.org/10.1016/0550-3213(93)90126-A

[14] Ji, H.C. (2021). Regularity Properties of Free Multiplicative Convolution on the Positive Line. *Int. Math. Res. Not. IMRN* **6** 4522–4563. MR4230404 https://doi.org/10.1093/imrn/rnaa152

[15] Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991 https://doi.org/10.1007/s002200050027

[16] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[17] Knowles, A. and Yin, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

[18] Kwak, J., Lee, J.O. and Park, J. (2021). Supplement to "Extremal eigenvalues of sample covariance matrices with general population." https://doi.org/10.3150/21-BEJ1329SUPP

[19] Lee, J.O. and Schnelli, K. (2013). Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *J. Math. Phys.* **54** 103504. MR3134604 https://doi.org/10.1063/1.4823718

[20] Lee, J.O. and Schnelli, K. (2016). Extremal eigenvalues and eigenvectors of deformed Wigner matrices. *Probab. Theory Related Fields* **164** 165–241. MR3449389 https://doi.org/10.1007/s00440-014-0610-8

[21] Lee, J.O. and Schnelli, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818 https://doi.org/10.1214/16-AAP1193

[22] Mahoney, M. and Martin, C. (2019). Traditional and heavy tailed self regularization in neural network models. In *International Conference on Machine Learning* 4284–4293.

[23] Marchenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457–483.

[24] Mo, M.Y. (2012). Rank 1 real Wishart spiked model. *Comm. Pure Appl. Math.* **65** 1528–1638. MR2969495 https://doi.org/10.1002/cpa.21415

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

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

[27] 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 https://doi.org/10.1006/jmva.1995.1058

[28] Wang, D. (2012). The largest eigenvalue of real symmetric, Hermitian and Hermitian self-dual random matrix models with rank one external source, Part I. *J. Stat. Phys.* **146** 719–761. MR2916094 https://doi.org/10.1007/s10955-012-0417-x

Bernoulli **27**(4), 2021, 2766–2803
https://doi.org/10.3150/21-BEJ1330

Asymptotic results for heavy-tailed Lévy processes and their exponential functionals

WEI XU

Department of Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. E-mail: xuwei@math.hu-berlin.de

In this paper, we first provide several conditional limit theorems for Lévy processes with negative drift and regularly varying tail. Then we apply them to study the asymptotic behavior of expectations of some exponential functionals of heavy-tailed Lévy processes. As the key point, we observe that the asymptotic mainly depends on the sample paths with early arrival of large jump. Both the polynomial decay rate and the exact expression of the limit coefficients are given. As an application, we give an exact description for the extinction speed of continuous-state branching processes in heavy-tailed Lévy random environment with stable branching mechanism.

Keywords: Lévy processes; regular variation; conditional limit theorem; exponential functional; branching process; random environment; survival probability

References

[1] Afanasyev, V.I., Geiger, J., Kersting, G. and Vatutin, V.A. (2005). Criticality for branching processes in random environment. *Ann. Probab.* **33** 645–673. MR2123206 https://doi.org/10.1214/009117904000000928

[2] Asmussen, S., Foss, S. and Korshunov, D. (2003). Asymptotics for sums of random variables with local subexponential behaviour. *J. Theoret. Probab.* **16** 489–518. MR1982040 https://doi.org/10.1023/A:1023535030388

[3] Bansaye, V., Caballero, M.-E. and Méléard, S. (2019). Scaling limits of population and evolution processes in random environment. *Electron. J. Probab.* **24** 19. MR3925459 https://doi.org/10.1214/19-EJP262

[4] Bansaye, V., Pardo, J.C. and Smadi, C. (2019). Extinction rate of continuous state branching processes in critical Lévy environments. Available at arXiv:1903.06058.

[5] Bansaye, V., Pardo Millan, J.C. and Smadi, C. (2013). On the extinction of continuous state branching processes with catastrophes. *Electron. J. Probab.* **18** 106. MR3151726 https://doi.org/10.1214/EJP.v18-2774

[6] Bansaye, V. and Vatutin, V. (2017). On the survival probability for a class of subcritical branching processes in random environment. *Bernoulli* **23** 58–88. MR3556766 https://doi.org/10.3150/15-BEJ723

[7] Barker, A. and Savov, M. (2021). Bivariate Bernstein-gamma functions and moments of exponential functionals of subordinators. *Stochastic Process. Appl.* **131** 454–497. MR4165648 https://doi.org/10.1016/j.spa.2020.09.017

[8] Bertoin, J. (1996). *Lévy Processes. Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge Univ. Press. MR1406564

[9] Bertoin, J. and Yor, M. (2005). Exponential functionals of Lévy processes. *Probab. Surv.* **2** 191–212. MR2178044 https://doi.org/10.1214/154957805100000122

[10] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. ed. New York: Wiley.

[11] Böinghoff, C. and Hutzenthaler, M. (2012). Branching diffusions in random environment. *Markov Process. Related Fields* **18** 269–310. MR2985138

[12] Carmona, P., Petit, F. and Yor, M. (1997). On the distribution and asymptotic results for exponential functionals of Lévy processes. In *Exponential Functionals and Principal Values Related to Brownian Motion. Bibl. Rev. Mat. Iberoamericana* 73–130. Madrid: Rev. Mat. Iberoamericana. MR1648657

[13] Cline, D.B.H. (1986). Convolution tails, product tails and domains of attraction. *Probab. Theory Related Fields* **72** 529–557. MR0847385 https://doi.org/10.1007/BF00344720

1350-7265

[14] Comtet, A., Monthus, C. and Yor, M. (1998). Exponential functionals of Brownian motion and disordered systems. *J. Appl. Probab.* **35** 255–271. MR1641852 https://doi.org/10.1017/s0021900200014923

[15] Denisov, D., Dieker, A.B. and Shneer, V. (2008). Large deviations for random walks under subexponentiality: The big-jump domain. *Ann. Probab.* **36** 1946–1991. MR2440928 https://doi.org/10.1214/07-AOP382

[16] Denisov, D., Foss, S. and Korshunov, D. (2010). Asymptotics of randomly stopped sums in the presence of heavy tails. *Bernoulli* **16** 971–994. MR2759165 https://doi.org/10.3150/10-BEJ251

[17] Denisov, D. and Shneer, V. (2013). Asymptotics for the first passage times of Lévy processes and random walks. *J. Appl. Probab.* **50** 64–84. MR3076773 https://doi.org/10.1239/jap/1363784425

[18] Denisov, D., Vatutin, V. and Wachtel, V. (2014). Local probabilities for random walks with negative drift conditioned to stay nonnegative. *Electron. J. Probab.* **19** 88. MR3263645 https://doi.org/10.1214/EJP.v19-3426

[19] Doney, R.A. (1985). Conditional limit theorems for asymptotically stable random walks. *Z. Wahrsch. Verw. Gebiete* **70** 351–360. MR0803677 https://doi.org/10.1007/BF00534868

[20] Doney, R.A. and Jones, E.M. (2012). Conditioned random walks and Lévy processes. *Bull. Lond. Math. Soc.* **44** 139–150. MR2881332 https://doi.org/10.1112/blms/bdr084

[21] Durrett, R. (1978). Conditioned limit theorems for some null recurrent Markov processes. *Ann. Probab.* **6** 798–828. MR0503953

[22] Durrett, R. (1980). Conditioned limit theorems for random walks with negative drift. *Z. Wahrsch. Verw. Gebiete* **52** 277–287. MR0576888 https://doi.org/10.1007/BF00538892

[23] Embrechts, P., Goldie, C.M. and Veraverbeke, N. (1979). Subexponentiality and infinite divisibility. *Z. Wahrsch. Verw. Gebiete* **49** 335–347. MR0547833 https://doi.org/10.1007/BF00535504

[24] Faÿ, G., González-Arévalo, B., Mikosch, T. and Samorodnitsky, G. (2006). Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Syst.* **54** 121–140. MR2268057 https://doi.org/10.1007/s11134-006-9348-z

[25] Fu, Z. and Li, Z. (2010). Stochastic equations of non-negative processes with jumps. *Stochastic Process. Appl.* **120** 306–330. MR2584896 https://doi.org/10.1016/j.spa.2009.11.005

[26] Geiger, J., Kersting, G. and Vatutin, V.A. (2003). Limit theorems for subcritical branching processes in random environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **39** 593–620. MR1983172 https://doi.org/10.1016/S0246-0203(02)00020-1

[27] Geman, H. and Yor, M. (1993). Bessel processes, Asian options, and perpetuities. *Math. Finance* **3** 349–375.

[28] Guivarc'h, Y. and Liu, Q. (2001). Propriétés asymptotiques des processus de branchement en environnement aléatoire. *C. R. Acad. Sci. Paris Sér. I Math.* **332** 339–344. MR1821473 https://doi.org/10.1016/S0764-4442(00)01783-3

[29] He, H., Li, Z. and Xu, W. (2018). Continuous-state branching processes in Lévy random environments. *J. Theoret. Probab.* **31** 1952–1974. MR3866603 https://doi.org/10.1007/s10959-017-0765-1

[30] Hult, H. and Lindskog, F. (2006). On regular variation for infinitely divisible random vectors and additive processes. *Adv. in Appl. Probab.* **38** 134–148. MR2213968 https://doi.org/10.1239/aap/1143936144

[31] Kawazu, K. and Tanaka, H. (1993). On the maximum of a diffusion process in a drifted Brownian environment. In *Séminaire de Probabilités, XXVII. Lecture Notes in Math.* **1557** 78–85. Berlin: Springer. MR1308554 https://doi.org/10.1007/BFb0087965

[32] Kurtz, T.G. (1978). Diffusion approximations for branching processes. In *Branching Processes (Conf., Saint Hippolyte, Que., 1976). Adv. Probab. Related Topics* **5** 269–292. New York: Dekker. MR0517538

[33] Kyprianou, A.E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications. Universitext*. Berlin: Springer. MR2250061

[34] Li, Z. and Xu, W. (2018). Asymptotic results for exponential functionals of Lévy processes. *Stochastic Process. Appl.* **128** 108–131. MR3729532 https://doi.org/10.1016/j.spa.2017.04.005

[35] Palau, S. and Pardo, J.C. (2017). Continuous state branching processes in random environment: The Brownian case. *Stochastic Process. Appl.* **127** 957–994. MR3605717 https://doi.org/10.1016/j.spa.2016.07.006

[36] Palau, S. and Pardo, J.C. (2018). Branching processes in a Lévy random environment. *Acta Appl. Math.* **153** 55–79. MR3745730 https://doi.org/10.1007/s10440-017-0120-7

[37] Palau, S., Pardo, J.C. and Smadi, C. (2016). Asymptotic behaviour of exponential functionals of Lévy processes with applications to random processes in random environment. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 1235–1258. MR3592365 https://doi.org/10.30757/alea.v13-42

[38] Pardo, J.C., Patie, P. and Savov, M. (2012). A Wiener–Hopf type factorization for the exponential functional of Lévy processes. *J. Lond. Math. Soc.* (2) **86** 930–956. MR3000836 https://doi.org/10.1112/jlms/jds028
[39] Patie, P. and Savov, M. (2012). Extended factorizations of exponential functionals of Lévy processes. *Electron. J. Probab.* **17** 38. MR2928721 https://doi.org/10.1214/EJP.v17-2057
[40] Patie, P. and Savov, M. (2018). Bernstein-gamma functions and exponential functionals of Lévy processes. *Electron. J. Probab.* **23** 75. MR3835481 https://doi.org/10.1214/18-EJP202
[41] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR1739520
[42] Stephenson, R. (2018). On the exponential functional of Markov additive processes, and applications to multi-type self-similar fragmentation processes and trees. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** 1257–1292. MR3867206 https://doi.org/10.30757/alea.v15-47
[43] Vatutin, V. and Zheng, X. (2012). Subcritical branching processes in a random environment without the Cramer condition. *Stochastic Process. Appl.* **122** 2594–2609. MR2926168 https://doi.org/10.1016/j.spa.2012.04.008
[44] Véchambre, G. (2019). Exponential functionals of spectrally one-sided Lévy processes conditioned to stay positive. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 620–660. MR3949948 https://doi.org/10.1214/18-aihp892
[45] Yor, M. (1992). On some exponential functionals of Brownian motion. *Adv. in Appl. Probab.* **24** 509–531. MR1174378 https://doi.org/10.2307/1427477

Bernoulli **27**(4), 2021, 2804–2826
https://doi.org/10.3150/21-BEJ1331

Estimating the inter-occurrence time distribution from superposed renewal processes

XIAO-YANG LI[1,*], ZHI-SHENG YE[1,†] and CHENG YONG TANG[2]

[1]*Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore. E-mail:* [*]*xiaoyangli@u.nus.edu;* [†]*yez@nus.edu.sg*
[2]*Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, USA. E-mail: yongtang@temple.edu*

Superposition of renewal processes is common in practice, and it is challenging to estimate the distribution of the individual inter-occurrence time associated with the renewal process. This is because with only aggregated event history, the link between the observed recurrence times and the respective renewal processes are completely missing, rendering existing theory and methods inapplicable. In this article, we propose a nonparametric procedure to estimate the inter-occurrence time distribution by properly deconvoluting the renewal equation with the empirical renewal function. By carefully controlling the discretization errors and properly handling challenges due to implicit and non-smooth mapping via the renewal equation, our theoretical analysis establishes the consistency and asymptotic normality of the nonparametric estimators. The proposed nonparametric distribution estimators are then utilized for developing theoretically valid and computationally efficient inferences when a parametric family is assumed for the individual renewal process. Comprehensive simulations show that compared with the existing maximum likelihood method, the proposed parametric estimation procedure is much faster, and the proposed estimators are more robust to round-off errors in the observed data.

Keywords: Superposition of renewal processes; aggregate recurrence data; Nelson–Aalen estimator; minimum distance estimation

References

[1] Aalen, O.O., Borgan, Ø. and Gjessing, H.K. (2008). *Survival and Event History Analysis*: *A Process Point of View*. *Statistics for Biology and Health*. New York: Springer. MR2449233 https://doi.org/10.1007/978-0-387-68560-1
[2] Aalen, O.O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Stat. Med*. **10** 1227–1240.
[3] Angus, J.E. and Hong, X. (1996). On the rate of convergence of the Ross approximation to the renewal function. *Probab. Engrg. Inform. Sci*. **10** 207–211. MR1386635 https://doi.org/10.1017/S0269964800004277
[4] Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference*: *The Minimum Distance Approach*. *Monographs on Statistics and Applied Probability* **120**. Boca Raton, FL: CRC Press. MR2830561
[5] Baxter, L.A. (1994). Estimation from quasi life tables. *Biometrika* **81** 567–577. MR1311098 https://doi.org/10.1093/biomet/81.3.567
[6] Bender, E.A. (1974). Asymptotic methods in enumeration. *SIAM Rev.* **16** 485–515. MR0376369 https://doi.org/10.1137/1016082
[7] Bratiychuk, M.S. and Kempa, W. (2003). Application of the superposition of renewal processes to the study of batch arrival queues. *Queueing Syst*. **44** 51–67. MR1989866 https://doi.org/10.1023/A:1024042823461
[8] Chen, P., Ye, Z.-S. and Zhao, X. (2017). Minimum distance estimation for the generalized Pareto distribution. *Technometrics* **59** 528–541. MR3740969 https://doi.org/10.1080/00401706.2016.1270857
[9] Cook, R.J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. *Statistics for Biology and Health*. New York: Springer. MR3822124
[10] Cox, D.R. and Isham, V. (1980). *Point Processes*. *Monographs on Applied Probability and Statistics*. London: CRC Press. MR0598033

1350-7265

[11] Cox, D.R. and Kartsonaki, C. (2012). The fitting of complex parametric models. *Biometrika* **99** 741–747. MR2966782 https://doi.org/10.1093/biomet/ass030
[12] Cox, D.R. and Smith, W.L. (1954). On the superposition of renewal processes. *Biometrika* **41** 91–99. MR0062995 https://doi.org/10.1093/biomet/41.1-2.91
[13] Dabrowska, D.M. (1995). Estimation of transition probabilities and bootstrap in a semiparametric Markov renewal model. *J. Nonparametr. Stat.* **5** 237–259. MR1377450 https://doi.org/10.1080/10485259508832646
[14] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR1100924
[15] Frees, E.W. (1986). Nonparametric renewal function estimation. *Ann. Statist.* **14** 1366–1378. MR0868305 https://doi.org/10.1214/aos/1176350163
[16] Gill, R.D. (1981). Testing with replacement and the product limit estimator. *Ann. Statist.* **9** 853–860. MR0619288
[17] Kella, O. and Stadje, W. (2006). Superposition of renewal processes and an application to multi-server queues. *Statist. Probab. Lett.* **76** 1914–1924. MR2271187 https://doi.org/10.1016/j.spl.2006.04.041
[18] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643
[19] Lam, C.Y.T. and Lehoczky, J.P. (1991). Superposition of renewal processes. *Adv. in Appl. Probab.* **23** 64–85. MR1091093 https://doi.org/10.2307/1427512
[20] Lawless, J.F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37** 158–168. MR1333194 https://doi.org/10.2307/1269617
[21] Li, X.-Y., Ye, Z.-S. and Tang, C.Y. (2021). Supplement to "Estimating the inter-occurrence time distribution from superposed renewal processes." https://doi.org/10.3150/21-BEJ1331SUPP
[22] Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 711–730. MR1796287 https://doi.org/10.1111/1467-9868.00259
[23] Lin, F. and Fine, J.P. (2009). Pseudomartingale estimating equations for modulated renewal process models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 3–23. MR2655521 https://doi.org/10.1111/j.1467-9868.2008.00680.x
[24] Lin, F.-C., Cai, J., Fine, J.P. and Lai, H.J. (2013). Nonparametric estimation of the mean function for recurrent event data with missing event category. *Biometrika* **100** 727–740. MR3094448 https://doi.org/10.1093/biomet/ast016
[25] Lin, F.-C., Truong, Y.K. and Fine, J.P. (2013). Robust analysis of semiparametric renewal process models. *Biometrika* **100** 709–726. MR3094447 https://doi.org/10.1093/biomet/ast011
[26] Peña, E.A., Strawderman, R.L. and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *J. Amer. Statist. Assoc.* **96** 1299–1315. MR1946578 https://doi.org/10.1198/016214501753381922
[27] Politis, K. and Pitts, S.M. (2000). Nonparametric estimation in renewal theory. II. Solutions of renewal-type equations. *Ann. Statist.* **28** 88–115. MR1762905 https://doi.org/10.1214/aos/1016120366
[28] Prabhu, N.U. (2007). *Stochastic Processes*: *Basic Theory and Its Applications*. Hackensack, NJ: World Scientific Co. Pte. Ltd. MR2364170 https://doi.org/10.1142/6431
[29] Rota, G.-C. (1964). The number of partitions of a set. *Amer. Math. Monthly* **71** 498–504. MR0161805 https://doi.org/10.2307/2312585
[30] Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. New York–Auckland–Düsseldorf: McGraw-Hill. MR0385023
[31] Spitoni, C., Verduijn, M. and Putter, H. (2012). Estimation and asymptotic theory for transition probabilities in Markov renewal multi-state models. *Int. J. Biostat.* **8** Art. 23. MR2997673 https://doi.org/10.1515/1557-4679.1375
[32] Strawderman, R.L. (2005). The accelerated gap times model. *Biometrika* **92** 647–666. MR2202652 https://doi.org/10.1093/biomet/92.3.647
[33] Sun, J. and Zhao, X. (2013). *Statistical Analysis of Panel Count Data*. *Statistics for Biology and Health*. New York: Springer. MR3136574 https://doi.org/10.1007/978-1-4614-8715-9
[34] Tobias, P.A. and Trindade, D.C. (2012). *Applied Reliability*, 3rd ed. Boca Raton, FL: CRC Press. MR3058437

[35] Tortorella, M. (1990). Closed Newton–Cotes quadrature rules for Stieltjes integrals and numerical convolution of life distributions. *SIAM J. Sci. Statist. Comput.* **11** 732–748. MR1054636 https://doi.org/10.1137/0911043

[36] Tortorella, M. (1996). Life estimation from pooled discrete renewal counts. In *Lifetime Data: Models in Reliability and Survival Analysis* (*Cambridge, MA*, 1994) 331–338. Dordrecht: Kluwer Academic. MR1425912

[37] Tortorella, M. (2005). Numerical solutions of renewal-type integral equations. *INFORMS J. Comput.* **17** 66–74. MR2129807 https://doi.org/10.1287/ijoc.1030.0054

[38] Trindade, D.C. and Haugh, L.D. (1980). Estimation of the reliability of computer components from field renewal data. *Microelectron. Reliab.* **20** 205–218.

[39] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[40] Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *J. Amer. Statist. Assoc.* **94** 146–153. MR1689220 https://doi.org/10.2307/2669690

[41] Xie, M., Preuss, W. and Cui, L.-R. (2003). Error analysis of some integration procedures for renewal equation and convolution integrals. *J. Stat. Comput. Simul.* **73** 59–70. MR1959245 https://doi.org/10.1080/00949650215727

[42] Yohai, V.J. and Zamar, R.H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83** 406–413. MR0971366

[43] Zhang, W., Tian, Y., Escobar, L.A. and Meeker, W.Q. (2017). Estimating a parametric component lifetime distribution from a collection of superimposed renewal processes. *Technometrics* **59** 202–214. MR3635043 https://doi.org/10.1080/00401706.2016.1172028