BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Nine Number Four November 2023 ISSN: 1350-7265

CONTENTS

(continued)

A list of forthcoming papers can be found online at https://www.bernoullisociety.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Nine Number Four November 2023 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 29 Number 4 November 2023 Pages 2599–3469 ISI/BS

Volume 29 Number 4 November 2023 ISSN 1350-7265

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Aims and Scope
Issued four times per year, BERNOULLI is the flagship journal of the Bernoulli Society for Mathematical Statistics and Probability. The journal aims at publishing original research contributions of the highest quality in all subfields of Mathematical Statistics and Probability. The main emphasis of Bernoulli is on theoretical work, yet discussion of interesting applications in relation to the proposed methodology is also welcome.

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: https://www.bernoullisociety.org/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 https://www.bernoullisociety.org/who-is-who

---

The papers published in Bernoulli are indexed or abstracted in Mathematical Reviews (*MathSciNet*), Zentralblatt MATH (*zbMATH Open*), Science Citation Index Expanded (*Web of Science*), *SCOPUS* and *Google Scholar*.

In 2023 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **29**(4), 2023, 2599–2626
https://doi.org/10.3150/22-BEJ1554

Bootstrap inference in functional linear regression models with scalar response

HYEMIN YEON[1,a], XIONGTAO DAI[2,c] and DANIEL J. NORDMAN[1,b]

[1]*Department of Statistics, Iowa State University, Ames, IA 50011, USA,* [a]*luckyhm@iastate.edu,* [b]*dnordman@iastate.edu*
[2]*Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, 94720, USA,* [c]*xdai@berkeley.edu*

In fitting linear regression models for functional data, a complicating factor with regressors as random curves is that regression estimators have complex distributions, due to issues in bias and scaling. Bias arises because the target slope function is infinite-dimensional, while finite-sample estimators necessarily involve truncations. To approximate sampling distributions, we develop a residual bootstrap method. Despite the parametric regression problem, the bootstrap for functional data requires a development that resembles resampling for nonparametric regression with multivariate regressors. Essentially, original- and bootstrap-data estimators require coordination in the truncation levels to remove bias (akin to tuning parameter choices). The resulting bootstrap has wide applicability for constructing both confidence and prediction regions at target regressor points, and with coverage properties even holding conditionally on data regressors; the method also extends to simultaneous regions. Establishment of the bootstrap further involves generalizing, refining, and correcting a foundational central limit theorem for functional linear regression. Numerical studies verify our theory, showing that the bootstrap performs better than normal approximations, and also suggest a rule of thumb for setting the truncation levels. The bootstrap method is illustrated with an application to wheat spectrum data.

Keywords: Central limit theorem; functional data analysis; residual bootstrap; prediction; scalar-on-function regression

References

[1] Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory*. *Springer Texts in Statistics*. New York: Springer. MR2247694
[2] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103
[3] Billingsley, P. (2012). *Probability and Measure*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2893652
[4] Cai, T.T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830
[5] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45** 11–22. MR1718346 https://doi.org/10.1016/S0167-7152(99)00036-X
[6] Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711 https://doi.org/10.1007/s00440-006-0025-2
[7] Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* **19** 2627–2651. MR3160566 https://doi.org/10.3150/12-BEJ469
[8] Dezeure, R., Bühlmann, P. and Zhang, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2
[9] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. New York: Springer. MR2229687

[10] Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20** 121–145. MR1150337 https://doi.org/10.1214/aos/1176348515

[11] Freedman, D.A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228. MR0630104

[12] Gervini, D. (2015). Dynamic retrospective regression for functional data. *Technometrics* **57** 26–34. MR3318346 https://doi.org/10.1080/00401706.2013.879076

[13] González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *J. Statist. Plann. Inference* **141** 453–461. MR2719509 https://doi.org/10.1016/j.jspi.2010.06.027

[14] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587 https://doi.org/10.1214/aos/1176348651

[15] Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711. MR1165588 https://doi.org/10.1214/aos/1176348652

[16] Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269 https://doi.org/10.1214/009053606000000957

[17] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

[18] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR3379106 https://doi.org/10.1002/9781118762547

[19] Kalivas, J.H. (1997). Two data sets of near infrared spectra. *Chemom. Intell. Lab. Syst.* **37** 255–259.

[20] Khademnoe, O. and Hosseini-Nasab, S.M.E. (2016). On asymptotic distribution of prediction in functional linear regression. *Statistics* **50** 974–990. MR3522384 https://doi.org/10.1080/02331888.2015.1135926

[21] Khademnoe, O. and Hosseini-Nasab, S.M.E. (2016). On properties of percentile bootstrap confidence intervals for prediction in functional linear regression. *J. Statist. Plann. Inference* **170** 129–143. MR3427865 https://doi.org/10.1016/j.jspi.2015.10.001

[22] Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. *Texts in Statistical Science Series*. Boca Raton, FL: CRC Press. MR3793167

[23] Kreiss, J.-P. and Franke, J. (1992). Bootstrapping stationary autoregressive moving-average models. *J. Time Series Anal.* **13** 297–317. MR1173561 https://doi.org/10.1111/j.1467-9892.1992.tb00109.x

[24] Morris, J.S. (2015). Functional regression. *Annu. Rev. Stat. Appl.* **2** 321–359.

[25] Pan, L. and Politis, D.N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *J. Statist. Plann. Inference* **177** 1–27. MR3511318 https://doi.org/10.1016/j.jspi.2014.10.003

[26] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

[27] Ratcliffe, S.J., Leader, L.R. and Heller, G.Z. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Stat. Med.* **21** 1103–1114. https://doi.org/10.1002/sim.1067

[28] Schlenker, W. and Roberts, M.J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc. Natl. Acad. Sci. USA* **106** 15594–15598. https://doi.org/10.1073/pnas.0906865106

[29] Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

[30] Wu, C.-F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. MR0868303 https://doi.org/10.1214/aos/1176350142

[31] Yeon, H., Dai, X. and Nordman, D.J. (2023). Supplement to "Bootstrap inference in functional linear regression models with scalar response." https://doi.org/10.3150/22-BEJ1554SUPPA, https://doi.org/10.3150/22-BEJ1554SUPPB

Bernoulli **29**(4), 2023, 2627–2651
https://doi.org/10.3150/22-BEJ1555

Stochastic integration with respect to local time of the Brownian sheet and regularising properties of Brownian sheet paths

ANTOINE-MARIE BOGSO[1,2,a,b], MOUSTAPHA DIEYE[2,3,c,d] and OLIVIER MENOUKEU PAMEN[2,4,e,f]

[1]*Department of Mathematics, University of Yaounde I, Yaounde, Cameroon,*
[a]*antoine.bogso@facsciences-uy1.cm,* [b]*antoine@aims.edu.gh*
[2]*African Institute of Mathematical Sciences, Accra, Ghana*
[3]*École Polytechnique de Thiès, Département Tronc commun, Thiès BP A10, Sénégal,* [c]*mdieye@ept.sn,*
[d]*moustapha@aims.edu.sn*
[4]*Department of Mathematical Sciences, University of Liverpool, Liverpool, United Kingdom,*
[e]*menoukeu@liverpool.ac.uk,* [f]*olivier@aims.edu.gh*

In this work, we generalise the stochastic local time space integration introduced in (*Potential Anal.* **13** (2000) 303–328) to the case of Brownian sheet. This allows us to prove a generalised two-parameter Itô formula and derive Davie type inequalities for the Brownian sheet. Such estimates are useful to obtain regularity bounds for some averaging type operators along Brownian sheet curves.

Keywords: Brownian sheet; regularisation by noise; SDE's on the plane; path-by-path uniqueness

References

[1] Barlow, M.T. and Yor, M. (1982). Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *J. Funct. Anal.* **49** 198–229. MR0680660 https://doi.org/10.1016/0022-1236(82)90080-5

[2] Beck, L., Flandoli, F., Gubinelli, M. and Maurelli, M. (2019). Stochastic ODEs and stochastic linear PDEs with critical drift: Regularity, duality and uniqueness. *Electron. J. Probab.* **24** Paper No. 136, 72 pp. MR4040996 https://doi.org/10.1214/19-ejp379

[3] Bogso, A.-M., Dieye, M. and Menoukeu Pamen, O. (2022). Path-by-path uniqueness of multidimensional SDE's on the plane with nondecreasing coefficients. *Electron. J. Probab.* **27** Paper No. 119, 26 pp. MR4479915 https://doi.org/10.1214/22-ejp844

[4] Bogso, A.-M., Dieye, M. and Menoukeu Pamen, O. (2023). Supplement to "Stochastic integration with respect to local time of the Brownian sheet and regularising properties of Brownian sheet paths." https://doi.org/10.3150/22-BEJ1555SUPP

[5] Bogso, A.-M. and Menoukeu Pamen, O. (2022). Malliavin differentiability of solutions of hyperbolic stochastic partial differential equations with irregular drifts. Preprint. Available at arXiv:2210.04694.

[6] Cairoli, R. and Walsh, J.B. (1975). Stochastic integrals in the plane. *Acta Math.* **134** 111–183. MR0420845 https://doi.org/10.1007/BF02392100

[7] Catellier, R. and Gubinelli, M. (2016). Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.* **126** 2323–2366. MR3505229 https://doi.org/10.1016/j.spa.2016.02.002

[8] Chouk, K. and Gubinelli, M. (2015). Nonlinear PDEs with modulated dispersion I: Nonlinear Schrödinger equations. *Comm. Partial Differential Equations* **40** 2047–2081. MR3418825 https://doi.org/10.1080/03605302.2015.1073300

[9] Christ, M. (1998). Convolution, curvature, and combinatorics: A case study. *Int. Math. Res. Not.* **19** 1033–1048. MR1654767 https://doi.org/10.1155/S1073792898000610

[10] Dalang, R.C. and Walsh, J.B. (2002). Time-reversal in hyperbolic s.p.d.e.'s. *Ann. Probab.* **30** 213–252. MR1894106 https://doi.org/10.1214/aop/1020107766

[11] Davie, A.M. (2007). Uniqueness of solutions of stochastic differential equations. *Int. Math. Res. Not. IMRN* **2007** Art. ID rnm124, 26 pp. MR2377011 https://doi.org/10.1093/imrn/rnm124

[12] Eisenbaum, N. (2000). Integration with respect to local time. *Potential Anal.* **13** 303–328. MR1804175 https://doi.org/10.1023/A:1026440719120

[13] Eisenbaum, N. (2006). Local time-space stochastic calculus for Lévy processes. *Stochastic Process. Appl.* **116** 757–778. MR2218334 https://doi.org/10.1016/j.spa.2005.12.002

[14] Eisenbaum, N. (2007). Local time-space calculus for reversible semimartingales. In *Séminaire de Probabilités XL. Lecture Notes in Math.* **1899** 137–146. Berlin: Springer. MR2409003 https://doi.org/10.1007/978-3-540-71189-6_6

[15] Elworthy, K.D., Truman, A. and Zhao, H. (2007). Generalized Itô formulae and space-time Lebesgue-Stieltjes integrals of local times. In *Séminaire de Probabilités XL. Lecture Notes in Math.* **1899** 117–136. Berlin: Springer. MR2409002 https://doi.org/10.1007/978-3-540-71189-6_5

[16] Feng, C. and Zhao, H. (2007). A generalized Itô's formula in two-dimensions and stochastic Lebesgue-Stieltjes integrals. *Electron. J. Probab.* **12** 1568–1599. MR2365878 https://doi.org/10.1214/EJP.v12-468

[17] 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

[18] Galeati, L. and Gubinelli, M. (2022). Noiseless regularisation by noise. *Rev. Mat. Iberoam.* **38** 433–502. MR4404773 https://doi.org/10.4171/rmi/1280

[19] Greenleaf, A. and Seeger, A. (1998). Fourier integral operators with cusp singularities. *Amer. J. Math.* **120** 1077–1119. MR1646055

[20] Greenleaf, A., Seeger, A. and Wainger, S. (1999). On X-ray transforms for rigid line complexes and integrals over curves in $\mathbb{R}^4$. *Proc. Amer. Math. Soc.* **127** 3533–3545. MR1670367 https://doi.org/10.1090/S0002-9939-99-05379-4

[21] Imkeller, P. (1984). Itô's formula for continuous (N,d)-processes. *Z. Wahrsch. Verw. Gebiete* **65** 535–562. MR0736145 https://doi.org/10.1007/BF00531838

[22] Imkeller, P. (1984). Local times for a class of multiparameter processes. *Stochastics* **12** 143–157. MR0747782 https://doi.org/10.1080/17442508408833297

[23] Imkeller, P. (1984). Stochastic analysis and local times for (N,d)-Wiener process. *Ann. Inst. Henri Poincaré Probab. Stat.* **20** 75–101. MR0740251

[24] Khoshnevisan, D. (2002). *Multiparameter Processes: An Introduction to Random Fields*. *Springer Monographs in Mathematics*. New York: Springer. MR1914748 https://doi.org/10.1007/b97363

[25] Littman, W. (1973). L^p–L^q estimates for singular integral operators. In *Proc. Symp. Pure Math.* **23** 479–481.

[26] Nualart, D. (1984). Une formule d'Itô pour les martingales continues à deux indices et quelques applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **20** 251–275. MR0762858

[27] Nualart, D. and Yeh, J. (1989). Existence and uniqueness of a strong solution to stochastic differential equations in the plane with stochastic boundary process. *J. Multivariate Anal.* **28** 149–171. MR0996988 https://doi.org/10.1016/0047-259X(89)90101-2

[28] Oberlin, D.M. (1987). Convolution estimates for some measures on curves. *Proc. Amer. Math. Soc.* **99** 56–60. MR0866429 https://doi.org/10.2307/2046270

[29] Oberlin, D.M. (1999). A convolution estimate for a measure on a curve in $\mathbb{R}^4$. II. *Proc. Amer. Math. Soc.* **127** 217–221. MR1476381 https://doi.org/10.1090/S0002-9939-99-04690-0

[30] Peskir, G. (2005). A change-of-variable formula with local time on curves. *J. Theoret. Probab.* **18** 499–535. MR2167640 https://doi.org/10.1007/s10959-005-3517-6

[31] 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

[32] Sanz, M. (1988). Local time for two-parameter continuous martingales with respect to the quadratic variation. *Ann. Probab.* **16** 778–792. MR0929078

[33] Tao, T. and Wright, J. (2003). L^p improving bounds for averages along curves. *J. Amer. Math. Soc.* **16** 605–638. MR1969206 https://doi.org/10.1090/S0894-0347-03-00420-X

[34] Walsh, J.B. (1978). The local time of the Brownian sheet. *Astérisque* **52** 47–61.

Bernoulli **29**(4), 2023, 2652–2690
https://doi.org/10.3150/22-BEJ1556

Efficient and consistent model selection procedures for time series

JEAN-MARC BARDET[1,a], KAMILA KARE[1,b] and WILLIAM KENGNE[2,c]

[1]*University Paris 1 Panthéon-Sorbonne, SAMM, Franc,* [a]*bardet@univ-paris1.fr,* [b]*kamilakare@gmail.com*
[2]*CY Cergy Paris Université, THEMA, France,* [c]*william.kengne@u-cergy.fr*

This paper studies the problem of model selection in a large class of causal time series models that includes ARMA or AR (∞) processes as well as GARCH or ARCH (∞), APARCH, ARMA-GARCH - and many other processes. First, we study the asymptotic behavior of the ideal penalty that minimizes the risk defined from a quasi-likelihood estimation among a finite family of models containing the true model. We then establish general conditions on the penalty term to obtain properties of consistency and efficiency. In particular, we prove that consistent model selection criteria outperform the classical AIC criterion in terms of efficiency. Finally, we derive the usual BIC criterion from a Bayesian approach and, retaining all second-order terms of the Laplace approximation, a data-driven criterion, which we call KC'. Monte Carlo experiments illustrate the asymptotic results obtained and show that the KC' criterion performs better than the AIC and BIC criteria in terms of consistency and efficiency.

Keywords: Model selection; affine causal processes; efficiency; data-driven criterion; consistency

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)* 267–281. Budapest: Akadémiai Kiadó. MR0483125

[2] Arlot, S. and Massart, P. (2009). Data-driven calibration of penalties for least-squares regression. *J. Mach. Learn. Res.* **10** 245–279.

[3] Bardet, J.-M., Boularouk, Y. and Djaballah, K. (2017). Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes. *Electron. J. Stat.* **11** 452–479. MR3619313 https://doi.org/10.1214/17-EJS1241

[4] Bardet, J.-M., Kamila, K. and Kengne, W. (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Electron. J. Stat.* **14** 2009–2052. MR4091861 https://doi.org/10.1214/20-EJS1709

[5] Bardet, J.-M. and Wintenberger, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* **37** 2730–2759. MR2541445 https://doi.org/10.1214/08-AOS674

[6] Bardet, J.M., Doukhan, P. and Wintenberger, O. (2020). Contrast estimation of general locally stationary processes using coupling. Preprint arXiv:2005.07397.

[7] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **52** 345–370. MR0914460 https://doi.org/10.1007/BF02294361

[8] Cavanaugh, J.E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statist. Probab. Lett.* **33** 201–208. MR1458291 https://doi.org/10.1016/S0167-7152(96)00128-9

[9] Chen, C.F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47** 540–546. MR0844485

[10] Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Process. Appl.* **106** 63–80. MR1983043 https://doi.org/10.1016/S0304-4149(03)00040-1

[11] Dedecker, J., Doukhan, P., Lang, G., León, J.R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. *Lecture Notes in Statistics* **190**. New York: Springer. MR2338725

1350-7265

[12] Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *J. Theoret. Probab.* **17** 861–885. MR2105738 https://doi.org/10.1007/s10959-004-0578-x

[13] Ding, J., Tarokh, V. and Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Process. Mag.* **35** 16–34.

[14] Ding, Z., Granger, C. and Engle, R.F. (1993). A long memory property of stock market returns and a new model. *J. Empir. Finance* **1** 83–106.

[15] Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* **118** 1997–2013. MR2462284 https://doi.org/10.1016/j.spa.2007.12.004

[16] Findley, D.F. and Wei, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *J. Multivariate Anal.* **83** 415–450. MR1945962 https://doi.org/10.1006/jmva.2001.2063

[17] Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: Wiley. MR3186556 https://doi.org/10.1002/9780470670057

[18] Hannan, E.J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081. MR0585705

[19] Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B* **41** 190–195. MR0547244

[20] Hsu, H.-L., Ing, C.-K. and Tong, H. (2019). On model selection from a finite family of possibly misspecified time series models. *Ann. Statist.* **47** 1061–1087. MR3909960 https://doi.org/10.1214/18-AOS1706

[21] Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76** 297–307. MR1016020 https://doi.org/10.1093/biomet/76.2.297

[22] Ing, C.-K., Sin, C. and Yu, S.-H. (2012). Model selection for integrated autoregressive processes of infinite order. *J. Multivariate Anal.* **106** 57–71. MR2887680 https://doi.org/10.1016/j.jmva.2011.10.008

[23] Ing, C.-K. and Wei, C.-Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474. MR2211091 https://doi.org/10.1214/009053605000000525

[24] Karagrigoriou, A. (1997). Asymptotic efficiency of the order selection of a non-Gaussian AR process. *Statist. Sinica* **7** 407–423. MR1466689

[25] Kashyap, R.L. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Trans. Pattern Anal. Mach. Intell.* **4** 99–104.

[26] Kass, R.E., Tierney, L. and Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's method. In *Essays in Honor of George Barnard* (S. Geisser, J.S. Hodges, S.J. Press and A. Zellner, eds.) 473–188. Amsterdam: North-Holland.

[27] Kengne, W. (2021). Strongly consistent model selection for general causal time series. *Statist. Probab. Lett.* **171** Paper No. 109000, 7. MR4207366 https://doi.org/10.1016/j.spl.2020.109000

[28] Li, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15** 958–975. MR0902239 https://doi.org/10.1214/aos/1176350486

[29] Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J. Econometrics* **140** 849–873. MR2408929 https://doi.org/10.1016/j.jeconom.2006.07.016

[30] Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19** 280–310. MR1966031 https://doi.org/10.1017/S0266466603192092

[31] Lv, J. and Liu, J.S. (2014). Model selection principles in misspecified models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 141–167. MR3153937 https://doi.org/10.1111/rssb.12023

[32] Mallows, C.L. (1973). Some comments on Cp. *Technometrics* **15** 661–675.

[33] Massart, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math.* **1896**. Berlin: Springer. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879

[34] McQuarrie, A.D.R. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. River Edge, NJ: World Scientific Co., Inc. MR1641582 https://doi.org/10.1142/9789812385451

[35] Papangelou, F. (1994). On a distributional bound arising in autoregressive model fitting. *J. Appl. Probab.* **31** 401–408. MR1274796 https://doi.org/10.1017/s0021900200044922

[36] Rao, C.R. and Wu, Y. (2001). On model selection. In *Model Selection*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **38** 1–64. Beachwood, OH: IMS. With discussion by Sadanori Konishi and Rahul Mukerjee and a rejoinder by the authors. MR2000751 https://doi.org/10.1214/lnms/1215540960

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

[38] Sclove, S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika* **52** 333–343.

[39] Shao, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. With comments and a rejoinder by the author. MR1466682

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

Bernoulli **29**(4), 2023, 2691–2718
https://doi.org/10.3150/22-BEJ1557

Element-wise estimation error of generalized Fused Lasso

TENG ZHANG[1,a] and SABYASACHI CHATTERJEE[2,b]

[1]*Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA,* [a]*teng.zhang@ucf.edu*
[2]*Department of Statistics, University of Illinois at Urbana Champaign, Champaign, IL 61820, USA,*
[b]*sc1706@illinois.edu*

The main result of this article is that we obtain an elementwise error bound for the Fused Lasso estimator for any general convex loss function ρ. We then focus on the special cases when either ρ is the square loss function (for mean regression) or is the quantile loss function (for quantile regression) for which we derive new pointwise error bounds. Even though error bounds for the usual Fused Lasso estimator and its quantile version have been studied before; our bound appears to be new. This is because all previous works bound a global loss function like the sum of squared error, or a sum of huber losses in the case of quantile regression in Padilla and Chatterjee (*Biometrika* **109** (2022) 751–768). Clearly, element wise bounds are stronger than global loss error bounds as it reveals how the loss behaves locally at each point. Our element wise error bound also has a clean and explicit dependence on the tuning parameter λ which informs the user of a good choice of λ. In addition, our bound is nonasymptotic with explicit constants and is able to recover almost all the known results for Fused Lasso (both mean and quantile regression) with additional improvements in some cases.

Keywords: Adaptive risk bounds; generalized Fused Lasso; law of iterated logarithm; nonasymptotic risk bounds; nonparametric quantile regression; pointwise risk bounds; total variation denoising

References

Barbero, Á. and Sra, S. (2018). Modular proximal optimization for multidimensional total-variation regularization. *J. Mach. Learn. Res.* **19** Paper No. 56, 82 pp. MR3899758

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841 https://doi.org/10.1214/10-AOS827

Chatterjee, S. and Goswami, S. (2021). New risk bounds for 2D total variation denoising. *IEEE Trans. Inf. Theory* **67** 4060–4091. MR4289366 https://doi.org/10.1109/TIT.2021.3059657

Chatterjee, S. and Sen, S. (2021). Regret Minimization in Isotonic, Heavy-Tailed Contextual Bandits via Adaptive Confidence Bands. Preprint. Available at arXiv:2110.10245.

Dalalyan, A.S., Hebiri, M. and Lederer, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756

Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

Guntuboyina, A., Lieu, D., Chatterjee, S. and Sen, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.* **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799

Hallac, D., Leskovec, J. and Boyd, S. (2015). Network lasso: Clustering and optimization in large graphs. In *KDD: Proceedings. International Conference on Knowledge Discovery & Data Mining 2015* 387–396. https://doi.org/10.1145/2783258.2783313

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565 https://doi.org/10.1198/jasa.2010.tm09181

He, X. and Shi, P.D. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparametr. Stat.* **3** 299–308. MR1291551 https://doi.org/10.1080/10485259408832589

Hütter, J.-C. and Rigollet, P. (2016). Optimal rates for total variation denoising. In *29th Annual Conference on Learning Theory* (V. Feldman, A. Rakhlin and O. Shamir, eds.). *Proc. Mach. Learn. Res.* **49** 1115–1146. New York, New York, USA: PMLR, Columbia Univ.

Jamieson, K., Malloy, M., Nowak, R. and Bubeck, S. (2014). lil' UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory* (M.F. Balcan, V. Feldman and C. Szepesvári, eds.). *Proc. Mach. Learn. Res.* **35** 423–439. Barcelona, Spain: PMLR.

Kim, S.-J., Koh, K., Boyd, S. and Gorinevsky, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

Lin, K., Sharpnack, J., Rinaldo, A. and Tibshirani, R.J. (2016). Approximate Recovery in Changepoint Problems, from ℓ_2 Estimation Error Rates. Preprint. Available at arXiv:1606.06746.

Lin, K., Sharpnack, J., Rinaldo, A. and Tibshirani, R.J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. *NIPS'17* 6887–6896. USA: Curran Associates Inc.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

Ortelli, F. and van de Geer, S. (2018). On the total variation regularized estimator over a class of tree graphs. *Electron. J. Stat.* **12** 4517–4570. MR3892703 https://doi.org/10.1214/18-ejs1519

Ortelli, F. and van de Geer, S. (2021). Prediction bounds for higher order total variation regularized least squares. *Ann. Statist.* **49** 2755–2773. MR4338382 https://doi.org/10.1214/21-aos2054

Padilla, O.H.M. and Chatterjee, S. (2022). Risk bounds for quantile trend filtering. *Biometrika* **109** 751–768. MR4472846 https://doi.org/10.1093/biomet/asab045

Rigollet, P. and Hütter, J.-C. (2017). High Dimensional Statistics. Lecture notes for course 18S997. Available at https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf.

Rinaldo, A. (2009). Properties and refinements of the fused lasso. *Ann. Statist.* **37** 2922–2952. MR2541451 https://doi.org/10.1214/08-AOS665

Rudin, L.I., Osher, S. and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D, Nonlinear Phenom.* **60** 259–268. MR3363401 https://doi.org/10.1016/0167-2789(92)90242-F

Steidl, G., Didas, S. and Neumann, J. (2006). Splines in higher order TV regularization. *Int. J. Comput. Vis.* **70** 241–255.

Tansey, W. and Scott, J.G. (2015). A fast and flexible algorithm for the graph-fused lasso. Preprint. Available at arXiv:1505.06475.

Tibshirani, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

Tibshirani, R.J. (2020). Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. Preprint. Available at arXiv:2003.03886.

Tibshirani, R. and Wang, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9** 18–29. https://doi.org/10.1093/biostatistics/kxm013

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. MR2136641 https://doi.org/10.1111/j.1467-9868.2005.00490.x

Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge: Cambridge Univ. Press. MR3967104 https://doi.org/10.1017/9781108627771

Wang, F., Madrid, O., Yu, Y. and Rinaldo, A. (2022). Denoising and change point localisation in piecewise-constant high-dimensional regression coefficients. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics* (G. Camps-Valls, F.J.R. Ruiz and I. Valera, eds.). *Proc. Mach. Learn. Res.* **151** 4309–4338. PMLR.

Bernoulli **29**(4), 2023, 2719–2744
https://doi.org/10.3150/22-BEJ1558

On the mean perimeter density of inhomogeneous random closed sets

ELENA VILLA[a]

Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy, [a]*elena.villa@unimi.it*

The computation of the mean perimeter density via the notion of mean covariogram for non-stationary Boolean models has been proposed as further work in Galerne (*Image Anal. Stereol.* **30** (2011) 39–51). We address this issue by considering here more general germ-grain models. Furthermore, we discuss similarities and differences with respect to the computation of the mean boundary density by means of the outer Minkowski content notion.

Keywords: Germ-grain model; mean density; Minkowski content; perimeter; variation

References

Ambrosio, L., Colesanti, A. and Villa, E. (2008). Outer Minkowski content for some classes of closed sets. *Math. Ann.* **342** 727–748. MR2443761 https://doi.org/10.1007/s00208-008-0254-z

Ambrosio, L., Fusco, N. and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. *Oxford Mathematical Monographs*. Oxford University Press, New York: The Clarendon Press. MR1857292

Baddeley, A., Bárány, I., Schneider, R. and Weil, W. (2007). *Stochastic Geometry*. *Lecture Notes in Math.* **1892**. Berlin: Springer. Lectures given at the C.I.M.E. Summer School held in Martina Franca, September 13–18, 2004, Edited by W. Weil. MR2327289

Camerlenghi, F. and Villa, E. (2015). Optimal bandwidth of the "Minkowski content"-based estimator of the mean density of random closed sets: Theoretical results and numerical experiments. *J. Math. Imaging Vision* **53** 264–287. MR3397099 https://doi.org/10.1007/s10851-015-0576-x

Coeurjolly, J.-F., Møller, J. and Waagepetersen, R. (2017). A tutorial on Palm distributions for spatial point processes. *Int. Stat. Rev.* **85** 404–420. MR3723609 https://doi.org/10.1111/insr.12205

Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. MR1950431

Daley, D.J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. MR2371524 https://doi.org/10.1007/978-0-387-49835-5

Evans, L.C. and Gariepy, R.F. (1992). *Measure Theory and Fine Properties of Functions*. *Studies in Advanced Mathematics*. Boca Raton, FL: CRC Press. MR1158660

Federer, H. (1969). *Geometric Measure Theory*. *Die Grundlehren der Mathematischen Wissenschaften* **153**. New York: Springer New York, Inc. MR0257325

Galerne, B. (2011). Computation of the perimeter of measurable sets via their covariogram. Applications to random sets. *Image Anal. Stereol.* **30** 39–51. MR2816305 https://doi.org/10.5566/ias.v30.p39-51

Galerne, B. (2016). Random fields of bounded variation and computation of their variation intensity. *Adv. in Appl. Probab.* **48** 947–971. MR3595761 https://doi.org/10.1017/apr.2016.60

Galerne, B. and Lachièze-Rey, R. (2015). Random measurable sets and covariogram realizability problems. *Adv. in Appl. Probab.* **47** 611–639. MR3406600 https://doi.org/10.1239/aap/1444308874

Heijmans, H.J.A.M. (1995). Mathematical morphology: A modern approach in image processing based on algebra and geometry. *SIAM Rev.* **37** 1–36. MR1327714 https://doi.org/10.1137/1037001

Hug, D. and Last, G. (2000). On support measures in Minkowski spaces and contact distributions in stochastic geometry. *Ann. Probab.* **28** 796–850. MR1782274 https://doi.org/10.1214/aop/1019160261

1350-7265

Kiderlen, M. and Rataj, J. (2018). Dilation volumes of sets of finite perimeter. *Adv. in Appl. Probab.* **50** 1095–1118. MR3881111 https://doi.org/10.1017/apr.2018.52

Maggi, F. (2012). *Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory*. *Cambridge Studies in Advanced Mathematics* **135**. Cambridge: Cambridge Univ. Press. MR2976521 https://doi.org/10.1017/CBO9781139108133

Matheron, G. (1975). *Random Sets and Integral Geometry*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0385969

Molchanov, I. (2005). *Theory of Random Sets*. *Probability and Its Applications (New York)*. London: Springer London, Ltd. MR2132405

Rataj, J. (2015). Random sets of finite perimeter. *Math. Nachr.* **288** 1047–1056. MR3357549 https://doi.org/10.1002/mana.201300341

Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry*. *Probability and Its Applications (New York)*. Berlin: Springer. MR2455326 https://doi.org/10.1007/978-3-540-78859-1

Stoyan, D., Kendall, W.S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications*. Chichester: Wiley. MR0895588

Villa, E. (2009). On the outer Minkowski content of sets. *Ann. Mat. Pura Appl. (4)* **188** 619–630. MR2533959 https://doi.org/10.1007/s10231-008-0093-2

Villa, E. (2010). Mean densities and spherical contact distribution function of inhomogeneous Boolean models. *Stoch. Anal. Appl.* **28** 480–504. MR2739571 https://doi.org/10.1080/07362991003708812

Villa, E. (2014). On the local approximation of mean densities of random closed sets. *Bernoulli* **20** 1–27. MR3160572 https://doi.org/10.3150/12-BEJ474

Zähle, M. (1982). Random processes of Hausdorff rectifiable closed sets. *Math. Nachr.* **108** 49–72. MR0695116 https://doi.org/10.1002/mana.19821080105

Bernoulli **29**(4), 2023, 2745–2770
https://doi.org/10.3150/22-BEJ1559

Invariance principle for fragmentation processes derived from conditioned stable Galton–Watson trees

GABRIEL BERZUNZA OJEDA[1,a] and CECILIA HOLMGREN[2,b]

[1]*Department of Mathematical Sciences, University of Liverpool, United Kingdom,*
[a]*gabriel.berzunza-ojeda@liverpool.ac.uk*
[2]*Department of Mathematics, Uppsala University, Sweden,* [b]*cecilia.holmgren@math.uu.se*

Aldous, Evans and Pitman (1998) studied the behavior of the fragmentation process derived from deleting the edges of a uniform random tree on n labelled vertices. In particular, they showed that, after proper rescaling, the above fragmentation process converges as $n \to \infty$ to the fragmentation process of the Brownian CRT obtained by cutting-down the Brownian CRT along its skeleton in a Poisson manner.

In this work, we continue the above investigation and study the fragmentation process obtained by deleting randomly chosen edges from a critical Galton–Watson tree $\mathbf{t}_n$ conditioned on having n vertices, whose offspring distribution belongs to the domain of attraction of a stable law of index $\alpha \in (1,2]$. Our main results establish that, after rescaling, the fragmentation process of $\mathbf{t}_n$ converges as $n \to \infty$ to the fragmentation process obtained by cutting-down at a rate proportional to the length measure on the skeleton of an α-stable Lévy tree. We further show that the latter can be constructed by considering the partitions of the unit interval induced by the normalized α-stable Lévy excursion with a deterministic drift studied by Miermont (2001). This extends the result of Bertoin (2000) on the fragmentation process of the Brownian CRT.

Keywords: Additive coalescent; fragmentation; Galton–Watson trees; spectrally positive stable Lévy processes; stable Lévy tree; Prim's algorithm

References

[1] Abraham, R., Delmas, J.-F. and Hoscheit, P. (2013). A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** 14. MR3035742 https://doi.org/10.1214/EJP.v18-2116
[2] Abraham, R. and Serlet, L. (2002). Poisson snake and fragmentation. *Electron. J. Probab.* **7** 17. MR1943890 https://doi.org/10.1214/EJP.v7-116
[3] Aldous, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226
[4] Aldous, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** 812–854. MR1434128 https://doi.org/10.1214/aop/1024404421
[5] Aldous, D. and Pitman, J. (1998). The standard additive coalescent. *Ann. Probab.* **26** 1703–1726. MR1675063 https://doi.org/10.1214/aop/1022855879
[6] Aldous, D. and Pitman, J. (2000). Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields* **118** 455–482. MR1808372 https://doi.org/10.1007/PL00008751
[7] Bertoin, J. (1996). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge Univ. Press. MR1406564
[8] Bertoin, J. (2000). A fragmentation process connected to Brownian motion. *Probab. Theory Related Fields* **117** 289–301. MR1771665 https://doi.org/10.1007/s004400050008
[9] Bertoin, J. (2001). Eternal additive coalescents and certain bridges with exchangeable increments. *Ann. Probab.* **29** 344–360. MR1825153 https://doi.org/10.1214/aop/1008956333

1350-7265

[10] Bertoin, J. (2002). Self-similar fragmentations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 319–340. MR1899456 https://doi.org/10.1016/S0246-0203(00)01073-6

[11] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. *Cambridge Studies in Advanced Mathematics* **102**. Cambridge: Cambridge Univ. Press. MR2253162 https://doi.org/10.1017/CBO9780511617768

[12] Bertoin, J. (2013). Almost giant clusters for percolation on large trees with logarithmic heights. *J. Appl. Probab.* **50** 603–611. MR3102504 https://doi.org/10.1239/jap/1378401225

[13] Bertoin, J. and Miermont, G. (2006). Asymptotics in Knuth's parking problem for caravans. *Random Structures Algorithms* **29** 38–55. MR2238028 https://doi.org/10.1002/rsa.20092

[14] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1700749 https://doi.org/10.1002/9780470316962

[15] Broutin, N. and Marckert, J.-F. (2016). A new encoding of coalescent processes: Applications to the additive and multiplicative cases. *Probab. Theory Related Fields* **166** 515–552. MR3547745 https://doi.org/10.1007/s00440-015-0665-1

[16] Chassaing, P. and Janson, S. (2001). A Vervaat-like path transformation for the reflected Brownian bridge conditioned on its local time at 0. *Ann. Probab.* **29** 1755–1779. MR1880241 https://doi.org/10.1214/aop/1015345771

[17] Chassaing, P. and Louchard, G. (2002). Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms* **21** 76–119. MR1913079 https://doi.org/10.1002/rsa.10039

[18] Chaumont, L. (1997). Excursion normalisée, méandre et pont pour les processus de Lévy stables. *Bull. Sci. Math.* **121** 377–403. MR1465814

[19] Duquesne, T. (2003). A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.* **31** 996–1027. MR1964956 https://doi.org/10.1214/aop/1048516543

[20] Duquesne, T. and Le Gall, J.-F. (2002). Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** vi+147. MR1954248

[21] Duquesne, T. and Le Gall, J.-F. (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131** 553–603. MR2147221 https://doi.org/10.1007/s00440-004-0385-4

[22] Evans, S.N. and Pitman, J. (1998). Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 339–383. MR1625867 https://doi.org/10.1016/S0246-0203(98)80015-0

[23] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. New York: Wiley. MR0270403

[24] Haas, B. and Miermont, G. (2012). Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.* **40** 2589–2666. MR3050512 https://doi.org/10.1214/11-AOP686

[25] 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

[26] 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

[27] Kingman, J.F.C. (1978). The representation of partition structures. *J. Lond. Math. Soc. (2)* **18** 374–380. MR0509954 https://doi.org/10.1112/jlms/s2-18.2.374

[28] Le Gall, J.-F. (2005). Random trees and applications. *Probab. Surv.* **2** 245–311. MR2203728 https://doi.org/10.1214/154957805100000140

[29] Lushnikov, A.A. (1978). Some new aspects of coagulation theory. *Izv. Akad. Nauk SSSR, Ser. Fiz. Atmosfer. I Okeana* **14** 738–743.

[30] Marckert, J.-F. and Wang, M. (2019). A new combinatorial representation of the additive coalescent. *Random Structures Algorithms* **54** 340–370. MR3912100 https://doi.org/10.1002/rsa.20775

[31] Marcus, A.H. (1968). Stochastic coalescence. *Technometrics* **10** 133–143. MR0223151 https://doi.org/10.2307/1266230

[32] Miermont, G. (2001). Ordered additive coalescent and fragmentations associated to Levy processes with no positive jumps. *Electron. J. Probab.* **6** 14. MR1844511 https://doi.org/10.1214/EJP.v6-87

[33] Miermont, G. (2005). Self-similar fragmentations derived from the stable tree. II. Splitting at nodes. *Probab. Theory Related Fields* **131** 341–375. MR2123249 https://doi.org/10.1007/s00440-004-0373-8

[34] Miermont, G. (2009). Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)* **42** 725–781. MR2571957 https://doi.org/10.24033/asens.2108

[35] Neveu, J. (1986). Arbres et processus de Galton-Watson. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 199–207. MR0850756

[36] Otter, R. (1949). The multiplicative process. *Ann. Math. Stat.* **20** 206–224. MR0030716 https://doi.org/10.1214/aoms/1177730031

[37] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448 https://doi.org/10.1007/BF01205234

[38] Pitman, J. (1999). Coalescent random forests. *J. Combin. Theory Ser. A* **85** 165–193. MR1673928 https://doi.org/10.1006/jcta.1998.2919

[39] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129 https://doi.org/10.1214/aop/1024404422

[40] Prim, R.C. (1957). Shortest connection networks and some generalizations. *Bell Syst. Tech. J.* **36** 1389–1401.

[41] Takács, L. (1967). On combinatorial methods in the theory of stochastic processes. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1* 431–447. Berkeley, CA: Univ. California Press. MR0214129

[42] Thévenin, P. (2021). A geometric representation of fragmentation processes on stable trees. *Ann. Probab.* **49** 2416–2476. MR4317709 https://doi.org/10.1214/21-aop1512

[43] Vervaat, W. (1979). A relation between Brownian bridge and Brownian excursion. *Ann. Probab.* **7** 143–149. MR0515820

Bernoulli **29**(4), 2023, 2771–2796
https://doi.org/10.3150/22-BEJ1561

Extremal clustering and cluster counting for spatial random fields

ANDERS RØNN-NIELSEN[a] and MADS STEHR[b]

Department of Finance, Center for Statistics, Copenhagen Business School, Frederiksberg, Denmark,
[a]*aro.fi@cbs.dk,* [b]*mast.fi@cbs.dk*

We consider a stationary random field indexed by an increasing sequence of subsets of $\mathbb{Z}^d$ obeying a very broad geometrical assumption on how the sequence expands. Under certain mixing and local conditions, we show how the tail distribution of the individual variables relates to the tail behavior of the maximum of the field over the index sets in the limit as the index sets expand.

In a framework where we let the increasing index sets be scalar multiplications of a fixed set C, potentially with different scalars in different directions, we use two cluster definitions to define associated cluster counting point processes on the rescaled index set C; one cluster definition divides the index set into more and more boxes and counts a box as a cluster if it contains an extremal observation. The other cluster definition that is more intuitive considers extremal points to be in the same cluster, if they are close in distance. We show that both cluster point processes converge to a Poisson point process on C. Additionally, we find a limit of the mean cluster size. Finally, we pay special attention to the case without clusters.

Keywords: Extreme value theory; spatial models; random fields; intrinsic volumes; extremal index; cluster counting process; limit theorems

References

[1] Chernick, M.R., Hsing, T. and McCormick, W.P. (1991). Calculating the extremal index for a class of stationary sequences. *Adv. in Appl. Probab.* **23** 835–850. MR1133731 https://doi.org/10.2307/1427679

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

[3] Ferreira, H. and Pereira, L. (2012). Point processes of exceedances by random fields. *J. Statist. Plann. Inference* **142** 773–779. MR2853583 https://doi.org/10.1016/j.jspi.2011.09.012

[4] Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78** 97–112. MR0940870 https://doi.org/10.1007/BF00718038

[5] Jakubowski, A. and Soja-Kukieła, N. (2019). Managing local dependencies in asymptotic theory for maxima of stationary random fields. *Extremes* **22** 293–315. MR3949047 https://doi.org/10.1007/s10687-018-0336-6

[6] Kallenberg, O. (2017). *Random Measures, Theory and Applications*. *Probability Theory and Stochastic Modelling* **77**. Cham: Springer. MR3642325 https://doi.org/10.1007/978-3-319-41598-7

[7] Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. MR0722133 https://doi.org/10.1007/BF00532484

[8] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. *Springer Series in Statistics*. New York-Berlin: Springer. MR0691492

[9] Ling, C. (2019). Extremes of stationary random fields on a lattice. *Extremes* **22** 391–411. MR3984355 https://doi.org/10.1007/s10687-019-00349-z

[10] O'Brien, G.L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.* **15** 281–291. MR0877604

[11] Pereira, L. and Ferreira, H. (2006). Limiting crossing probabilities of random fields. *J. Appl. Probab.* **43** 884–891. MR2274809 https://doi.org/10.1239/jap/1158784955

[12] Pereira, L., Martins, A.P. and Ferreira, H. (2017). Clustering of high values in random fields. *Extremes* **20** 807–838. MR3737386 https://doi.org/10.1007/s10687-017-0291-7

[13] Rønn-Nielsen, A. and Stehr, M. (2022). Extremes of Lévy-driven spatial random fields with regularly varying Lévy measure. *Stochastic Process. Appl.* **150** 19–49. MR4419568 https://doi.org/10.1016/j.spa.2022.04.007

[14] Rønn-Nielsen, A. and Stehr, M. (2023). Supplement to "Extremal clustering and cluster counting for spatial random fields." https://doi.org/10.3150/22-BEJ1561SUPP

[15] Schneider, R. (1993). *Convex Bodies: The Brunn-Minkowski Theory*. *Encyclopedia of Mathematics and Its Applications* **44**. Cambridge: Cambridge Univ. Press. MR1216521 https://doi.org/10.1017/CBO9780511526282

[16] Soja-Kukieła, N. (2019). On maxima of stationary fields. *J. Appl. Probab.* **56** 1217–1230. MR4041457 https://doi.org/10.1017/jpr.2019.69

[17] Stehr, M. and Rønn-Nielsen, A. (2021). Extreme value theory for spatial random fields—with application to a Lévy-driven field. *Extremes* **24** 753–795. MR4331056 https://doi.org/10.1007/s10687-021-00415-5

[18] Stehr, M. and Rønn-Nielsen, A. (2022). Extremes of subexponential Lévy-driven random fields in the Gumbel domain of attraction. *Extremes* **25** 79–105. MR4376585 https://doi.org/10.1007/s10687-021-00428-0

Bernoulli **29**(4), 2023, 2797–2827
https://doi.org/10.3150/22-BEJ1562

Concentration bounds for the empirical angular measure with statistical learning applications

STÉPHAN CLÉMENÇON[1,a], HAMID JALALZAI[1,b], STÉPHANE LHAUT[2,c], ANNE SABOURIN[3,e] and JOHAN SEGERS[2,d]

[1] *LTCI, Télécom Paris, Palaiseau, France,* [a] *stephan.clemencon@telecom-paris.fr,*
[b] *hamid.jalalzai@telecom-paris.fr*
[2] *LIDAM/ISBA, UCLouvain, Louvain-la-Neuve, Belgium,* [c] *stephane.lhaut@uclouvain.be,*
[d] *johan.segers@uclouvain.be*
[3] *Université Paris Cité, CNRS, MAP5, F-75006 Paris, France,* [e] *anne.sabourin@u-paris.fr*

The angular measure on the unit sphere characterizes the first-order dependence structure of the components of a random vector in extreme regions and is defined in terms of standardized margins. Its statistical recovery is an important step in learning problems involving observations far away from the center. In the common situation that the components of the vector have different distributions, the rank transformation offers a convenient and robust way of standardizing data in order to build an empirical version of the angular measure based on the most extreme observations. However, the study of the sampling distribution of the resulting empirical angular measure is challenging. It is the purpose of the paper to establish finite-sample bounds for the maximal deviations between the empirical and true angular measures, uniformly over classes of Borel sets of controlled combinatorial complexity. The bounds are valid with high probability and, up to logarithmic factors, scale as the square root of the effective sample size. The bounds are applied to provide performance guarantees for two statistical learning procedures tailored to extreme regions of the input space and built upon the empirical angular measure: binary classification in extreme regions through empirical risk minimization and unsupervised anomaly detection through minimum-volume sets of the sphere.

Keywords: Angular measure; classification; concentration inequality; extreme value analysis; minimum-volume sets; ranks

References

[1] Beirlant, J., Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2016). Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* **143** 453–466. MR3431445 https://doi.org/10.1016/j.jmva.2015.10.006

[2] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. With contributions from Daniel De Waal and Chris Ferro. MR2108013 https://doi.org/10.1002/0470012382

[3] Blanchard, G., Lee, G. and Scott, C. (2010). Semi-supervised novelty detection. *J. Mach. Learn. Res.* **11** 2973–3009. MR2746544

[4] Boldi, M.-O. and Davison, A.C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 217–229. MR2325273 https://doi.org/10.1111/j.1467-9868.2007.00585.x

[5] Boucheron, S. and Thomas, M. (2015). Tail index estimation, concentration and adaptivity. *Electron. J. Stat.* **9** 2751–2792. MR3435810 https://doi.org/10.1214/15-EJS1088

[6] Bousquet, O., Boucheron, S. and Lugosi, G. (2004). Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning* (O. Bousquet, U. von Luxburg and G. Rätsch, eds.). *Lecture Notes in Artificial Intelligence* **3176** 169–207. Berlin: Springer.

1350-7265

[7] Bücher, A., Segers, J. and Volgushev, S. (2014). When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *Ann. Statist.* **42** 1598–1634. MR3262462 https://doi.org/10.1214/14-AOS1237

[8] Castro Camilo, D. and de Carvalho, M. (2017). Spectral density regression for bivariate extremes. *Stoch. Environ. Res. Risk Assess.* **31** 1603–1613.

[9] Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electron. J. Stat.* **9** 383–418. MR3323204 https://doi.org/10.1214/15-EJS1002

[10] Clémençon, S., Jalalzai, H., Lhaut, S., Sabourin, A. and Segers, J. (2023). Supplement to "Concentration bounds for the empirical angular measure with statistical learning applications." https://doi.org/10.3150/22-BEJ1562SUPP

[11] Coles, S.G. and Tawn, J.A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

[12] Cooley, D., Davis, R.A. and Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *J. Multivariate Anal.* **101** 2103–2117. MR2671204 https://doi.org/10.1016/j.jmva.2010.04.007

[13] Cooley, D. and Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika* **106** 587–604. MR3992391 https://doi.org/10.1093/biomet/asz028

[14] Das, B., Mitra, A. and Resnick, S. (2013). Living on the multidimensional edge: Seeking hidden risks using regular variation. *Adv. in Appl. Probab.* **45** 139–163. MR3077544

[15] de Carvalho, M. and Davison, A.C. (2014). Spectral density ratio models for multivariate extremes. *J. Amer. Statist. Assoc.* **109** 764–776. MR3223748 https://doi.org/10.1080/01621459.2013.872651

[16] de Carvalho, M., Oumow, B., Segers, J. and Warchoł, M. (2013). A Euclidean likelihood estimator for bivariate tail dependence. *Comm. Statist. Theory Methods* **42** 1176–1192. MR3031275 https://doi.org/10.1080/03610926.2012.709905

[17] de Haan, L. and de Ronde, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes* **1** 7–45. MR1652944 https://doi.org/10.1023/A:1009909800311

[18] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3

[19] 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

[20] de Haan, L. and Sinha, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.* **27** 732–759. MR1714710 https://doi.org/10.1214/aos/1018031214

[21] Drees, H. and Sabourin, A. (2021). Principal component analysis for multivariate extremes. *Electron. J. Stat.* **15** 908–943. MR4255291 https://doi.org/10.1214/21-ejs1803

[22] Einmahl, J.H.J., de Haan, L. and Krajina, A. (2013). Estimating extreme bivariate quantile regions. *Extremes* **16** 121–145. MR3057193 https://doi.org/10.1007/s10687-012-0156-z

[23] Einmahl, J.H.J., de Haan, L. and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.* **34** 1987–2014. MR2283724 https://doi.org/10.1214/009053606000000434

[24] Einmahl, J.H.J., de Haan, L. and Piterbarg, V.I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29** 1401–1423. MR1873336 https://doi.org/10.1214/aos/1013203459

[25] Einmahl, J.H.J., de Haan, L. and Sinha, A.K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70** 143–171. MR1475660 https://doi.org/10.1016/S0304-4149(97)00065-3

[26] Einmahl, J.H.J., Krajina, A. and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

[27] Einmahl, J.H.J. and Mason, D.M. (1992). Generalized quantile processes. *Ann. Statist.* **20** 1062–1078. MR1165606 https://doi.org/10.1214/aos/1176348670

[28] Einmahl, J.H.J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* **37** 2953–2989. MR2541452 https://doi.org/10.1214/08-AOS677

[29] Engelke, S. and Ivanovs, J. (2017). Robust bounds in multivariate extremes. *Ann. Appl. Probab.* **27** 3706–3734. MR3737936 https://doi.org/10.1214/17-AAP1294

[30] Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. MR4243547 https://doi.org/10.1146/annurev-statistics-040620-041554

[31] Fougères, A.-L., de Haan, L. and Mercadier, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43** 903–934. MR3325714 https://doi.org/10.1214/14-AOS1305

[32] Goix, N., Sabourin, A. and Clémençon, S. (2015). Learning the dependence structure of rare events: A nonasymptotic study. In *Proceedings of the International Conference on Learning Theory, COLT'15*.

[33] Goix, N., Sabourin, A. and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *J. Multivariate Anal.* **161** 12–31. MR3698112 https://doi.org/10.1016/j.jmva.2017.06.010

[34] Guillou, A., Naveau, P. and You, A. (2015). A folding methodology for multivariate extremes: Estimation of the spectral probability measure and actuarial applications. *Scand. Actuar. J.* **7** 549–572. MR3399703 https://doi.org/10.1080/03461238.2013.864326

[35] Haussler, D. (1995). Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *J. Combin. Theory Ser. A* **69** 217–232. MR1313896 https://doi.org/10.1016/0097-3165(95)90052-7

[36] Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)* **80** 121–140. MR2281910 https://doi.org/10.2298/PIM0694121H

[37] Jalalzai, H., Clémençon, S. and Sabourin, A. (2018). On binary classification in extreme regions. In *Advances in Neural Information Processing Systems* 3092–3100.

[38] Janßen, A. and Wan, P. (2020). k-means clustering of extremes. *Electron. J. Stat.* **14** 1211–1233. MR4071364 https://doi.org/10.1214/20-EJS1689

[39] Koltchinskii, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization (with discussion). *Ann. Statist.* **34** 2593–2706.

[40] Lehtomaa, J. and Resnick, S.I. (2020). Asymptotic independence and support detection techniques for heavy-tailed multivariate data. *Insurance Math. Econom.* **93** 262–277. MR4109082 https://doi.org/10.1016/j.insmatheco.2020.05.002

[41] Lhaut, S., Sabourin, A. and Segers, J. (2022). Uniform concentration bounds for frequencies of rare events. *Statist. Probab. Lett.* **189** Paper No. 109610, 7. MR4452775 https://doi.org/10.1016/j.spl.2022.109610

[42] Lugosi, G. (2002). Pattern classification and learning theory. In *Principles of Nonparametric Learning (Udine, 2001)*. *CISM Courses and Lect.* **434** 1–56. Vienna: Springer. MR1987656

[43] McDiarmid, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*. *Algorithms Combin.* **16** 195–248. Berlin: Springer. MR1678578 https://doi.org/10.1007/978-3-662-12788-9_6

[44] Meyer, N. and Wintenberger, O. (2021). Sparse regular variation. *Adv. in Appl. Probab.* **53** 1115–1148. MR4342579 https://doi.org/10.1017/apr.2021.14

[45] Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018). *Foundations of Machine Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. Second edition of [MR3057769]. MR3931734

[46] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. New York: Springer. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

[47] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424

[48] Sabourin, A. (2015). Semi-parametric modeling of excesses above high multivariate thresholds with censored data. *J. Multivariate Anal.* **136** 126–146. MR3321484 https://doi.org/10.1016/j.jmva.2015.01.014

[49] Sabourin, A. and Naveau, P. (2014). Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Comput. Statist. Data Anal.* **71** 542–567. MR3131989 https://doi.org/10.1016/j.csda.2013.04.021

[50] Scott, C.D. and Nowak, R.D. (2006). Learning minimum volume sets. *J. Mach. Learn. Res.* **7** 665–704. MR2274383

[51] 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

[52] Thomas, A., Clemencon, S., Gramfort, A. and Sabourin, A. (2017). Anomaly detection in extreme regions via empirical MV-sets on the sphere. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.). *Proceedings of Machine Learning Research* **54** 1011–1019. Fort Lauderdale, FL, USA: PMLR.

[53] 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

[54] Wadsworth, J.L., Tawn, J.A., Davison, A.C. and Elton, D.M. (2017). Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 149–175. MR3597968 https://doi.org/10.1111/rssb.12157

Bernoulli **29**(4), 2023, 2828–2853
https://doi.org/10.3150/22-BEJ1565

Semiparametric regression of panel count data with informative terminal event

XIANGBIN HU[1,a], LI LIU[2,c], YING ZHANG[3,d] and XINGQIU ZHAO[1,b]

[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China,* [a]*xiang-bin.hu@connect.polyu.hk,* [b]*xingqiu.zhao@polyu.edu.hk*
[2]*School of Mathematics and Statistics, Wuhan University, Wuhan, China,* [c]*lliu.math@whu.edu.cn*
[3]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, USA,* [d]*ying.zhang@unmc.edu*

We study a semiparametric model for robust analysis of panel count data with an informative terminal event. To explore the explicit effect of the terminal event on recurrent events of interest, we propose a conditional mean model for a reversed counting process anchoring at the terminal event. Treating the distribution function of the terminal event as a nuisance functional parameter, we develop a predicted least squares-based two-stage estimation procedure with the spline-based sieve estimation technique, and derive the convergence rate of the proposed estimator. Furthermore, overcoming the difficulties caused by the convergence rate slower than $1/\sqrt{n}$, we establish the asymptotic normality for the estimator of the finite-dimensional parameter and a functional of the estimator of the infinite-dimensional parameter. The proposed method is evaluated through extensive simulation studies and illustrated with an application to the Longitudinal Healthy Longevity Survey study on elder people in China.

Keywords: Asymptotic normality; counting process; empirical process; panel count data; predicted least squares; terminal event; two-stage estimation

References

Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *Ann. Statist.* **37** 1112–1149. MR2509069 https://doi.org/10.1214/08-AOS599

Breslow, N.E. (1972). Discussion of the paper by D. R. Cox. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **34** 216–217.

Chan, K.C.G. and Wang, M.-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *Ann. Appl. Stat.* **4** 1602–1620. MR2758343 https://doi.org/10.1214/09-AOAS319

Chiou, S.H., Huang, C.-Y., Xu, G. and Yan, J. (2019). Semiparametric regression analysis of panel count data: A practical review. *Int. Stat. Rev.* **87** 24–43. MR3940137 https://doi.org/10.1111/insr.12271

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

Hu, X., Liu, L., Zhang, Y. and Zhao, X. (2023). Supplement to "Semiparametric regression of panel count data with informative terminal event." https://doi.org/10.3150/22-BEJ1565SUPP

Kalbfleisch, J.D. and Lawless, J.F. (1985). The analysis of panel data under a Markov assumption. *J. Amer. Statist. Assoc.* **80** 863–871. MR0819585

Kong, S., Nan, B., Kalbfleisch, J.D., Saran, R. and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *J. Amer. Statist. Assoc.* **113** 357–368. MR3803470 https://doi.org/10.1080/01621459.2016.1255637

Lange, K. (2001). *Numerical Analysis for Statisticians*. *Statistics and Computing*. New York: Springer. MR1681963

Li, L., Wu, C.-H., Ning, J., Huang, X., Shih, Y.-C.T. and Shen, Y. (2018). Semiparametric estimation of longitudinal medical cost trajectory. *J. Amer. Statist. Assoc.* **113** 582–592. MR3832210 https://doi.org/10.1080/01621459.2017.1361329

Li, Z., Frost, H.R., Tosteson, T.D., Zhao, L., Liu, L., Lyons, K., Chen, H., Cole, B., Currow, D. and Bakitas, M. (2017). A semiparametric joint model for terminal trend of quality of life and survival in palliative care research. *Stat. Med.* **36** 4692–4704. MR3731248 https://doi.org/10.1002/sim.7445

1350-7265

Liu, L., Su, W., Yin, G., Zhao, X. and Zhang, Y. (2022). Nonparametric inference for reversed mean models with panel count data. *Bernoulli* **28** 2968–2997. MR4474569 https://doi.org/10.3150/21-bej1444

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94** 705–718. MR2410018 https://doi.org/10.1093/biomet/asm057

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *J. Amer. Statist. Assoc.* **104** 1060–1070. MR2750237 https://doi.org/10.1198/jasa.2009.tm08086

Shen, B., Chen, C., Liu, D., Datta, S., Ghahramani, N., Chinchilli, V.M. and Wang, M. (2021). Joint modeling of longitudinal data with informative cluster size adjusted for zero-inflation and a dependent terminal event. *Stat. Med.* **40** 4582–4596. MR4315439 https://doi.org/10.1002/sim.9081

Sun, J. and Kalbfleisch, J.D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5** 279–289. MR1329298

Sun, J., Tong, X. and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics* **63** 1053–1059. MR2414582 https://doi.org/10.1111/j.1541-0420.2007.00808.x

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

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Amer. Statist. Assoc.* **107** 688–700. MR2980077 https://doi.org/10.1080/01621459.2012.682528

Tong, X., He, X., Sun, L. and Sun, J. (2009). Variable selection for panel count data via non-concave penalized estimating function. *Scand. J. Stat.* **36** 620–635. MR2572579 https://doi.org/10.1111/j.1467-9469.2009.00658.x

van der Vaart, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1999)*. *Lecture Notes in Math.* **1781** 331–457. Berlin: Springer. MR1915446

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

Wellner, J.A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. MR1792787 https://doi.org/10.1214/aos/1015951998

Wellner, J.A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35** 2106–2142. MR2363965 https://doi.org/10.1214/009053607000000181

Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y. and Zhang, C. Chinese Longitudinal Healthy Longevity Survey (CLHLS) 1998–2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] 2017-04-11. https://doi.org/10.3886/ICPSR36692.v1

Zhang, H., Sun, J. and Wang, D. (2013). Variable selection and estimation for multivariate panel count data via the seamless-L_0 penalty. *Canad. J. Statist.* **41** 368–385. MR3061885 https://doi.org/10.1002/cjs.11172

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89** 39–48. MR1888344 https://doi.org/10.1093/biomet/89.1.39

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika* **93** 777–790. MR2285071 https://doi.org/10.1093/biomet/93.4.777

Zhao, H., Li, Y. and Sun, J. (2013a). Analyzing panel count data with a dependent observation process and a terminal event. *Canad. J. Statist.* **41** 174–191. MR3030791 https://doi.org/10.1002/cjs.11143

Zhao, H., Li, Y. and Sun, J. (2013b). Semiparametric analysis of multivariate panel count data with dependent observation processes and a terminal event. *J. Nonparametr. Stat.* **25** 379–394. MR3056091 https://doi.org/10.1080/10485252.2012.758724

Zhao, X. and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* **67** 770–779. MR2829131 https://doi.org/10.1111/j.1541-0420.2010.01504.x

Zhao, X. and Zhang, Y. (2017). Asymptotic normality of nonparametric M-estimators with applications to hypothesis testing for panel count data. *Statist. Sinica* **27** 931–950. MR3675037

Zhou, J., Zhang, H., Sun, L. and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Anal.* **23** 560–584. MR3705825 https://doi.org/10.1007/s10985-016-9375-y

Bernoulli **29**(4), 2023, 2854–2878
https://doi.org/10.3150/22-BEJ1566

Cramér type moderate deviations for the Grenander estimator near the boundaries of the support

FUQING GAO[1,a], HUI JIANG[2,b] and XINGQIU ZHAO[3,c]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, China,* [a]*fqgao@whu.edu.cn*
[2]*School of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, China,* [b]*huijiang@nuaa.edu.cn*
[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China,* [c]*xingqiu.zhao@polyu.edu.hk*

The Grenander estimator (nonparametric maximum likelihood estimator) of a decreasing density function is not consistent at boundaries of support. This phenomenon has great influences on the global measures of deviation between density function and its Grenander estimator. In this article, by strong approximation technique and comparison method, we establish Cramér type moderate deviations for the Grenander estimator near the boundaries of the support. The obtained results provide a nice uniform comparison between tail probability of the estimator and limiting fluctuation distribution.

Keywords: Empirical process; Grenander estimator; Cramér type moderate deviations; strong approximation

References

[1] Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A. and Wellner, J. (2011). On the Grenander estimator at zero. *Statist. Sinica* **21** 873–899. MR2829859 https://doi.org/10.5705/ss.2011.038a
[2] Bhattacharya, P.K. and Brockwell, P.J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. Verw. Gebiete* **37** 51–75. MR0426166 https://doi.org/10.1007/BF00536298
[3] Chang, J., Shao, Q.-M. and Zhou, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *Ann. Statist.* **44** 1931–1956. MR3546439 https://doi.org/10.1214/15-AOS1375
[4] Durot, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122** 222–240. MR1894068 https://doi.org/10.1007/s004400100171
[5] Durot, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497
[6] Durot, C., Kulikov, V.N. and Lopuhaä, H.P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036 https://doi.org/10.1214/12-AOS1015
[7] Durot, C. and Lopuhaä, H.P. (2014). A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electron. J. Stat.* **8** 2479–2513. MR3285873 https://doi.org/10.1214/14-EJS958
[8] 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
[9] Fan, X., Grama, I., Liu, Q. and Shao, Q.-M. (2019). Self-normalized Cramér type moderate deviations for martingales. *Bernoulli* **25** 2793–2823. MR4003565 https://doi.org/10.3150/18-BEJ1071
[10] Fang, X., Luo, L. and Shao, Q.-M. (2020). A refined Cramér-type moderate deviation for sums of local statistics. *Bernoulli* **26** 2319–2352. MR4091111 https://doi.org/10.3150/20-BEJ1195

[11] Gao, F., Xiong, J. and Zhao, X. (2018). Moderate deviations and nonparametric inference for monotone functions. *Ann. Statist.* **46** 1225–1254. MR3798002 https://doi.org/10.1214/17-AOS1583

[12] Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.* **39** 1211–1240. MR2816352 https://doi.org/10.1214/10-AOS865

[13] Gao, F., Jiang, H. and Zhao, X. (2023). Supplement to "Cramér type moderate deviations for the Grenander estimator near the boundaries of the support." https://doi.org/10.3150/22-BEJ1566SUPP

[14] 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

[15] Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[16] Groeneboom, P. (2021). Grenander functionals and Cauchy's formula. *Scand. J. Stat.* **48** 275–294. MR4233173 https://doi.org/10.1111/sjos.12449

[17] Groeneboom, P., Hooghiemstra, G. and Lopuhaä, H.P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. MR1740109 https://doi.org/10.1214/aos/1017938928

[18] 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

[19] Groeneboom, P. and Jongbloed, G. (2018). Some developments in the theory of shape constrained inference. *Statist. Sci.* **33** 473–492. MR3881204 https://doi.org/10.1214/18-STS657

[20] Groeneboom, P. and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Basel: Birkhäuser. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[21] Hooghiemstra, G. and Lopuhaä, H.P. (1998). An extremal limit theorem for the argmax process of Brownian motion minus a parabolic drift. *Extremes* **1** 215–240. MR1814624 https://doi.org/10.1023/A:1009962823531

[22] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[23] Kulikov, V.N. and Lopuhaä, H.P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. MR2211085 https://doi.org/10.1214/009053605000000462

[24] Kulikov, V.N. and Lopuhaä, H.P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34** 742–768. MR2283391 https://doi.org/10.1214/009053606000000100

[25] Liu, W. and Shao, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363 https://doi.org/10.1214/09-AOS774

[26] Lopuhaä, H.P. and Musta, E. (2019). Central limit theorems for the L_p-error of smooth isotonic estimators. *Electron. J. Stat.* **13** 1031–1098. MR3935844 https://doi.org/10.1214/19-ejs1550

[27] Mukherjee, R. and Sen, B. (2019). On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density. *Electron. J. Stat.* **13** 4416–4448. MR4028511 https://doi.org/10.1214/19-ejs1629

[28] Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

[29] Shao, Q.-M. and Zhou, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719

[30] Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plann. Inference* **52** 143–159. MR1392133 https://doi.org/10.1016/0378-3758(95)00114-X

[31] 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

[32] Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3** 501–515. MR1243398

Bernoulli **29**(4), 2023, 2879–2901
https://doi.org/10.3150/22-BEJ1567

Dating the break in high-dimensional data

RUNMIN WANG[1,a] and XIAOFENG SHAO[2,b]

[1]*Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.,* [a]*runminw@tamu.edu*
[2]*Department of Statistics, University of Illinois at Urbana-Champaign, Illinois 61820, U.S.A.,* [b]*xshao@illinos.edu*

This paper is concerned with estimation and inference for the location of a change point in the mean of independent high-dimensional data. Our change point location estimator maximizes a new U-statistic based objective function, and its convergence rate and asymptotic distribution after suitable centering and normalization are obtained under mild assumptions. Our estimator turns out to have better efficiency as compared to the least squares based counterpart in the literature. Based on the asymptotic theory, we construct a confidence interval by plugging in consistent estimates of several quantities in the normalization. We also provide a bootstrap-based confidence interval and state its asymptotic validity under suitable conditions. Through simulation studies, we demonstrate favorable finite sample performance of the new change point location estimator as compared to its least squares based counterpart, and our bootstrap-based confidence intervals, as compared to several existing competitors. The asymptotic theory based on high-dimensional U-statistic is substantially different from those developed in the literature and is of independent interest. The usefulness of our bootstrap-based confidence interval is illustrated in a genomics data set.

Keywords: Change point detection; high dimension; structural break; U-statistic

References

[1] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x
[2] Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* **15** 453–472. MR1292161 https://doi.org/10.1111/j.1467-9892.1994.tb00204.x
[3] Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. MR1455175 https://doi.org/10.1017/S0266466600005831
[4] Bai, J. (1997). Estimation of a change point in multiple regression models. *Rev. Econ. Stat.* **4** 551–563.
[5] Bai, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020
[6] Bai, J., Lumsdaine, R.L. and Stock, J.H. (1998). Testing for and dating common breaks in multivariate time series. *Rev. Econ. Stud.* **65** 395–432. MR1635358 https://doi.org/10.1111/1467-937X.00051
[7] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540
[8] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305
[9] Bhattacharjee, M., Banerjee, M. and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *J. Mach. Learn. Res.* **21** Paper No. 107, 59. MR4119175
[10] Bhattacharya, P.K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case. *J. Multivariate Anal.* **23** 183–208. MR0918253 https://doi.org/10.1016/0047-259X(87)90152-7
[11] Bredel, M., Bredel, C., Juric, D., Harsh, G.R., Vogel, H., Recht, L.D. and Sikic, B.I. (2005). High-resolution genome-wide mapping of genetic alterations in human glial brain tumors. *Cancer Res.* **65** 4088–4096.
[12] Chen, L., Wang, W. and Wu, W. (2019). Inference of break-points in high-dimensional time series. Available at SSRN 3378221.
[13] Chen, S.X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

[14] Cho, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. MR3522667 https://doi.org/10.1214/16-EJS1155
[15] Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079
[16] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2743035
[17] Dette, H., Pan, G. and Yang, Q. (2022). Estimating a change point in a sequence of very high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **117** 444–454. MR4399097 https://doi.org/10.1080/01621459.2020.1785477
[18] Eilers, P.H. and de Menezes, R.X. (2005). Quantile smoothing of array cgh data. *Bioinformatics* **21** 1146–1153.
[19] Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *Ann. Statist.* **47** 2051–2079. MR3953444 https://doi.org/10.1214/18-AOS1740
[20] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245
[21] Hinkley, D.V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727 https://doi.org/10.1093/biomet/57.1.1
[22] Hinkley, D.V. (1972). Time-ordered classification. *Biometrika* **59** 509–523. MR0368317 https://doi.org/10.1093/biomet/59.3.509
[23] Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x
[24] Jeng, X.J., Cai, T.T. and Li, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.* **105** 1156–1166. MR2752611 https://doi.org/10.1198/jasa.2010.tm10083
[25] Jirak, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347
[26] Killick, R. and Eckley, I. (2014). Changepoint: An R package for changepoint analysis. *J. Stat. Softw.* **58** 1–19.
[27] Lai, T.L., Xing, H. and Zhang, N. (2008). Stochastic segmentation models for array-based comparative genomic hybridization data analysis. *Biostatistics* **9** 290–307.
[28] Lai, W.R., Johnson, M.D., Kucherlapati, R. and Park, P.J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* **21** 3763–3770.
[29] Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics* **1** 278–352.
[30] Picard, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867. MR0809433 https://doi.org/10.2307/1427090
[31] Picard, F., Robin, S., Lavielle, M., Vaisse, C. and Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinform.* **6** 1–14.
[32] Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. MR2758116 https://doi.org/10.1111/j.1467-9868.2009.00737.x
[33] Shao, X. (2015). Self-normalization for time series: A review of recent developments. *J. Amer. Statist. Assoc.* **110** 1797–1817. MR3449074 https://doi.org/10.1080/01621459.2015.1050493
[34] Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103
[35] Stryhn, H. (1996). The location of the maximum of asymmetric two-sided Brownian motion with triangular drift. *Statist. Probab. Lett.* **29** 279–284. MR1411427 https://doi.org/10.1016/0167-7152(95)00183-2
[36] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[37] Wang, R. and Shao, X. (2020). Hypothesis testing for high-dimensional time series via self-normalization. *Ann. Statist.* **48** 2728–2758. MR4152119 https://doi.org/10.1214/19-AOS1904
[38] Wang, R. and Shao, X. (2023). Supplement to “Dating the break in high-dimensional data.” https://doi.org/10.3150/22-BEJ1567SUPP

[39] Wang, R., Zhu, C., Volgushev, S. and Shao, X. (2022). Inference for change points in high dimensional data via self-normalization. *Ann. Statist.* **50** 781–806. MR4405366 https://doi.org/10.1214/21-AOS2127
[40] Wang, T. and Samworth, R.J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243
[41] Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Statist.* **15** 1321–1328. MR0902262 https://doi.org/10.1214/aos/1176350509
[42] Yu, M. and Chen, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406
[43] Zhang, N.R. and Siegmund, D.O. (2012). Model selection for high-dimensional, multi-sequence change-point problems. *Statist. Sinica* **22** 1507–1538. MR3027097
[44] Zhang, N.R., Siegmund, D.O., Ji, H. and Li, J.Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025
[45] Zhang, Y., Wang, R. and Shao, X. (2022). Adaptive inference for change points in high-dimensional data. *J. Amer. Statist. Assoc.* **117** 1751–1762. MR4528468 https://doi.org/10.1080/01621459.2021.1884562

Bernoulli **29**(4), 2023, 2902–2926
https://doi.org/10.3150/22-BEJ1568

The uniform infinite cubic planar graph

BENEDIKT STUFLER[a]

Vienna University of Technology, Institute of Discrete Mathematics and Geometry, Vienna, Austria,
[a]*benedikt.stufler@tuwien.ac.at*

We prove that the random simple connected cubic planar graph C_n with an even number n of vertices admits a novel uniform infinite cubic planar graph (UICPG) as quenched local limit. We describe how the limit may be constructed by a series of random blow-up operations applied to the dual map of the type III Uniform Infinite Planar Triangulation established by Angel and Schramm (*Comm. Math. Phys.* **241** (2003) 191–213). Our main technical lemma is a contiguity relation between C_n and a model where the networks inserted at the links of the largest 3-connected component of C_n are replaced by independent copies of a specific Boltzmann network. We prove that the number of vertices of the largest 3-connected component concentrates at κn for $\kappa \approx 0.85085$, with Airy-type fluctuations of order $n^{2/3}$. The second-largest component is shown to have significantly smaller size $O_p(n^{2/3})$.

Keywords: Local limits; planar graphs; random graphs

References

[1] Addario-Berry, L. and Albenque, M. (2017). The scaling limit of random simple triangulations and random simple quadrangulations. *Ann. Probab.* **45** 2767–2825. MR3706731 https://doi.org/10.1214/16-AOP1124
[2] Ambjørn, J. and Budd, T.G. (2016). Multi-point functions of weighted cubic maps. *Ann. Inst. Henri Poincaré D* **3** 1–44. MR3462628 https://doi.org/10.4171/AIHPD/23
[3] Angel, O. and Schramm, O. (2003). Uniform infinite planar triangulations. *Comm. Math. Phys.* **241** 191–213. MR2013797 https://doi.org/10.1007/978-1-4419-9675-6_16
[4] Banderier, C., Flajolet, P., Schaeffer, G. and Soria, M. (2001). Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures Algorithms* **19** 194–246. Analysis of algorithms (Krynica Morska, 2000). MR1871555 https://doi.org/10.1002/rsa.10021
[5] Banderier, C., Kuba, M. and Wallner, M. (2021). Phase transitions of composition schemes: Mittag-Leffler and mixed Poisson distributions. Available at arXiv:2103.03751.
[6] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1700749 https://doi.org/10.1002/9780470316962
[7] Björnberg, J.E. and Stefánsson, S.Ö. (2014). Recurrence of bipartite planar maps. *Electron. J. Probab.* **19** no. 31, 40. MR3183575 https://doi.org/10.1214/EJP.v19-3102
[8] Bodirsky, M., Kang, M., Löffler, M. and McDiarmid, C. (2007). Random cubic planar graphs. *Random Structures Algorithms* **30** 78–94. MR2283223 https://doi.org/10.1002/rsa.20149
[9] Budzinski, T. and Louf, B. (2021). Local limits of uniform triangulations in high genus. *Invent. Math.* **223** 1–47. MR4199439 https://doi.org/10.1007/s00222-020-00986-3
[10] Budzinski, T. and Louf, B. (2022). Local limits of bipartite maps with prescribed face degrees in high genus. *Ann. Probab.* **50** 1059–1126. MR4413212 https://doi.org/10.1214/21-aop1554
[11] Curien, N., Ménard, L. and Miermont, G. (2013). A view from infinity of the uniform infinite planar quadrangulation. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 45–88. MR3083919
[12] Drmota, M., Noy, M., Requilé, C. and Rué, J. Random cubic planar maps. Available at arXiv:2209.14799.
[13] Fang, W., Kang, M., Moßhammer, M. and Sprüssel, P. (2018). Cubic graphs and related triangulations on orientable surfaces. *Electron. J. Combin.* **25** Paper No. 1.30, 52. MR3785009 https://doi.org/10.37236/5989
[14] Gao, Z. and Wormald, N.C. (1999). The size of the largest components in random planar maps. *SIAM J. Discrete Math.* **12** 217–228. MR1686830 https://doi.org/10.1137/S0895480195292053

1350-7265

[15] Gao, Z. and Wormald, N.C. (2004). Asymptotic normality determined by high moments, and submap counts of random maps. *Probab. Theory Related Fields* **130** 368–376. MR2095934 https://doi.org/10.1007/s00440-004-0356-9

[16] Giménez, O., Noy, M. and Rué, J. (2013). Graph classes with given 3-connected components: Asymptotic enumeration and random graphs. *Random Structures Algorithms* **42** 438–479. MR3068033 https://doi.org/10.1002/rsa.20421

[17] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff Publishing. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman. MR0322926

[18] Janson, S. (2012). Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.* **9** 103–252. MR2908619 https://doi.org/10.1214/11-PS188

[19] Kang, M. and Missethan, M. (2021). Local limit of sparse random planar graphs. Available at arXiv:2101.11910.

[20] McDiarmid, C. (2009). Random graphs from a minor-closed class. *Combin. Probab. Comput.* **18** 583–599. MR2507738 https://doi.org/10.1017/S0963548309009717

[21] Ménard, L. and Nolin, P. (2014). Percolation on uniform infinite planar maps. *Electron. J. Probab.* **19** no. 79, 27. MR3256879 https://doi.org/10.1214/EJP.v19-2675

[22] Noy, M., Requilé, C. and Rué, J. (2017). Enumeration of labeled 4-regular planar graphs. In *Extended Abstracts of the Ninth European Conference on Combinatorics, Graph Theory and Applications, EuroComb 2017, Vienna, Austria, August 28 – September 1, 2017* 933–939. Amsterdam: Elsevier. https://doi.org/10.1016/j.endm.2017.07.056

[23] Noy, M., Requilé, C. and Rué, J. (2020). Further results on random cubic planar graphs. *Random Structures Algorithms* **56** 892–924. MR4084194 https://doi.org/10.1002/rsa.20893

[24] Noy, M., Requilé, C. and Rué, J. (2022). On the expected number of perfect matchings in cubic planar graphs. *Publ. Mat.* **66** 325–353. MR4366217 https://doi.org/10.5565/publmat6612213

[25] Poulalhon, D. and Schaeffer, G. (2006). Optimal coding and sampling of triangulations. *Algorithmica* **46** 505–527. MR2291967 https://doi.org/10.1007/s00453-006-0114-8

[26] Requilé, C. and Rué, J. (2015). Triangles in random cubic planar graphs. In *Extended Abstracts of the Eight European Conference on Combinatorics, Graph Theory and Applications, EuroComb 2015, Bergen, Norway, August 31 – September 4, 2015* 383–391. Amsterdam: Elsevier. https://doi.org/10.1016/j.endm.2015.06.054

[27] Stephenson, R. (2018). Local convergence of large critical multi-type Galton-Watson trees and applications to random maps. *J. Theoret. Probab.* **31** 159–205. MR3769811 https://doi.org/10.1007/s10959-016-0707-3

[28] Stufler, B. (2018). Gibbs partitions: The convergent case. *Random Structures Algorithms* **53** 537–558. MR3854044 https://doi.org/10.1002/rsa.20771

[29] Stufler, B. (2021). Local convergence of random planar graphs. *J. Eur. Math. Soc. (JEMS)*, published online first. https://doi.org/10.4171/JEMS/1174

[30] Stufler, B. (2022). Quenched local convergence of Boltzmann planar maps. *J. Theoret. Probab.* **35** 1324–1342. MR4414421 https://doi.org/10.1007/s10959-021-01089-2

[31] Tutte, W.T. (1962). A census of planar triangulations. *Canad. J. Math.* **14** 21–38. MR0130841 https://doi.org/10.4153/CJM-1962-002-9

[32] Tutte, W.T. (1973). The enumerative theory of planar maps. In *A Survey of Combinatorial Theory (Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1971)* 437–448. Amsterdam: North-Holland. MR0364004

Bernoulli **29**(4), 2023, 2927–2960
https://doi.org/10.3150/22-BEJ1569

Exponential concentration for geometric-median-of-means in non-positive curvature spaces

HO YUN[a] and BYEONG U. PARK[b]

Department of Statistics, Seoul National University, Seoul, South Korea, [a]*zrose0921@gmail.com,* [b]*bupark@snu.ac.kr*

In Euclidean spaces, the empirical mean vector as an estimator of the population mean is known to have polynomial concentration unless a strong tail assumption is imposed on the underlying probability measure. The idea of median-of-means tournament has been considered as a way of overcoming the sub-optimality of the empirical mean vector. In this paper, to address the sub-optimal performance of the empirical mean in a more general setting, we consider general Polish spaces with a general metric, which are allowed to be non-compact and of infinite-dimension. We discuss the estimation of the associated population Fréchet mean, and for this we extend the existing notion of median-of-means to this general setting. We devise several new notions and inequalities associated with the geometry of the underlying metric, and using them we study the concentration properties of the extended notions of median-of-means as the estimators of the population Fréchet mean. We show that the new estimators achieve exponential concentration under only a second moment condition on the underlying distribution, while the empirical Fréchet mean has polynomial concentration. We focus our study on spaces with non-positive Alexandrov curvature since they afford slower rates of convergence than spaces with positive curvature. We note that this is the first work that derives non-asymptotic concentration inequalities for extended notions of the median-of-means in non-vector spaces with a general metric.

Keywords: Concentration inequalities; Fréchet mean; median-of-means estimators; non-Euclidean geometry; NPC spaces; power transform metric

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985 https://doi.org/10.1214/EJP.v13-521

[2] Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0

[3] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. MR2288028 https://doi.org/10.1137/050637996

[4] Bačák, M. (2014). Computing medians and means in Hadamard spaces. *SIAM J. Optim.* **24** 1542–1566. MR3264572 https://doi.org/10.1137/140953393

[5] Bačák, M. (2014). *Convex Analysis and Optimization in Hadamard Spaces*. *De Gruyter Series in Nonlinear Analysis and Applications* **22**. Berlin: de Gruyter. MR3241330 https://doi.org/10.1515/9783110361629

[6] Bačák, M. (2018). Old and new challenges in Hadamard spaces. ArXiv preprint. Available at arXiv:1807.01355.

[7] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

[8] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

1350-7265

[9] Billera, L.J., Holmes, S.P. and Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

[10] Bolley, F., Guillin, A. and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** 541–593. MR2280433 https://doi.org/10.1007/s00440-006-0004-7

[11] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[12] Catoni, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

[13] Catoni, O. and Giulini, I. (2018). Dimension-free PAC-Bayesian bounds for the estimation of the mean of a random vector. ArXiv preprint. Available at arXiv:1802.04308.

[14] Chen, Y., Lin, Z. and Müller, H.-G. (2021). Wasserstein regression. *J. Amer. Statist. Assoc.* **116** 1–14.

[15] Cherapanamjeri, Y., Flammarion, N. and Bartlett, P.L. (2019). Fast mean estimation with sub-Gaussian rates. In *Conference on Learning Theory* 786–806. PMLR.

[16] Depersin, J. and Lecué, G. (2021). On the robustness to adversarial corruption and to heavy-tailed data of the Stahel-Donoho median of means. ArXiv preprint. Available at arXiv:2101.09117.

[17] Depersin, J. and Lecué, G. (2022). Robust sub-Gaussian estimation of a mean vector in nearly linear time. *Ann. Statist.* **50** 511–536. MR4382026 https://doi.org/10.1214/21-aos2118

[18] Devroye, L., Lerasle, M., Lugosi, G. and Oliveira, R.I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440

[19] Fillard, P., Arsigny, V., Pennec, X., Hayashi, K.M., Thompson, P.M. and Ayache, N. (2007). Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *NeuroImage* **34** 639–650. https://doi.org/10.1016/j.neuroimage.2006.09.027

[20] Fillard, P., Arsigny, V., Pennec, X., Thompson, P.M. and Ayache, N. (2005). Extrapolation of sparse tensor fields: Application to the modeling of brain variability. In *Biennial International Conference on Information Processing in Medical Imaging* 27–38. Springer.

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

[22] Gallot, S., Hulin, D. and Lafontaine, J. (1990). *Riemannian Geometry*, 2nd ed. *Universitext*. Berlin: Springer. MR1083149 https://doi.org/10.1007/978-3-642-97242-3

[23] Ganea, O.-E., Bécigneul, G. and Hofmann, T. (2018). Hyperbolic neural networks. In *NeurIPS 2018*.

[24] Ghodrati, L. and Panaretos, V.M. (2022). Distribution-on-distribution regression via optimal transport maps. *Biometrika* **109** 957–974. MR4519110 https://doi.org/10.1093/biomet/asac005

[25] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics, [40]*. New York: Cambridge Univ. Press. MR3588285 https://doi.org/10.1017/CBO9781107337862

[26] Hopkins, S.B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843

[27] Horváth, L., Kokoszka, P. and Wang, S. (2021). Monitoring for a change point in a sequence of distributions. *Ann. Statist.* **49** 2271–2291. MR4319250 https://doi.org/10.1214/20-aos2036

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

[29] Kloeckner, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **9** 297–323. MR2731158

[30] Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. *Lecture Notes in Math.* **2033**. Heidelberg: Springer. MR2829871 https://doi.org/10.1007/978-3-642-22147-7

[31] Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer Science & Business Media.

[32] Le Gouic, T. and Loubes, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917. MR3663634 https://doi.org/10.1007/s00440-016-0727-z

[33] Lecué, G. and Lerasle, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. MR4102681 https://doi.org/10.1214/19-AOS1828

[34] Lederer, J. and van de Geer, S. (2014). New concentration inequalities for suprema of empirical processes. *Bernoulli* **20** 2020–2038. MR3263097 https://doi.org/10.3150/13-BEJ549
[35] Lei, Z., Luh, K., Venkat, P. and Zhang, F. (2020). A fast spectral algorithm for mean estimation with sub-Gaussian rates. In *Conference on Learning Theory* 2598–2612. PMLR.
[36] Lerasle, M., Szabó, Z., Mathieu, T. and Lecué, G. (2019). MONK outlier-robust mean embedding estimation by median-of-means. In *International Conference on Machine Learning* 3782–3793. PMLR.
[37] Lin, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. MR4032859 https://doi.org/10.1137/18M1221084
[38] Lin, Z., Müller, H.-G. and Park, B.U. (2021). Additive models for symmetric positive-definite matrices, Riemannian manifolds and Lie groups. ArXiv preprint. Available at arXiv:2009.08789.
[39] Lugosi, G. and Mendelson, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950 https://doi.org/10.1214/17-AOS1639
[40] Lugosi, G. and Mendelson, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x
[41] Lugosi, G. and Mendelson, S. (2021). Robust multivariate mean estimation: The optimality of trimmed mean. *Ann. Statist.* **49** 393–410. MR4206683 https://doi.org/10.1214/20-AOS1961
[42] Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645
[43] Nemirovsky, A.S. and Yudin, D.B. (1983). *Problem Complexity and Method Efficiency in Optimization*. *Wiley-Interscience Series in Discrete Mathematics*. New York: Wiley. MR0702836
[44] Panaretos, V.M. and Zemel, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. *SpringerBriefs in Probability and Mathematical Statistics*. Cham: Springer. MR4350694 https://doi.org/10.1007/978-3-030-38438-8
[45] Pennec, X., Sommer, S. and Fletcher, T. (2019). *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press.
[46] Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624
[47] Schötz, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618
[48] Sturm, K.-T., Coulhon, T. and Grigor'yan, A. (2003). Probability measures on metric spaces of nonpositive curvature. Heat kernels and analysis on manifolds, graphs, and metric spaces, contemporary mathematics. *Am. Math. Soc.* **358**.
[49] Tifrea, A., Bécigneul, G. and Ganea, O.-E. (2018). Poincaré glove: Hyperbolic word embeddings. ArXiv preprint. Available at arXiv:1810.06546.
[50] van de Geer, S. and Lederer, J. (2013). The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields* **157** 225–250. MR3101846 https://doi.org/10.1007/s00440-012-0455-y
[51] Van Handel, R. (2014). Probability in high dimension Technical Report Princeton Univ NJ.
[52] 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
[53] Yun, H. and Park, B.U. (2023). Supplement to "Exponential Concentration for Geometric-Median-of-Means in Non-Positive Curvature Spaces." https://doi.org/10.3150/22-BEJ1569SUPP
[54] Zhang, C., Kokoszka, P. and Petersen, A. (2022). Wasserstein autoregressive models for density time series. *J. Time Series Anal.* **43** 30–52. MR4400283 https://doi.org/10.1111/jtsa.12590

Bernoulli **29**(4), 2023, 2961–2986
https://doi.org/10.3150/22-BEJ1570

Inference for partially observed Riemannian Ornstein–Uhlenbeck diffusions of covariance matrices

MAI NGOC BUI[1,a], YVO POKERN[2,b] and PETROS DELLAPORTAS[2,3,c]

[1]*Department of Applied Health Research, University College London, Gower Street, London, WC1E 6BT, UK,* [a]*mai.bui@ucl.ac.uk*
[2]*Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK,* [b]*y.pokern@ucl.ac.uk*
[3]*Department of Statistics, Athens University of Economics and Business, Athens 10434, Greece,* [c]*p.dellaportas@ucl.ac.uk*

We construct a generalization of the Ornstein–Uhlenbeck processes on the cone of covariance matrices endowed with the Log-Euclidean and the Affine-Invariant metrics. Our development exploits the Riemannian geometric structure of symmetric positive definite matrices viewed as a differential manifold. We then provide Bayesian inference for discretely observed diffusion processes of covariance matrices based on an MCMC algorithm built with the help of a novel diffusion bridge sampler accounting for the geometric structure. Our proposed algorithm is illustrated with a real data financial application.

Keywords: Affine-Invariant metric; Log-Euclidean metric; Ornstein–Uhlenbeck process; Riemannian manifold; partially observed diffusion

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001). The distribution of realized stock return volatility. *J. Financ. Econ.* **61** 43–76.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55. MR1952727 https://doi.org/10.1198/016214501750332965

[3] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **56** 411–421.

[4] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2006). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. MR2288028 https://doi.org/10.1137/050637996

[5] Asai, M. and So, M.K.P. (2013). Stochastic covariance models. *J. Japan Statist. Soc.* **43** 127–162. MR3236278 https://doi.org/10.14490/jjss.43.127

[6] Ball, F.G., Dryden, I.L. and Golalizadeh, M. (2008). Brownian motion and Ornstein-Uhlenbeck processes in planar shape space. *Methodol. Comput. Appl. Probab.* **10** 1–22. MR2394032 https://doi.org/10.1007/s11009-007-9042-6

[7] Barndorff-Nielsen, O.E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925. MR2051439 https://doi.org/10.1111/j.1468-0262.2004.00515.x

[8] Barndorff-Nielsen, O.E. and Stelzer, R. (2007). Positive-definite matrix processes of finite variation. *Probab. Math. Statist.* **27** 3–43. MR2353270

[9] Bauer, G.H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *J. Econometrics* **160** 93–101. MR2745870 https://doi.org/10.1016/j.jeconom.2010.03.021

[10] Baxendale, P. (1976). Measures and Markov processes on function spaces. *Mém. Soc. Math. Fr.* **46** 131–141. MR0440013 https://doi.org/10.24033/msmf.190
[11] Beskos, A., Kalogeropoulos, K. and Pazos, E. (2013). Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.* **123** 1415–1453. MR3016228 https://doi.org/10.1016/j.spa.2012.12.001
[12] Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 395–410. MR2155345 https://doi.org/10.1111/j.1467-9868.2005.00508.x
[13] Boothby, W.M. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed. *Pure and Applied Mathematics* **120**. Orlando, FL: Academic Press. MR0861409
[14] Bru, M.-F. (1991). Wishart processes. *J. Theoret. Probab.* **4** 725–751. MR1132135 https://doi.org/10.1007/BF01259552
[15] Bui, M.N., Pokern, Y. and Dellaportas, P. (2023). Supplement to "Inference for partially observed Riemannian Ornstein–Uhlenbeck diffusions of covariance matrices." https://doi.org/10.3150/22-BEJ1570SUPP
[16] Buraschi, A., Porchia, P. and Trojani, F. (2010). Correlation risk and optimal portfolio choice. *J. Finance* **65** 393–420.
[17] Caseiro, R., Martins, P., Henriques, J.F. and Batista, J. (2012). A nonparametric Riemannian framework on tensor field with application to foreground segmentation. *Pattern Recognit.* **45** 3997–4017.
[18] Chefd'hotel, C., Tschumperlé, D., Deriche, R. and Faugeras, O. (2004). Regularizing flows for constrained matrix-valued images. *J. Math. Imaging Vision* **20** 147–162. MR2049787 https://doi.org/10.1023/B:JMIV.0000011324.14508.fb
[19] Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *J. Appl. Econometrics* **26** 922–947. MR2843112 https://doi.org/10.1002/jae.1152
[20] Cook, S.R., Gelman, A. and Rubin, D.B. (2006). Validation of software for Bayesian models using posterior quantiles. *J. Comput. Graph. Statist.* **15** 675–692. MR2291268 https://doi.org/10.1198/106186006X136976
[21] FirstRate Data (2019). Historical Intraday Index Price Data. Available at http://firstratedata.com/it/index.
[22] Del Moral, P. and Murray, L.M. (2015). Sequential Monte Carlo with highly informative observations. *SIAM/ASA J. Uncertain. Quantificat.* **3** 969–997. MR3403141 https://doi.org/10.1137/15M1011214
[23] Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116** 1660–1675. MR2269221 https://doi.org/10.1016/j.spa.2006.04.004
[24] do Carmo, M.P. (1992). *Riemannian Geometry*. *Mathematics: Theory & Applications*. Boston, MA: Birkhäuser, Inc. MR1138207 https://doi.org/10.1007/978-1-4757-2201-7
[25] Dryden, I.L., Koloydenko, A. and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249
[26] Durham, G.B. and Gallant, A.R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. MR1939904 https://doi.org/10.1198/073500102288618397
[27] Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375 https://doi.org/10.1111/1468-0262.00226
[28] Elworthy, K.D. (1982). *Stochastic Differential Equations on Manifolds*. *London Mathematical Society Lecture Note Series* **70**. Cambridge: Cambridge Univ. Press. MR0675100
[29] Epps, T.W. (1979). Comovements in stock prices in the very short run. *J. Amer. Statist. Assoc.* **74** 291–298.
[30] Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19** 177–191. MR1939708 https://doi.org/10.1198/073500101316970403
[31] Fletcher, P.T. and Joshi, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.* **87** 250–262.
[32] Gangolli, R. (1964). On the construction of certain diffusions on a differentiable manifold. *Z. Wahrsch. Verw. Gebiete* **2** 406–419. MR0165590 https://doi.org/10.1007/BF00533608
[33] Golightly, A. and Wilkinson, D.J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. MR2422763 https://doi.org/10.1016/j.csda.2007.05.019
[34] Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *J. Econometrics* **167** 211–223. MR2885447 https://doi.org/10.1016/j.jeconom.2011.11.004

[35] Gourieroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Rev.* **25** 177–217. MR2256286 https://doi.org/10.1080/07474930600713234

[36] Gourieroux, C., Jasiak, J. and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *J. Econometrics* **150** 167–181. MR2535514 https://doi.org/10.1016/j.jeconom.2008.12.016

[37] Gourieroux, C. and Sufana, R. (2010). Derivative pricing with Wishart multivariate stochastic volatility. *J. Bus. Econom. Statist.* **28** 438–451. MR2723611 https://doi.org/10.1198/jbes.2009.08105

[38] Hansen, P.R., Lunde, A. and Voev, V. (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *J. Appl. Econometrics* **29** 774–799. MR3258063 https://doi.org/10.1002/jae.2389

[39] Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Rev. Financ. Stud.* **18** 37–84.

[40] Hsu, E.P. (2002). *Stochastic Analysis on Manifolds*. *Graduate Studies in Mathematics* **38**. Providence, RI: Amer. Math. Soc. MR1882015 https://doi.org/10.1090/gsm/038

[41] Huckemann, S.F., Kim, P.T., Koo, J.-Y. and Munk, A. (2010). Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Statist.* **38** 2465–2498. MR2676895 https://doi.org/10.1214/09-AOS783

[42] Jin, X. and Maheu, J.M. (2013). Modeling realized covariances and returns. *J. Financ. Econom.* **11** 335–369.

[43] Jin, X., Maheu, J.M. and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *J. Appl. Econometrics* **34** 641–660. MR4002445 https://doi.org/10.1002/jae.2685

[44] Jørgensen, E. (1978). Construction of the Brownian motion and the Ornstein-Uhlenbeck process in a Riemannian manifold on basis of the Gangolli-McKean injection scheme. *Z. Wahrsch. Verw. Gebiete* **44** 71–87. MR0503466 https://doi.org/10.1007/BF00534142

[45] Jost, J. (2008). *Riemannian Geometry and Geometric Analysis*, 5th ed. *Universitext*. Berlin: Springer. MR2431897

[46] Jung, S., Schwartzman, A. and Groisser, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **36** 1180–1201. MR3379023 https://doi.org/10.1137/140967040

[47] Kalogeropoulos, K., Roberts, G.O. and Dellaportas, P. (2010). Inference for stochastic volatility models using time change transformations. *Ann. Statist.* **38** 784–807. MR2604696 https://doi.org/10.1214/09-AOS702

[48] Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541. MR0442975 https://doi.org/10.1002/cpa.3160300502

[49] Lin, M., Chen, R. and Mykland, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105** 820–838. MR2724864 https://doi.org/10.1198/jasa.2010.tm09057

[50] Lindström, E. (2012). A regularized bridge sampler for sparsely sampled diffusions. *Stat. Comput.* **22** 615–623. MR2865039 https://doi.org/10.1007/s11222-011-9255-y

[51] Lovett, S. (2010). *Differential Geometry of Manifolds*. Natick, MA: A K Peters, Ltd. MR2676438

[52] Manton, J.H. (2013). A primer on stochastic differential geometry for signal processing. *IEEE J. Sel. Top. Signal Process.* **7** 681–699.

[53] Matsumoto, H. (2001). Closed form formulae for the heat kernels and the Green functions for the Laplacians on the symmetric spaces of rank one. *Bull. Sci. Math.* **125** 553–581. MR1869991 https://doi.org/10.1016/S0007-4497(01)01099-5

[54] McKean, H.P. Jr. (1960). Brownian motions on the 3-dimensional rotation group. *Mem. Coll. Sci., Univ. Kyoto, Ser. A: Math.* **33** 25–38. MR0133869 https://doi.org/10.1215/kjm/1250776060

[55] Mider, M., Jenkins, P.A., Pollock, M., Roberts, G.O. and Sørensen, M. (2019). Simulating bridges using confluent diffusions. ArXiv preprint. Available at arXiv:1903.10184.

[56] Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26** 735–747. MR2137480 https://doi.org/10.1137/S0895479803436937

[57] Moakher, M. and Zéraï, M. (2011). The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *J. Math. Imaging Vision* **40** 171–187. MR2782125 https://doi.org/10.1007/s10851-010-0255-x

[58] Nagano, Y., Yamaguchi, S., Fujita, Y. and Koyama, M. (2019). A wrapped normal distribution on hyperbolic space for gradient-based learning. In *International Conference on Machine Learning* 4693–4702.

[59] Noureldin, D., Shephard, N. and Sheppard, K. (2012). Multivariate high-frequency-based volatility (heavy) models. *J. Appl. Econometrics* **27** 907–933. MR2993282 https://doi.org/10.1002/jae.1260
[60] Nye, T.M.W. (2011). Principal components analysis in the space of phylogenetic trees. *Ann. Statist.* **39** 2716–2739. MR2906884 https://doi.org/10.1214/11-AOS915
[61] Pedersen, A.R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1** 257–279. MR1363541 https://doi.org/10.2307/3318480
[62] Pennec, X. (2006). Statistical computing on manifolds for computational anatomy Ph.D. thesis Université Nice Sophia Antipolis.
[63] Pfaffel, O. (2012). Wishart processes. ArXiv preprint. Available at arXiv:1201.3256.
[64] Roberts, G.O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88** 603–621. MR1859397 https://doi.org/10.1093/biomet/88.3.603
[65] Said, S., Bombrun, L., Berthoumieu, Y. and Manton, J.H. (2017). Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Trans. Inf. Theory* **63** 2153–2170. MR3626862 https://doi.org/10.1109/TIT.2017.2653803
[66] Schauer, M., van der Meulen, F. and van Zanten, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. MR3648050 https://doi.org/10.3150/16-BEJ833
[67] Skovgaard, L.T. (1984). A Riemannian geometry of the multivariate normal model. *Scand. J. Stat.* **11** 211–223. MR0793171
[68] Sommer, S., Arnaudon, A., Kühnel, L. and Joshi, S. (2017). Bridge simulation and metric estimation on landmark manifolds. ArXiv preprint. Available at arXiv:1705.10943.
[69] Staneva, V. and Younes, L. (2017). Learning shape trends: Parameter estimation in diffusions on shape manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 38–46.
[70] Stramer, O., Bognar, M. and Schneider, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. *J. Financ. Econom.* **8** 450–480.
[71] Tschumperle, D. and Deriche, R. (2001). Diffusion tensor regularization with constraints preservation. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* **1** I–I. IEEE.
[72] van der Meulen, F. and Schauer, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. MR3656495 https://doi.org/10.1214/17-EJS1290
[73] Whitaker, G.A., Golightly, A., Boys, R.J. and Sherlock, C. (2017). Improved bridge constructs for stochastic differential equations. *Stat. Comput.* **27** 885–900. MR3627552 https://doi.org/10.1007/s11222-016-9660-3
[74] Yu, P.L.H., Li, W.K. and Ng, F.C. (2017). The generalized conditional autoregressive Wishart model for multivariate realized volatility. *J. Bus. Econom. Statist.* **35** 513–527. MR3716040 https://doi.org/10.1080/07350015.2015.1096788
[75] Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *J. Econometrics* **160** 33–47. MR2745865 https://doi.org/10.1016/j.jeconom.2010.03.012

Bernoulli **29**(4), 2023, 2987–3012
https://doi.org/10.3150/22-BEJ1571

SPDEs with non-Lipschitz coefficients and nonhomogeneous boundary conditions

JIE XIONG[1,a] and XU YANG[2,b]

[1]*Department of Mathematics and SUSTech International center for Mathematics, Southern University of Science and Technology, Shenzhen, China,* [a]*xiongj@sustech.edu.cn*
[2]*School of Mathematics and Information Science, North Minzu University, Yinchuan, China,* [b]*xuyang@mail.bnu.edu.cn*

In this article, we consider a stochastic partial differential equation (SPDE) driven by Gaussian colored noise with Dirichlet, Neumann or mixed nonhomogeneous random boundary conditions when the drift and diffusion coefficients are non-Lipschitz. We prove the existence of a unique strong solution to this SPDE and obtain a comparison theorem between such SPDEs. We also study the Hölder continuity of the solution in both time and space variables, and find the dependence of the Hölder exponent on that of the Dirichlet boundary.

Keywords: Boundary conditions; comparison theorem; Hölder continuity; non-Lipschitz coefficients; pathwise uniqueness; stochastic partial differential equation

References

Bally, V., Gyöngy, I. and Pardoux, É. (1994). White noise driven parabolic SPDEs with measurable drift. *J. Funct. Anal.* **120** 484–510. MR1266318 https://doi.org/10.1006/jfan.1994.1040
Dalang, R.C., Mueller, C. and Zambotti, L. (2006). Hitting properties of parabolic s.p.d.e.'s with reflection. *Ann. Probab.* **34** 1423–1450. MR2257651 https://doi.org/10.1214/009117905000000792
Dalang, R.C. and Zhang, T. (2013). Hölder continuity of solutions of SPDEs with reflection. *Commun. Math. Stat.* **1** 133–142. MR3197857 https://doi.org/10.1007/s40304-013-0009-3
Dawson, D.A. (1993). Measure-valued Markov processes. In *École D'Été de Probabilités de Saint-Flour XXI—1991*. *Lecture Notes in Math.* **1541** 1–260. Berlin: Springer. MR1242575 https://doi.org/10.1007/BFb0084190
Dawson, D.A. and Li, Z. (2006). Skew convolution semigroups and affine Markov processes. *Ann. Probab.* **34** 1103–1142. MR2243880 https://doi.org/10.1214/009117905000000747
Dawson, D.A. and Li, Z. (2012). Stochastic equations, flows and measure-valued processes. *Ann. Probab.* **40** 813–857. MR2952093 https://doi.org/10.1214/10-AOP629
Dawson, D.A., Vaillancourt, J. and Wang, H. (2000). Stochastic partial differential equations for a class of interacting measure-valued diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **36** 167–180. MR1751657 https://doi.org/10.1016/S0246-0203(00)00121-7
Donati-Martin, C. and Pardoux, É. (1993). White noise driven SPDEs with reflection. *Probab. Theory Related Fields* **95** 1–24. MR1207304 https://doi.org/10.1007/BF01197335
Donnelly, P. and Kurtz, T.G. (1999). Particle representations for measure-valued population models. *Ann. Probab.* **27** 166–205. MR1681126 https://doi.org/10.1214/aop/1022677258
Dynkin, E.B. (1994). *An Introduction to Branching Measure-Valued Processes*. *CRM Monograph Series* **6**. Providence, RI: Amer. Math. Soc. MR1280712 https://doi.org/10.1090/crmm/006
Etheridge, A.M. (2000). *An Introduction to Superprocesses*. *University Lecture Series* **20**. Providence, RI: Amer. Math. Soc. MR1779100 https://doi.org/10.1090/ulect/020
Gomez, A., Lee, K., Mueller, C., Wei, A. and Xiong, J. (2013). Strong uniqueness for an SPDE via backward doubly stochastic differential equations. *Statist. Probab. Lett.* **83** 2186–2190. MR3093800 https://doi.org/10.1016/j.spl.2013.06.010

1350-7265

Gyöngy, I. (1995). On non-degenerate quasi-linear stochastic partial differential equations. *Potential Anal.* **4** 157–171. MR1323824 https://doi.org/10.1007/BF01275588

Gyöngy, I. and Pardoux, É. (1993a). On quasi-linear stochastic partial differential equations. *Probab. Theory Related Fields* **94** 413–425. MR1201552 https://doi.org/10.1007/BF01192556

Gyöngy, I. and Pardoux, É. (1993b). On the regularization effect of space-time white noise on quasi-linear parabolic partial differential equations. *Probab. Theory Related Fields* **97** 211–229. MR1240724 https://doi.org/10.1007/BF01199321

Hsu, P. (1984). *Reflecting Brownian Motion, Boundary Local Time and the Neumann Problem*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)–Stanford University. MR2633560

Hu, Y., Lu, F. and Nualart, D. (2013). Hölder continuity of the solutions for a class of nonlinear SPDE's arising from one dimensional superprocesses. *Probab. Theory Related Fields* **156** 27–49. MR3055251 https://doi.org/10.1007/s00440-012-0419-2

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

Kalsi, J. (2020). Existence of invariant measures for reflected stochastic partial differential equations. *J. Theoret. Probab.* **33** 1755–1767. MR4125974 https://doi.org/10.1007/s10959-019-00906-z

Konno, N. and Shiga, T. (1988). Stochastic partial differential equations for some measure-valued diffusions. *Probab. Theory Related Fields* **79** 201–225. MR0958288 https://doi.org/10.1007/BF00320919

Kurtz, T.G. (1998). Martingale problems for conditional distributions of Markov processes. *Electron. J. Probab.* **3** no. 9, 29. MR1637085 https://doi.org/10.1214/EJP.v3-31

Kurtz, T.G. (2007). The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities. *Electron. J. Probab.* **12** 951–965. MR2336594 https://doi.org/10.1214/EJP.v12-431

Li, Z. (2011). *Measure-Valued Branching Markov Processes*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2760602 https://doi.org/10.1007/978-3-642-15004-3

Li, Z., Wang, H., Xiong, J. and Zhou, X. (2012). Joint continuity of the solutions to a class of nonlinear SPDEs. *Probab. Theory Related Fields* **153** 441–469. MR2948683 https://doi.org/10.1007/s00440-011-0351-x

Lieberman, G.M. (1996). *Second Order Parabolic Differential Equations*. River Edge, NJ: World Scientific Co., Inc. MR1465184 https://doi.org/10.1142/3302

Mitoma, I. (1985). An ∞-dimensional inhomogeneous Langevin's equation. *J. Funct. Anal.* **61** 342–359. MR0820620 https://doi.org/10.1016/0022-1236(85)90027-8

Mytnik, L. and Perkins, E. (2011). Pathwise uniqueness for stochastic heat equations with Hölder continuous coefficients: The white noise case. *Probab. Theory Related Fields* **149** 1–96. MR2773025 https://doi.org/10.1007/s00440-009-0241-7

Mytnik, L., Perkins, E. and Sturm, A. (2006). On pathwise uniqueness for stochastic heat equations with non-Lipschitz coefficients. *Ann. Probab.* **34** 1910–1959. MR2271487 https://doi.org/10.1214/009117906000000331

Mytnik, L. and Xiong, J. (2015). Well-posedness of the martingale problem for superprocess with interaction. *Illinois J. Math.* **59** 485–497. MR3499521

Neuman, E. (2018). Pathwise uniqueness of the stochastic heat equation with spatially inhomogeneous white noise. *Ann. Probab.* **46** 3090–3187. MR3857853 https://doi.org/10.1214/17-AOP1239

Nualart, D. and Pardoux, É. (1992). White noise driven quasilinear SPDEs with reflection. *Probab. Theory Related Fields* **93** 77–89. MR1172940 https://doi.org/10.1007/BF01195389

Perkins, E. (1995). On the martingale problem for interactive measure-valued branching diffusions. *Mem. Amer. Math. Soc.* **115** vi+89. MR1249422 https://doi.org/10.1090/memo/0549

Perkins, E. (2002). Dawson-Watanabe superprocesses and measure-valued diffusions. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1999)*. *Lecture Notes in Math.* **1781** 125–324. Berlin: Springer. MR1915445

Reimers, M. (1989). One-dimensional stochastic partial differential equations and the branching measure diffusion. *Probab. Theory Related Fields* **81** 319–340. MR0983088 https://doi.org/10.1007/BF00340057

Rippl, T. and Sturm, A. (2013). New results on pathwise uniqueness for the heat equation with colored noise. *Electron. J. Probab.* **18** no. 77, 46. MR3101643 https://doi.org/10.1214/EJP.v18-2506

Rosen, J. (1987). Joint continuity of the intersection local times of Markov processes. *Ann. Probab.* **15** 659–675. MR0885136

Strauss, W.A. (2008). *Partial Differential Equations: An Introduction*, 2nd ed. Chichester: Wiley. MR2398759

Sturm, A. (2003). On convergence of population processes in random environments to the stochastic heat equation with colored noise. *Electron. J. Probab.* **8** no. 6, 39. MR1986838 https://doi.org/10.1214/EJP.v8-129

Walsh, J.B. (1986). An introduction to stochastic partial differential equations. In *École D'été de Probabilités de Saint-Flour, XIV—1984*. *Lecture Notes in Math.* **1180** 265–439. Berlin: Springer. MR0876085 https://doi.org/10.1007/BFb0074920

Xiong, J. (2013a). *Three Classes of Nonlinear Stochastic Partial Differential Equations*. Hackensack, NJ: World Scientific Co. Pte. Ltd. MR3235846 https://doi.org/10.1142/8728

Xiong, J. (2013b). Super-Brownian motion as the unique strong solution to an SPDE. *Ann. Probab.* **41** 1030–1054. MR3077534 https://doi.org/10.1214/12-AOP789

Xiong, J. and Yang, X. (2017). Strong existence and uniqueness to a class of nonlinear SPDEs driven by Gaussian colored noises. *Statist. Probab. Lett.* **129** 113–119. MR3688523 https://doi.org/10.1016/j.spl.2017.05.012

Xiong, J. and Yang, X. (2019). Existence and pathwise uniqueness to an SPDE driven by α-stable colored noise. *Stochastic Process. Appl.* **129** 2681–2722. MR3980141 https://doi.org/10.1016/j.spa.2018.08.003

Xu, T. and Zhang, T. (2009). White noise driven SPDEs with reflection: Existence, uniqueness and large deviation principles. *Stochastic Process. Appl.* **119** 3453–3470. MR2568282 https://doi.org/10.1016/j.spa.2009.06.005

Yang, J. and Zhang, T. (2014). Existence and uniqueness of invariant measures for SPDEs with two reflecting walls. *J. Theoret. Probab.* **27** 863–877. MR3245987 https://doi.org/10.1007/s10959-012-0448-x

Zhang, T. (2010). White noise driven SPDEs with reflection: Strong Feller properties and Harnack inequalities. *Potential Anal.* **33** 137–151. MR2658979 https://doi.org/10.1007/s11118-009-9162-4

Zhang, T. (2011). Systems of stochastic partial differential equations with reflection: Existence and uniqueness. *Stochastic Process. Appl.* **121** 1356–1372. MR2794980 https://doi.org/10.1016/j.spa.2011.02.003

Bernoulli **29**(4), 2023, 3013–3032
https://doi.org/10.3150/22-BEJ1572

Gibbsianness and non-Gibbsianness for Bernoulli lattice fields under removal of isolated sites

BENEDIKT JAHNEL[1,2,a] and CHRISTOF KÜLSKE[3,b]

[1]*Technische Universität Braunschweig, Institut für Mathematische Stochastik, Universitätsplatz 2, 38106 Braunschweig, Germany*
[2]*Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39 10117 Berlin, Germany,*
[a]*Benedikt.Jahnel@wias-berlin.de*
[3]*Ruhr-Universität Bochum, Fakultät für Mathematik, 44801 Bochum, Germany,*
[b]*Christof.Kuelske@ruhr-uni-bochum.de*

We consider the i.i.d. Bernoulli field μ_p on $\mathbb{Z}^d$ with occupation density $p \in [0,1]$. To each realization of the set of occupied sites we apply a thinning map that removes all occupied sites that are isolated in graph distance. We show that, while this map seems non-invasive for large p, as it changes only a small fraction $p(1-p)^{2d}$ of sites, there is $p(d) < 1$ such that for all $p \in (p(d), 1)$ the resulting measure is a non-Gibbsian measure, i.e., it does not possess a continuous version of its finite-volume conditional probabilities. On the other hand, for small p, the Gibbs property is preserved.

Keywords: Gibbsianness; Gibbs-uniqueness; Bernoulli field; local thinning; two-layer representation; Dobrushin uniqueness; Peierls' argument

References

[1] Adler, J. (1991). Bootstrap percolation. *Phys. A, Stat. Mech. Appl.* **171** 453–470.
[2] Baccelli, F. and Bermolen, P. (2012). Extremal versus additive Matérn point processes. *Queueing Syst.* **71** 179–197. MR2925795 https://doi.org/10.1007/s11134-012-9280-3
[3] Ball, K. (2005). Poisson thinning by monotone factors. *Electron. Commun. Probab.* **10** 60–69. MR2133893 https://doi.org/10.1214/ECP.v10-1134
[4] Beffara, V. and Duminil-Copin, H. (2013). Planar percolation with a glimpse of Schramm-Loewner evolution. *Probab. Surv.* **10** 1–50. MR3161674 https://doi.org/10.1214/11-PS186
[5] Bergmann, S., Kissel, S. and Külske, C. (2023). Dynamical Gibbs–non-Gibbs transitions in Widom–Rowlinson models on trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **59** 325–344. https://doi.org/10.1214/22-AIHP1242
[6] Bethuelsen, S.A. and Conache, D. (2018). One-sided continuity properties for the Schonmann projection. *J. Stat. Phys.* **172** 1147–1163. MR3830302 https://doi.org/10.1007/s10955-018-2092-z
[7] Blaszczyszyn, B. and Keeler, H. (2019). Determinantal thinning of point processes with network learning applications. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)* 1–8. IEEE.
[8] Bollobás, B. and Riordan, O. (2006). *Percolation*. New York: Cambridge Univ. Press. MR2283880 https://doi.org/10.1017/CBO9781139167383
[9] Bovier, A. (2006). *Statistical Mechanics of Disordered Systems*. *Cambridge Series in Statistical and Probabilistic Mathematics* **18**. Cambridge: Cambridge Univ. Press. MR2252929 https://doi.org/10.1017/CBO9780511616808
[10] Brémaud, P. (1979). Optimal thinning of a point process. *SIAM J. Control Optim.* **17** 222–230. MR0525023 https://doi.org/10.1137/0317017

[11] Bricmont, J., Kupiainen, A. and Lefevere, R. (1998). Renormalization group pathologies and the definition of Gibbs states. *Comm. Math. Phys.* **194** 359–388. MR1627665 https://doi.org/10.1007/s002200050362
[12] Brown, T. (1979). Position dependent and stochastic thinning of point processes. *Stochastic Process. Appl.* **9** 189–193. MR0548838 https://doi.org/10.1016/0304-4149(79)90030-9
[13] Elliott, R.J., Aggoun, L. and Moore, J.B. (1995). *Hidden Markov Models*. *Applications of Mathematics (New York)* **29**. New York: Springer. MR1323178
[14] Engler, N., Jahnel, B. and Külske, C. (2022). Gibbsianness of locally thinned random fields. *Markov Process. Related Fields* **28** 185–214.
[15] Fernández, R., den Hollander, F. and Martínez, J. (2014). Variational description of Gibbs-non-Gibbs dynamical transitions for spin-flip systems with a Kac-type interaction. *J. Stat. Phys.* **156** 203–220. MR3215620 https://doi.org/10.1007/s10955-014-1004-0
[16] Geman, D. (1990). Random fields and inverse problems in imaging. In *École D'été de Probabilités de Saint-Flour XVIII—1988*. *Lecture Notes in Math.* **1427** 113–193. Berlin: Springer. MR1100283 https://doi.org/10.1007/BFb0103042
[17] Georgii, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. Berlin: de Gruyter. MR2807681 https://doi.org/10.1515/9783110250329
[18] Griffiths, R.B. and Pearce, P.A. (1978). Position-space renormalization-group transformations: Some proofs and some problems. *Phys. Rev. Lett.* **41** 917.
[19] Griffiths, R.B. and Pearce, P.A. (1979). Mathematical properties of position-space renormalization-group transformations. *J. Stat. Phys.* **20** 499–545. MR0533527 https://doi.org/10.1007/BF01012897
[20] Grimmett, G. (1997). Percolation and disordered systems. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1996)*. *Lecture Notes in Math.* **1665** 153–300. Berlin: Springer. MR1490045 https://doi.org/10.1007/BFb0092620
[21] Grimmett, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **321**. Berlin: Springer. MR1707339 https://doi.org/10.1007/978-3-662-03981-6
[22] Guyon, X. (1995). *Random Fields on a Network*. *Probability and Its Applications (New York)*. New York: Springer. MR1344683
[23] Häggström, O. and Külske, C. (2004). Gibbs properties of the fuzzy Potts model on trees and in mean field. *Markov Process. Related Fields* **10** 477–506. MR2097868
[24] Henning, F., Kraaij, R.C. and Külske, C. (2019). Gibbs–non-Gibbs transitions in the fuzzy Potts model with a Kac-type interaction: Closing the Ising gap. *Bernoulli* **25** 2051–2074. MR3961240 https://doi.org/10.3150/18-BEJ1045
[25] Isham, V. (1980). Dependent thinning of point processes. *J. Appl. Probab.* **17** 987–995. MR0587199 https://doi.org/10.2307/3213208
[26] Jahnel, B. and König, W. (2020). *Probabilistic Methods in Telecommunications*. *Compact Textbooks in Mathematics*. Cham: Birkhäuser/Springer. MR4182426 https://doi.org/10.1007/978-3-030-36090-0
[27] Jahnel, B. and Külske, C. (2017). Sharp thresholds for Gibbs–non-Gibbs transitions in the fuzzy Potts model with a Kac-type interaction. *Bernoulli* **23** 2808–2827. MR3648046 https://doi.org/10.3150/16-BEJ828
[28] Jahnel, B. and Külske, C. (2017). The Widom-Rowlinson model under spin flip: Immediate loss and sharp recovery of quasilocality. *Ann. Appl. Probab.* **27** 3845–3892. MR3737939 https://doi.org/10.1214/17-AAP1298
[29] Kraaij, R., Redig, F. and van Zuijlen, W. (2021). A Hamilton-Jacobi point of view on mean-field Gibbs–non-Gibbs transitions. *Trans. Amer. Math. Soc.* **374** 5287–5329. MR4293773 https://doi.org/10.1090/tran/8408
[30] Külske, C. and Le Ny, A. (2007). Spin-flip dynamics of the Curie-Weiss model: Loss of Gibbsianness with possibly broken symmetry. *Comm. Math. Phys.* **271** 431–454. MR2287911 https://doi.org/10.1007/s00220-007-0201-y
[31] Külske, C., Le Ny, A. and Redig, F. (2004). Relative entropy and variational properties of generalized Gibbsian measures. *Ann. Probab.* **32** 1691–1726. MR2060315 https://doi.org/10.1214/009117904000000342
[32] Külske, C. and Rozikov, U.A. (2017). Fuzzy transformations and extremality of Gibbs measures for the Potts model on a Cayley tree. *Random Structures Algorithms* **50** 636–678. MR3660523 https://doi.org/10.1002/rsa.20671

[33] Kumagai, T. (2014). *Random Walks on Disordered Media and Their Scaling Limits*. *Lecture Notes in Math.* **2101**. Cham: Springer. MR3156983 https://doi.org/10.1007/978-3-319-03152-1

[34] Last, G. (1993). On dependent marking and thinning of point processes. *Stochastic Process. Appl.* **45** 73–94. MR1204862 https://doi.org/10.1016/0304-4149(93)90061-8

[35] Le Ny, A. (2013). Almost Gibbsianness and parsimonious description of the decimated 2d-Ising model. *J. Stat. Phys.* **152** 305–335. MR3082651 https://doi.org/10.1007/s10955-013-0773-1

[36] Le Ny, A. and Redig, F. (2002). Short time conservation of Gibbsianness under local stochastic evolutions. *J. Stat. Phys.* **109** 1073–1090. MR1938286 https://doi.org/10.1023/A:1020428711181

[37] Matérn, B. (1960). Spatial variation stochastic models and their application to some problems in forest surveys and other sampling investigations. *Medd. Statens Skogsforsk. Inst.* **49**.

[38] Møller, J., Huber, M.L. and Wolpert, R.L. (2010). Perfect simulation and moment properties for the Matérn type III process. *Stochastic Process. Appl.* **120** 2142–2158. MR2684739 https://doi.org/10.1016/j.spa.2010.06.002

[39] Møller, J. and Schoenberg, F.P. (2010). Thinning spatial point processes into Poisson processes. *Adv. in Appl. Probab.* **42** 347–358. MR2675106 https://doi.org/10.1239/aap/1275055232

[40] Rœlly, S. and Ruszel, W.M. (2014). Propagation of Gibbsianness for infinite-dimensional diffusions with space-time interaction. *Markov Process. Related Fields* **20** 653–674. MR3308572

[41] Rolski, T. and Szekli, R. (1991). Stochastic ordering and thinning of point processes. *Stochastic Process. Appl.* **37** 299–312. MR1102876 https://doi.org/10.1016/0304-4149(91)90049-I

[42] Schonmann, R.H. (1989). Projections of Gibbs measures may be non-Gibbsian. *Comm. Math. Phys.* **124** 1–7. MR1012855

[43] van Enter, A.C.D., Ermolaev, V.N., Iacobelli, G. and Külske, C. (2012). Gibbs-non-Gibbs properties for evolving Ising models on trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 774–791. MR2976563 https://doi.org/10.1214/11-AIHP421

[44] van Enter, A.C.D., Fernández, R., den Hollander, F. and Redig, F. (2002). Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures. *Comm. Math. Phys.* **226** 101–130. MR1889994 https://doi.org/10.1007/s002200200605

[45] van Enter, A.C.D., Fernández, R. and Sokal, A.D. (1993). Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory. *J. Stat. Phys.* **72** 879–1167. MR1241537 https://doi.org/10.1007/BF01048183

[46] van Enter, A.C.D., and Le Ny, A. (2017). Decimation of the Dyson-Ising ferromagnet. *Stochastic Process. Appl.* **127** 3776–3791. MR3707245 https://doi.org/10.1016/j.spa.2017.03.007

[47] van Enter, A.C.D., Maes, C., Schonmann, R.H. and Shlosman, S. (2000). The Griffiths singularity random field. In *On Dobrushin's Way. From Probability Theory to Statistical Physics*. *Amer. Math. Soc. Transl. Ser. 2* **198** 51–58. Providence, RI: Amer. Math. Soc. MR1766342 https://doi.org/10.1090/trans2/198/04

[48] Wilson, K. and Kogut, J. (1974). The renormalization group and the ε expansion. *Phys. Rep.* **12** 75–199.

Bernoulli **29**(4), 2023, 3033–3058
https://doi.org/10.3150/22-BEJ1573

Estimation for the reaction term in semi-linear SPDEs under small diffusivity

SASCHA GAUDLITZ[a] and MARKUS REISS[b]

Institut für Mathematik, Humboldt-Universität zu Berlin, [a]*sascha.gaudlitz@hu-berlin.de,*
[b]*mreiss@math.hu-berlin.de*

We consider the estimation of a non-linear reaction term in the stochastic heat or more generally in a semi-linear stochastic partial differential equation (SPDE). Consistent inference is achieved by studying a small diffusivity level, which is realistic in applications. Our main result is a central limit theorem for the estimation error of a parametric estimator, from which confidence intervals can be constructed. Statistical efficiency is demonstrated by establishing local asymptotic normality. The estimation method is extended to local observations in time and space, which allows for non-parametric estimation of a reaction intensity varying in time and space. Furthermore, discrete observations in time and space can be handled. The statistical analysis requires advanced tools from stochastic analysis like Malliavin calculus for SPDEs, the infinite-dimensional Gaussian Poincaré inequality and regularity results for SPDEs in L^p-interpolation spaces.

Keywords: Discrete observations; fractional heat equation; LAN property; maximum likelihood estimation; non-parametric estimation; Poincaré inequality; splitting trick

References

Alonso, S., Stange, M. and Beta, C. (2018). Modeling random crawling, membrane deformation and intracellular polarity of motile amoeboid cells. *PLoS ONE* **13** e0201977. https://doi.org/10.1371/journal.pone.0201977

Altmeyer, R., Cialenco, I. and Pasemann, G. (2020). Parameter estimation for semilinear SPDEs from local measurements. Available at arXiv:2004.14728v3.

Altmeyer, R. and Reiß, M. (2021). Nonparametric estimation for linear SPDEs from local measurements. *Ann. Appl. Probab.* **31** 1–38. MR4254472 https://doi.org/10.1214/20-aap1581

Altmeyer, R., Bretschneider, T., Janák, J. and Reiß, M. (2022). Parameter estimation in an SPDE model for cell repolarization. *SIAM/ASA J. Uncertain. Quantificat.* **10** 179–199. MR4376300 https://doi.org/10.1137/20M1373347

Bibinger, M. and Trabs, M. (2020). Volatility estimation for stochastic PDEs using high-frequency observations. *Stochastic Process. Appl.* **130** 3005–3052. MR4080737 https://doi.org/10.1016/j.spa.2019.09.002

Chong, C. (2020). High-frequency analysis of parabolic stochastic PDEs. *Ann. Statist.* **48** 1143–1167. MR4102691 https://doi.org/10.1214/19-AOS1841

Cialenco, I. (2018). Statistical inference for SPDEs: An overview. *Stat. Inference Stoch. Process.* **21** 309–329. MR3824970 https://doi.org/10.1007/s11203-018-9177-9

Cialenco, I. and Glatt-Holtz, N. (2011). Parameter estimation for the stochastically perturbed Navier-Stokes equations. *Stochastic Process. Appl.* **121** 701–724. MR2770904 https://doi.org/10.1016/j.spa.2010.12.007

Cialenco, I. and Huang, Y. (2020). A note on parameter estimation for discretely sampled SPDEs. *Stoch. Dyn.* **20** 2050016. MR4101083 https://doi.org/10.1142/S0219493720500161

Da Prato, G. and Debussche, A. (2003). Strong solutions to the stochastic quantization equations. *Ann. Probab.* **31** 1900–1916. MR2016604 https://doi.org/10.1214/aop/1068646370

Da Prato, G. and Zabczyk, J. (2014). *Stochastic Equations in Infinite Dimensions*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **152**. Cambridge: Cambridge Univ. Press. MR3236753 https://doi.org/10.1017/CBO9781107295513

1350-7265

Di Nezza, E., Palatucci, G. and Valdinoci, E. (2012). Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136** 521–573. MR2944369 https://doi.org/10.1016/j.bulsci.2011.12.004

Flemming, S., Font, F., Alonso, S. and Beta, C. (2020). How cortical waves drive fission of motile cells. *Proc. Natl. Acad. Sci. USA* **117** 6330–6338. https://doi.org/10.1073/pnas.1912428117

Gaudlitz, S. and Reiß, M. (2023). Supplement to "Estimation for the reaction term in semi-linear SPDEs under small diffusivity." https://doi.org/10.3150/22-BEJ1573SUPP

Hairer, M. (2009). An introduction to stochastic PDEs. Available at arXiv:0907.4178v1.

Hildebrandt, F. and Trabs, M. (2021). Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations. Available at arXiv:2102.13415v1.

Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. MR1902892 https://doi.org/10.1214/aos/1021379858

Huebner, M., Khasminskii, R. and Rozovskii, B.L. (1993). Two examples of parameter estimation for stochastic partial differential equations. In *Stochastic Processes* 149–160. New York: Springer. MR1427311

Huebner, M. and Rozovskiĭ, B.L. (1995). On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's. *Probab. Theory Related Fields* **103** 143–163. MR1355054 https://doi.org/10.1007/BF01204212

Ibragimov, I.A. (2003). An estimation problem for quasilinear stochastic partial differential equations. *Problemy Peredachi Informatsii* **39** 58–87. MR2101666 https://doi.org/10.1023/A:1023630515274

Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. *Applications of Mathematics* **16**. New York-Berlin: Springer. MR0620321

Ibragimov, I.A. and Khas'minskii, R.Z. (1999). Estimation problems for coefficients of stochastic partial differential equations. Part I. *Theory Probab.* **43** 370–387.

Ibragimov, I.A. and Khas'minskii, R.Z. (2000). Estimation problems for coefficients of stochastic partial differential equations. Part II. *Theory Probab.* **44** 469–494.

Ibragimov, I.A. and Khas'minskii, R.Z. (2001). Estimation problems for coefficients of stochastic partial differential equations. Part III. *Theory Probab.* **45** 210–232.

Janson, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge: Cambridge Univ. Press. MR1474726 https://doi.org/10.1017/CBO9780511526169

Kutoyants, Y.A. (2013). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics,* **167**. London: Springer.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. MR0126903

Liptser, R.S. and Shiryayev, A.N. (1989). *Theory of Martingales*. *Mathematics and Its Applications (Soviet Series)* **49**. Dordrecht: Kluwer Academic. MR1022664 https://doi.org/10.1007/978-94-009-2438-3

Lord, G.J., Powell, C.E. and Shardlow, T. (2014). *An Introduction to Computational Stochastic PDEs*. *Cambridge Texts in Applied Mathematics*. New York: Cambridge Univ. Press. MR3308418 https://doi.org/10.1017/CBO9781139017329

Lototsky, S.V. and Rozovsky, B.L. (2017). *Stochastic Partial Differential Equations*. *Universitext*. Cham: Springer. MR3674586 https://doi.org/10.1007/978-3-319-58647-2

Markussen, B. (2003). Likelihood inference for a discretely observed stochastic partial differential equation. *Bernoulli* **9** 745–762. MR2047684 https://doi.org/10.3150/bj/1066418876

Nourdin, I., Peccati, G. and Reinert, G. (2009). Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.* **257** 593–609. MR2527030 https://doi.org/10.1016/j.jfa.2008.12.017

Nualart, D. (2009). Application of Malliavin calculus to stochastic partial differential equations. In *A Minicourse on Stochastic Partial Differential Equations*. *Lecture Notes in Math.* **1962** 73–109. Berlin: Springer. MR2508774 https://doi.org/10.1007/978-3-540-85994-9_3

Pasemann, G. and Stannat, W. (2020). Drift estimation for stochastic reaction-diffusion systems. *Electron. J. Stat.* **14** 547–579. MR4056266 https://doi.org/10.1214/19-EJS1665

Pasemann, G., Flemming, S., Alonso, S., Beta, C. and Stannat, W. (2021). Diffusivity estimation for activator-inhibitor models: Theory and application to intracellular dynamics of the actin cytoskeleton. *J. Nonlinear Sci.* **31** 59. MR4255684 https://doi.org/10.1007/s00332-021-09714-4

Sanz-Solé, M. (2005). *Malliavin Calculus, with Applications to Stochastic Partial Differential Equations*. *Fundamental Sciences*. EPFL Press: Lausanne. MR2167213

Sanz-Solé, M. and Süß, A. (2013). The stochastic wave equation in high dimensions: Malliavin differentiability and absolute continuity. *Electron. J. Probab.* **18** 64. MR3078023 https://doi.org/10.1214/EJP.v18-2341

Soh, S., Byrska, M., Kandere-Grzybowska, K. and Grzybowski, B.A. (2010). Reaction-diffusion systems in intracellular molecular transport and control. *Angew. Chem., Int. Ed. Engl.* **49** 4170–4198. https://doi.org/10.1002/anie.200905513

Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359 https://doi.org/10.1007/b13794

Yagi, A. (2010). *Abstract Parabolic Evolution Equations and Their Applications*. *Springer Monographs in Mathematics*. Berlin: Springer. MR2573296 https://doi.org/10.1007/978-3-642-04631-5

Bernoulli **29**(4), 2023, 3059–3092
https://doi.org/10.3150/22-BEJ1575

Additive regression with parametric help

HYERIM HONG[1,a], YOUNG KYUNG LEE[2,c] and BYEONG U. PARK[1,b]

[1]*Seoul National University, Seoul, South Korea,* [a]*hhong@snu.ac.kr,* [b]*bupark@snu.ac.kr*
[2]*Kangwon National University, Chuncheon, South Korea,* [c]*youngklee@kangwon.ac.kr*

Additive models have been studied as a way of overcoming theoretical and practical difficulties in estimating a multivariate nonparametric regression function. Several methods have been proposed that ensure the optimal univariate rate one can achieve in estimating univariate nonparametric functions. In this paper a new method is proposed which reduces the constant factor in the first-order approximation of the average squared error of the most successful existing method. The new estimator is based on an orthogonal decomposition of the underlying regression function, with an arbitrarily chosen parametric family, under a special inner product structure arising from the bias formula of the estimator. It is shown that the proposed method entails reduction in the constant factor of the leading bias of the existing method while it retains the same first-order variance. These theoretical findings are confirmed in Monte Carlo experiments.

Keywords: Additive model; bias reduction; local linear smoothing; parametric help; smooth backfitting

References

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

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

Blot, J. and Cieutat, P. (2016). Completeness of sums of subspaces of bounded functions and applications. *Commun. Math. Anal.* **19** 43–61. MR3501515 https://doi.org/10.1177/003754977201900208

Boente, G. and Martínez, A. (2017). Marginal integration M-estimators for additive models. *TEST* **26** 231–260. MR3650526 https://doi.org/10.1007/s11749-016-0508-0

Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115

Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi-likelihood with a parametric guide. *Ann. Statist.* **37** 4153–4183. MR2572456 https://doi.org/10.1214/09-AOS713

Febrero-Bande, M. and González-Manteiga, W. (2013). Generalized additive models for functional data. *TEST* **22** 278–292. MR3062258 https://doi.org/10.1007/s11749-012-0308-0

Glad, I.K. (1998). Parametrically guided non-parametric regression. *Scand. J. Stat.* **25** 649–668. MR1666776 https://doi.org/10.1111/1467-9469.00127

Gozalo, P. and Linton, O. (2000). Local nonlinear least squares: Using parametric information in nonparametric regression. *J. Econometrics* **99** 63–106. MR1793389 https://doi.org/10.1016/S0304-4076(00)00031-2

Han, K., Müller, H.-G. and Park, B.U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233–1265. MR3706793 https://doi.org/10.3150/16-BEJ898

Han, K., Müller, H.-G. and Park, B.U. (2020). Additive functional regression for densities as responses. *J. Amer. Statist. Assoc.* **115** 997–1010. MR4107695 https://doi.org/10.1080/01621459.2019.1604365

Han, K. and Park, B.U. (2018). Smooth backfitting for errors-in-variables additive models. *Ann. Statist.* **46** 2216–2250. MR3845016 https://doi.org/10.1214/17-AOS1617

Hjort, N.L. and Glad, I.K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* **23** 882–904. MR1345205 https://doi.org/10.1214/aos/1176324627

1350-7265

Huang, L.-S. and Yu, C.-H. (2019). Classical backfitting for smooth-backfitting additive models. *J. Comput. Graph. Statist.* **28** 386–400. MR3974888 https://doi.org/10.1080/10618600.2018.1530120
Jeon, J.M. and Park, B.U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697. MR4152117 https://doi.org/10.1214/19-AOS1902
Jeon, J.M., Park, B.U. and Van Keilegom, I. (2021a). Additive regression for non-Euclidean responses and predictors. *Ann. Statist.* **49** 2611–2641. MR4338377 https://doi.org/10.1214/21-aos2048
Jeon, J.M., Park, B.U. and Van Keilegom, I. (2021b). Additive regression for predictors of various natures and possibly incomplete Hilbertian responses. *Electron. J. Stat.* **15** 1473–1548. MR4255309 https://doi.org/10.1214/21-ejs1823
Jeon, J.M., Lee, Y.K., Mammen, E. and Park, B.U. (2022). Locally polynomial Hilbertian additive regression. *Bernoulli* **28** 2034–2066. MR4411521 https://doi.org/10.3150/21-bej1410
Lee, Y.K. (2004). On marginal integration method in nonparametric regression. *J. Korean Statist. Soc.* **33** 435–447. MR2126371
Lee, Y.K., Mammen, E. and Park, B.U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883. MR2722458 https://doi.org/10.1214/10-AOS808
Lee, Y.K., Mammen, E. and Park, B.U. (2012). Flexible generalized varying coefficient regression models. *Ann. Statist.* **40** 1906–1933. MR3015048 https://doi.org/10.1214/12-AOS1026
Lee, Y.K., Mammen, E., Nielsen, J.P. and Park, B.U. (2020). Nonparametric regression with parametric help. *Electron. J. Stat.* **14** 3845–3868. MR4164866 https://doi.org/10.1214/20-EJS1760
Lee, Y.K., Park, B.U., Hong, H. and Kim, D. (2022). Estimation of Hilbertian varying coefficient models. *Stat. Interface* **15** 129–149. MR4363348
Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93
Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Statist.* **36** 686–718. MR2396812 https://doi.org/10.1214/009053607000000848
Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490. MR1742496 https://doi.org/10.1214/aos/1017939137
Mammen, E. and Park, B.U. (2006). A simple smooth backfitting method for additive models. *Ann. Statist.* **34** 2252–2271. MR2291499 https://doi.org/10.1214/009053606000000696
Nielsen, J.P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 43–61. MR2136638 https://doi.org/10.1111/j.1467-9868.2005.00487.x
Opsomer, J.D. (2000). Asymptotic properties of backfitting estimators. *J. Multivariate Anal.* **73** 166–179. MR1763322 https://doi.org/10.1006/jmva.1999.1868
Opsomer, J.D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25** 186–211. MR1429922 https://doi.org/10.1214/aos/1034276626
Park, B.U., Mammen, E., Lee, Y.K. and Lee, E.R. (2015). Varying coefficient regression models: A review and new developments. *Int. Stat. Rev.* **83** 36–64. MR3341079 https://doi.org/10.1111/insr.12029
Park, B.U., Chen, C.-J., Tao, W. and Müller, H.-G. (2018). Singular additive models for function to function regression. *Statist. Sinica* **28** 2497–2520. MR3839871
Sperlich, S., Linton, O. and Härdle, W. (1999). Integration and backfitting methods in additive models – Finite sample properties and comparison. *TEST* **8** 419–458.
Talamakrouni, M., El Ghouch, A. and Van Keilegom, I. (2015). Guided censored regression. *Scand. J. Stat.* **42** 214–233. MR3318033 https://doi.org/10.1111/sjos.12103
Talamakrouni, M., Van Keilegom, I. and El Ghouch, A. (2016). Parametrically guided nonparametric density and hazard estimation with censored data. *Comput. Statist. Data Anal.* **93** 308–323. MR3406214 https://doi.org/10.1016/j.csda.2015.01.009
Xu, J. and Zikatanov, L. (2002). The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Amer. Math. Soc.* **15** 573–597. MR1896233 https://doi.org/10.1090/S0894-0347-02-00398-3
Yang, S.J. and Park, B.U. (2014). Efficient estimation for partially linear varying coefficient models when coefficient functions have different smoothing variables. *J. Multivariate Anal.* **126** 100–113. MR3173084 https://doi.org/10.1016/j.jmva.2014.01.004

Yu, K., Mammen, E. and Park, B.U. (2011). Semi-parametric regression: Efficiency gains from modeling the nonparametric part. *Bernoulli* **17** 736–748. MR2787613 https://doi.org/10.3150/10-BEJ296

Yu, K., Park, B.U. and Mammen, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260. MR2387970 https://doi.org/10.1214/009053607000000596

Zhang, X., Park, B.U. and Wang, J.-L. (2013). Time-varying additive models for longitudinal data. *J. Amer. Statist. Assoc.* **108** 983–998. MR3174678 https://doi.org/10.1080/01621459.2013.778776

Bernoulli **29**(4), 2023, 3093–3113
https://doi.org/10.3150/22-BEJ1576

Riemannian Langevin algorithm for solving semidefinite programs

MUFAN (BILL) LI[1,a] and MURAT A. ERDOGDU[2,b]

[1]*Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [a]*mufan.li@mail.utoronto.ca*
[2]*Department of Computer Science and Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [b]*erdogdu@cs.toronto.edu*

We propose a Langevin diffusion-based algorithm for non-convex optimization and sampling on a product manifold of spheres. Under a logarithmic Sobolev inequality, we establish a guarantee for finite iteration convergence to the Gibbs distribution in terms of Kullback–Leibler divergence. We show that with an appropriate temperature choice, the suboptimality gap to the global minimum is guaranteed to be arbitrarily small with high probability.

As an application, we consider the Burer–Monteiro approach for solving a semidefinite program (SDP) with diagonal constraints, and analyze the proposed Langevin algorithm for optimizing the non-convex objective. In particular, we establish a logarithmic Sobolev inequality for the Burer–Monteiro problem when there are no spurious local minima, but under the presence saddle points. Combining the results, we then provide a global optimality guarantee for the SDP and the Max-Cut problem. More precisely, we show that the Langevin algorithm achieves ϵ accuracy with high probability in $\widetilde{\Omega}(\epsilon^{-4.5})$ iterations.

Keywords: Riemannian Langevin algorithm; non-convex optimization; logarithmic Sobolev inequality; semidefinite programs; Burer–Monteiro problem

References

Absil, P.-A., Mahony, R. and Sepulchre, R. (2009). *Optimization Algorithms on Matrix Manifolds*. Princeton: Princeton Univ. Press.

Alizadeh, F., Haeberly, J.-P.A. and Overton, M.L. (1997). Complementarity and nondegeneracy in semidefinite programming. *Math. Program.* **77** 111–128.

Arora, S., Hazan, E. and Kale, S. (2005). Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)* 339–348. IEEE.

Arora, S. and Kale, S. (2007). A combinatorial, primal-dual approach to semidefinite programs [extended abstract]. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* 227–236. New York: ACM. MR2402446 https://doi.org/10.1145/1250790.1250823

Bakry, D. and Émery, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités, XIX, 1983/84. Lecture Notes in Math.* **1123** 177–206. Berlin: Springer. MR0889476 https://doi.org/10.1007/BFb0075847

Bakry, D., Gentil, I. and LeDoux, M. (2013). *Analysis and Geometry of Markov Diffusion Operators*. Berlin: Springer.

Bakry, D., Barthe, F., Cattiaux, P. and Guillin, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66. MR2386063 https://doi.org/10.1214/ECP.v13-1352

Barvinok, A.I. (1995). Problems of distance geometry and convex properties of quadratic maps. *Discrete Comput. Geom.* **13** 189–202. MR1314962 https://doi.org/10.1007/BF02574037

Belloni, A., Liang, T., Narayanan, H. and Rakhlin, A. (2015). Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory* 240–265.

Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855 https://doi.org/10.3150/bj/1165269151

1350-7265

Betancourt, M.J. (2013). Generalizing the no-u-turn sampler to riemannian manifolds. arXiv preprint, arXiv:1304.1920.

Block, A., Mroueh, Y., Rakhlin, A. and Ross, J. (2020). Fast mixing of multi-scale langevin dynamics underthe manifold hypothesis. arXiv preprint, arXiv:2006.11166.

Boumal, N., Voroninski, V. and Bandeira, A. (2016). The non-convex burer-monteiro approach works on smooth semidefinite programs. In *Advances in Neural Information Processing Systems* 2757–2765.

Bovier, A. and den Hollander, F. (2016). *Metastability: A Potential-Theoretic Approach* **351**. Berlin: Springer.

Brosse, N., Durmus, A., Moulines, É. and Pereyra, M. (2017). Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo. In *Conference on Learning Theory* 319–342. PMLR.

Bubeck, S., Eldan, R. and Lehec, J. (2018). Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* **59** 757–783. MR3802303 https://doi.org/10.1007/s00454-018-9992-1

Burer, S. and DC Monteiro, R. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.* **95** 329–357.

Byrne, S. and Girolami, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scand. J. Stat.* **40** 825–845. MR3145120 https://doi.org/10.1111/sjos.12036

Cattiaux, P. and Guillin, A. (2017). Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity. *J. Funct. Anal.* **272** 2361–2391. MR3603301 https://doi.org/10.1016/j.jfa.2016.10.003

Cattiaux, P., Guillin, A. and Wu, L.-M. (2010). A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* **148** 285–304. MR2653230 https://doi.org/10.1007/s00440-009-0231-9

Cattiaux, P., Guillin, A. and Zitt, P.A. (2013). Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 95–118. MR3060149 https://doi.org/10.1214/11-AIHP447

Cheng, X. and Bartlett, P. (2018). Convergence of Langevin mcmc in kl-divergence. In *Algorithmic Learning Theory* 186–211. PMLR.

Cheng, X., Chatterji, N.S., Abbasi-Yadkori, Y., Bartlett, P.L. and Jordan, M.I. (2018). Sharp convergence rates for langevin dynamics in the nonconvex setting. arXiv preprint, arXiv:1805.01648.

Cifuentes, D. (2021). On the Burer-Monteiro method for general semidefinite programs. *Optim. Lett.* **15** 2299–2309. MR4300033 https://doi.org/10.1007/s11590-021-01705-4

Cifuentes, D. and Moitra, A. (2019). Polynomial time guarantees for the burer-monteiro method. arXiv preprint, arXiv:1912.01745.

Cobb, A.D., Güneş Baydin, A., Markham, A. and Roberts, S.J. (2019). Introducing an explicit symplectic integration scheme for riemannian manifold hamiltonian monte carlo. arXiv preprint, arXiv:1910.06243.

Dalalyan, A. (2017). Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Conference on Learning Theory* 678–689. PMLR.

Dalalyan, A.S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.* **129** 5278–5311. MR4025705 https://doi.org/10.1016/j.spa.2019.02.016

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov Chains*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

Durmus, A., Majewski, S. and Miasojedow, B. (2019). Analysis of Langevin Monte Carlo via convex optimization. *J. Mach. Learn. Res.* **20** 2666–2711.

Durmus, A. and Moulines, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

Erdogdu, M.A., Deshpande, Y. and Montanari, A. (2017). Inference in graphical models via semidefinite programming hierarchies. *Adv. Neural Inf. Process. Syst.* **30**.

Erdogdu, M.A. and Hosseinzadeh, R. (2020). On the convergence of langevin monte carlo: The interplay between tail growth and smoothness. arXiv preprint, arXiv:2005.13097.

Erdogdu, M.A., Mackey, L. and Shamir, O. (2018). Global non-convex optimization with discretized diffusions. In *Advances in Neural Information Processing Systems* 9671–9680.

Erdogdu, M.A., Ozdaglar, A., Parrilo, P.A. and Vanli, N.D. (2022). Convergence rate of block-coordinate maximization Burer-Monteiro method for solving large SDPs. *Math. Program.* **195** 243–281. MR4499058 https://doi.org/10.1007/s10107-021-01686-3

Garber, D. and Hazan, E. (2011). Approximating semidefinite programs in sublinear time. In *Advances in Neural Information Processing Systems* 1080–1088.

Gelfand, S.B. and Mitter, S.K. (1991). Recursive stochastic algorithms for global optimization in $\mathbf{R}^d$. *SIAM J. Control Optim.* **29** 999–1018. MR1110084 https://doi.org/10.1137/0329055

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

Goemans, M.X. and Williamson, D.P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.* **42** 1115–1145. MR1412228 https://doi.org/10.1145/227683.227684

Holbrook, A., Lan, S., Vandenberg-Rodes, A. and Shahbaba, B. (2018). Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: With application to Bayesian spectral density estimation. *J. Stat. Comput. Simul.* **88** 982–1002. MR3750634 https://doi.org/10.1080/00949655.2017.1416470

Holley, R. and Stroock, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** 1159–1194. MR0893137 https://doi.org/10.1007/BF01011161

Hsu, E.P. (2002). *Stochastic Analysis on Manifolds. Contemporary Mathematics*. Providence: Amer. Math. Soc. ISBN 9780821808023.

Hu, J., Liu, X., Wen, Z.-W. and Yuan, Y.-X. (2020). A brief introduction to manifold optimization. *J. Oper. Res. Soc. China* **8** 199–248. MR4114364 https://doi.org/10.1007/s40305-020-00295-9

Javanmard, A., Montanari, A. and Ricci-Tersenghi, F. (2016). Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA* **113** E2218–E2223. MR3494080 https://doi.org/10.1073/pnas.1523097113

Jenkins, P.A. and Spanò, D. (2017). Exact simulation of the Wright-Fisher diffusion. *Ann. Appl. Probab.* **27** 1478–1509. MR3678477 https://doi.org/10.1214/16-AAP1236

Jordan, R., Kinderlehrer, D. and Otto, F. (1998). The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** 1–17. MR1617171 https://doi.org/10.1137/S0036141096303359

Kalai, A.T. and Vempala, S. (2006). Simulated annealing for convex optimization. *Math. Oper. Res.* **31** 253–266. MR2233996 https://doi.org/10.1287/moor.1060.0194

Lee, J.M. (2019). *Introduction to Riemannian Manifolds*. *Graduate Texts in Mathematics*. Berlin: Springer.

Li, M. and Erdogdu, M.A. (2023). Supplement to "Riemannian Langevin Algorithm for Solving Semidefinite Programs." https://doi.org/10.3150/22-BEJ1576SUPP

Li, X., Wu, Y., Mackey, L. and Erdogdu, M.A. (2019). Stochastic Runge-Kutta accelerates Langevin Monte Carlo and beyond. In *Advances in Neural Information Processing Systems* 7748–7760.

Liu, C., Zhu, J. and Song, Y. (2016). Stochastic gradient geodesic mcmc methods. In *Advances in Neural Information Processing Systems* 3009–3017.

Łojasiewicz, S. (1963). A topological property of real analytic subsets. *Les équations aux Dérivées Partielles* 87–89.

Majumdar, A., Hall, G. and Ahmadi, A.A. (2019). A survey of recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. arXiv preprint, arXiv:1908.05209.

Mei, S., Bai, Y. and Montanari, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

Mei, S., Misiakiewicz, T., Montanari, A. and Oliveira, R.I. (2017). Solving sdps for synchronization and maxcut problems via the grothendieck inequality. arXiv preprint, arXiv:1703.08729.

Menz, G. and Schlichting, A. (2014). Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.* **42** 1809–1884. MR3262493 https://doi.org/10.1214/14-AOP908

Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2509253 https://doi.org/10.1017/CBO9780511626630

Mijatović, A., Mramor, V. and Uribe Bravo, G. (2020). A note on the exact simulation of spherical Brownian motion. *Statist. Probab. Lett.* **165** 108836. MR4113846 https://doi.org/10.1016/j.spl.2020.108836

Moitra, A. and Risteski, A. (2020). Fast convergence for langevin diffusion with matrix manifold structure. arXiv preprint, arXiv:2002.05576.

Nesterov, Y. (2013). *Introductory Lectures on Convex Optimization: A Basic Course* **87**. Berlin: Springer.

Pataki, G. (1998). On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.* **23** 339–358. MR1626662 https://doi.org/10.1287/moor.23.2.339

Patterson, S. and Whye Teh, Y. (2013). Stochastic gradient Riemannian Langevin dynamics on the probability simplex. In *Advances in Neural Information Processing Systems* 3102–3110.

Poljak, B.T. (1963). Gradient methods for minimizing functionals. *Zh. Vychisl. Mat. Mat. Fiz.* **3** 643–653. MR0158568

Pumir, T., Jelassi, S. and Boumal, N. (2018). Smoothed analysis of the low-rank approach for smooth semidefinite programs. In *Advances in Neural Information Processing Systems* 2281–2290.

Raginsky, M., Rakhlin, A. and Telgarsky, M. (2017). Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, the Netherlands, 7–10 July 2017* 1674–1703.

Steurer, D. (2010). Fast SDP algorithms for constraint satisfaction problems. In *Proceedings of the Twenty-First Annual ACM-Siam Symposium on Discrete Algorithms* 684–697. Philadelphia, PA: SIAM. MR2768625

Tat Lee, Y. and Padmanabhan, S. (2019). An $\widetilde{O}(m/\varepsilon^{3.5})$-cost algorithm for semidefinite programs with diagonal constraints. arXiv preprint, arXiv:1903.01859.

Toscani, G. and Villani, C. (2000). On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds. *J. Stat. Phys.* **98** 1279–1309. MR1751701 https://doi.org/10.1023/A:1018623930325

Vempala, S.S. and Wibisono, A. (2019). Rapid convergence of the unadjusted langevin algorithm: Log-sobolev suffices. arXiv preprint, arXiv:1903.08568.

Waldspurger, I. and Waters, A. (2020). Rank optimality for the Burer-Monteiro factorization. *SIAM J. Optim.* **30** 2577–2602. MR4152680 https://doi.org/10.1137/19M1255318

Wang, F.-Y. (1997a). On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups. *Probab. Theory Related Fields* **108** 87–101. MR1452551 https://doi.org/10.1007/s004400050102

Wang, F.-Y. (1997b). Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Probab. Theory Related Fields* **109** 417–424. MR1481127 https://doi.org/10.1007/s004400050137

Wibisono, A. (2019). Proximal langevin algorithm: Rapid convergence under isoperimetry. arXiv preprint, arXiv:1911.01469.

Zhang, K.S., Peyré, G., Fadili, J. and Pereyra, M. (2020). Wasserstein control of mirror langevin monte carlo. arXiv preprint, arXiv:2002.04363.

Bernoulli **29**(4), 2023, 3114–3140
https://doi.org/10.3150/22-BEJ1577

Sequential Gaussian approximation for nonstationary time series in high dimensions

FABIAN MIES[a] and ANSGAR STELAND[b]

Institute of Statistics, RWTH Aachen University, Wüllnerstr. 3, Aachen, Germany,
[a]*mies@stochastik.rwth-aachen.de,* [b]*steland@stochastik.rwth-aachen.de*

Gaussian couplings of partial sum processes are derived for the high-dimensional regime $d = o(n^{1/3})$. The coupling is derived for sums of independent random vectors and subsequently extended to nonstationary time series. Our inequalities depend explicitly on the dimension and on a measure of nonstationarity, and are thus also applicable to arrays of random vectors. To enable high-dimensional statistical inference, a feasible Gaussian approximation scheme is proposed. Applications to sequential testing and change-point detection are described.

Keywords: Strong approximation; Rosenthal inequality; physical dependence measure; bounded variation

References

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

Bai, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020

Berkes, I., Liu, W. and Wu, W.B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley. MR0233396

Bonis, T. (2020). Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem. *Probab. Theory Related Fields* **178** 827–860. MR4168389 https://doi.org/10.1007/s00440-020-00989-4

Bours, M. and Steland, A. (2021). Large-sample approximations and change testing for high-dimensional covariance matrices of multivariate linear time series and factor models. *Scand. J. Stat.* **48** 610–654. MR4275159 https://doi.org/10.1111/sjos.12508

Breiman, L. (1967). On the tail behavior of sums of independent random variables. *Z. Wahrsch. Verw. Gebiete* **9** 20–25. MR0226707 https://doi.org/10.1007/BF00535464

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813 https://doi.org/10.1214/aos/1176350057

Chang, J., Chen, X. and Wu, M. (2021). Central limit theorems for high dimensional dependent data. arXiv:2104.12929 [math, stat].

Chen, Y. (2021). High-Dimensional Online Changepoint Detection. Ph.D. thesis, Univ. Cambridge.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. MR4009488 https://doi.org/10.1093/restud/rdy065

Chernozhukov, V., Chetverikov, D. and Koike, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. arXiv:2012.09513.

1350-7265

Chernozhukov, V., Chetverikov, D., Kato, K. and Koike, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. https://doi.org/10.1214/22-AOS2193

Csörgő, M. and Révész, P. (1975). A new method to prove Strassen type laws of invariance principle. I, II. *Z. Wahrsch. Verw. Gebiete* **31** 255–259. MR0375411 https://doi.org/10.1007/BF00532865

Cui, Y., Levine, M. and Zhou, Z. (2021). Estimation and inference of time-varying auto-covariance under complex trend: A difference-based approach. *Electron. J. Stat.* **15** 4264–4294. MR4312204 https://doi.org/10.1214/21-ejs1893

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

Dahlhaus, R., Richter, S. and Wu, W.B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

Dębicki, K., Hashorva, E. and Ji, L. (2016). Extremes of a class of nonhomogeneous Gaussian random fields. *Ann. Probab.* **44** 984–1012. MR3474465 https://doi.org/10.1214/14-AOP994

Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

Dette, H. and Gösmann, J. (2018). Relevant change points in high dimensional time series. *Electron. J. Stat.* **12** 2578–2636. MR3849896 https://doi.org/10.1214/18-EJS1464

Dette, H., Wu, W. and Zhou, Z. (2019). Change point analysis of correlation in non-stationary time series. *Statist. Sinica* **29** 611–643. MR3931381

Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR0996984 https://doi.org/10.1016/0047-259X(89)90097-3

Eldan, R., Mikulincer, D. and Zhai, A. (2020). The CLT in high dimensions: Quantitative bounds via martingale embedding. *Ann. Probab.* **48** 2494–2524. MR4152649 https://doi.org/10.1214/20-AOP1429

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *Ann. Statist.* **47** 2051–2079. MR3953444 https://doi.org/10.1214/18-AOS1740

Gösmann, J., Stoehr, C., Heiny, J. and Dette, H. (2022). Sequential change point detection in high dimensional time series. *Electron. J. Stat.* **16** 3608–3671. MR4444665 https://doi.org/10.1214/22-ejs2027

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

Jirak, M. (2012). Change-point analysis in increasing dimension. *J. Multivariate Anal.* **111** 136–159. MR2944411 https://doi.org/10.1016/j.jmva.2012.05.007

Jirak, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

Karatzas, I. and Shreve, S.E. (1998). *Brownian Motion and Stochastic Calculus*. New York: Springer. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

Karmakar, S. and Wu, W.B. (2020). Optimal Gaussian approximation for multiple time series. *Statist. Sinica* **30** 1399–1417. MR4257539 https://doi.org/10.5705/ss.202017.0303

Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent *RV*'s and the sample *DF*. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

Kurisu, D., Kato, K. and Shao, X. (2021). Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data. arXiv:2103.10720.

Li, J. and Liao, Z. (2020). Uniform nonparametric inference for time series. *J. Econometrics* **219** 38–51. MR4152784 https://doi.org/10.1016/j.jeconom.2019.09.011

Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280. MR2485027 https://doi.org/10.1016/j.spa.2008.01.012

Liu, W., Xiao, H. and Wu, W.B. (2013). Probability and moment inequalities under dependence. *Statist. Sinica* **23** 1257–1272. MR3114713

Mies, F. (2021). Functional estimation and change detection for nonstationary time series. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2021.1969239

Mies, F. and Steland, A. (2023). Supplement to "Sequential Gaussian approximation for nonstationary time series in high dimensions." https://doi.org/10.3150/22-BEJ1577SUPP

Peligrad, M. and Shao, Q.M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *J. Multivariate Anal.* **52** 140–157. MR1325375 https://doi.org/10.1006/jmva.1995.1008

Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. MR1331198

Schmidt, S.K., Wornowizki, M., Fried, R. and Dehling, H. (2021). An asymptotic test for constancy of the variance under short-range dependence. *Ann. Statist.* **49** 3460–3481. MR4352537 https://doi.org/10.1214/21-aos2092

Steland, A. (2020). Testing and estimating change-points in the covariance matrix of a high-dimensional time series. *J. Multivariate Anal.* **177** 104582, 24. MR4068598 https://doi.org/10.1016/j.jmva.2019.104582

Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

Wu, W.B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528 https://doi.org/10.5705/ss.2008.223

Zaitsev, A.Y. (2007). Estimates for the rate of strong approximation in the multidimensional invariance principle. *J. Math. Sci.* **145** 4856–4865. MR2355400 https://doi.org/10.1007/s10958-007-0319-7

Zaitsev, A.Y. (2013). The accuracy of strong Gaussian approximation for sums of independent random vectors. *Russian Math. Surveys* **68** 721–761. MR3154816 https://doi.org/10.1070/rm2013v068n04abeh004851

Zhai, A. (2018). A high-dimensional CLT in W_2 distance with near optimal convergence rate. *Probab. Theory Related Fields* **170** 821–845. MR3773801 https://doi.org/10.1007/s00440-017-0771-3

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697 https://doi.org/10.3150/17-BEJ939

Zhang, D. and Wu, W.B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

Zhang, N.R., Siegmund, D.O., Ji, H. and Li, J.Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636

Bernoulli **29**(4), 2023, 3141–3170
https://doi.org/10.3150/22-BEJ1578

Diffusion means in geometric spaces

BENJAMIN ELTZNER[1,a], PERNILLE E.H. HANSEN[2,b],
STEPHAN F. HUCKEMANN[3,d] and STEFAN SOMMER[2,c]

[1]*Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany,* [a]*benjamin.eltzner@mpinat.mpg.de*
[2]*Department of Computer Science, University of Copenhagen, Copenhagen, Denmark,* [b]*pehh@di.ku.dk,*
[c]*sommer@di.ku.dk*
[3]*Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Göttingen, Germany,*
[d]*huckeman@math.uni-goettingen.de*

We introduce a location statistic for distributions on non-linear geometric spaces, the diffusion mean, serving as an extension and an alternative to the Fréchet mean. The diffusion mean arises as the generalization of Gaussian maximum likelihood analysis to non-linear spaces by maximizing the likelihood of a Brownian motion. The diffusion mean depends on a time parameter t, which admits the interpretation of the allowed variance of the diffusion. The diffusion t-mean of a distribution X is the most likely origin of a Brownian motion at time t, given the end-point distribution X. We give a detailed description of the asymptotic behavior of the diffusion estimator and provide sufficient conditions for the diffusion estimator to be strongly consistent. Particularly, we present a smeary central limit theorem for diffusion means and we show that joint estimation of the mean and diffusion variance rules out smeariness in all directions simultaneously in general situations. Furthermore, we investigate properties of the diffusion mean for distributions on the sphere $\mathcal{S}^m$. Experimentally, we consider simulated data and data from magnetic pole reversals, all indicating similar or improved convergence rate compared to the Fréchet mean. Here, we additionally estimate t and consider its effects on smeariness and uniqueness of the diffusion mean for distributions on the sphere.

Keywords: Diffusion mean; generalized Fréchet mean; geometric statistics; maximum likelihood estimation; spherical statistics

References

Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5

Alonso-Orán, D., Chamizo, F., Martínez, A.D. and Mas, A. (2019). Pointwise monotonicity of heat kernels. *Rev. Math. Complut.* https://doi.org/10.1007/s13163-021-00417-8

Arede, M.T. (1991). Heat kernels on Lie groups. In *Stochastic Analysis and Applications (Lisbon, 1989). Progress in Probability* **26** 52–62. Boston, MA: Birkhäuser. MR1168067

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

Chavel, I. (1984). *Eigenvalues in Riemannian Geometry*. *Pure and Applied Mathematics* **115**. Orlando, FL: Academic Press. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk. MR0768584

Cheng, S.Y., Li, P. and Yau, S.T. (1981). On the upper estimate of the heat kernel of a complete Riemannian manifold. *Amer. J. Math.* **103** 1021–1063. MR0630777 https://doi.org/10.2307/2374257

Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116** 1660–1675. MR2269221 https://doi.org/10.1016/j.spa.2006.04.004

Dryden, I.L. and Mardia, K.V. (2016). *Statistical Shape Analysis*, 2nd ed. Chichester: Wiley. MR1646114

Eltzner, B. (2022). Geometrical smeariness—a new phenomenon of Fréchet means. *Bernoulli* **28** 239–254. MR4337704 https://doi.org/10.3150/21-bej1340

1350-7265

Eltzner, B. and Huckemann, S. (2017). Bootstrapping descriptors for non-Euclidean data. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **10589** 12–19. Cham: Springer. MR3737707 https://doi.org/10.1007/978-3-319-68445-1_2

Eltzner, B. and Huckemann, S.F. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47** 3360–3381. MR4025745 https://doi.org/10.1214/18-AOS1781

Eltzner, B., Hansen, P.E.H., Huckemann, S.F. and Sommer, S. (2023). Supplement to "Diffusion means in geometric spaces." https://doi.org/10.3150/22-BEJ1578SUPP

Evans, S.N. and Jaffe, A.Q. (2020). Strong laws of large numbers for Fréchet means. arXiv:2012.12859 [math, stat].

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

Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725 https://doi.org/10.1007/BF02291478

Grigor'yan, A. and Noguchi, M. (1998). The heat kernel on hyperbolic space. *Bull. Lond. Math. Soc.* **30** 643–650. MR1642767 https://doi.org/10.1112/S0024609398004780

Groisser, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics* **77** 31–60. MR2138772 https://doi.org/10.1080/17442500512331341059

Hansen, P., Eltzner, B. and Sommer, S. (2021). Diffusion means and heat kernel on manifolds. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **12829** 111–118. Cham: Springer. MR4424315 https://doi.org/10.1007/978-3-030-80209-7_13

Hendriks, H. and Landsman, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statist. Probab. Lett.* **26** 169–178. MR1381468 https://doi.org/10.1016/0167-7152(95)00007-0

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** 227–243. MR1659156 https://doi.org/10.1006/jmva.1998.1776

Hotz, T. (2013). Extrinsic vs intrinsic means on the circle. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **8085** 433–440. Heidelberg: Springer. MR3126073 https://doi.org/10.1007/978-3-642-40020-9_47

Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.* **67** 177–193. MR3297863 https://doi.org/10.1007/s10463-013-0444-7

Hsu, E.P. (2002). *Stochastic Analysis on Manifolds*. *Graduate Studies in Mathematics* **38**. Providence, RI: Amer. Math. Soc. MR1882015 https://doi.org/10.1090/gsm/038

Huckemann, S.F. (2011a). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349 https://doi.org/10.1214/10-AOS862

Huckemann, S. (2011b). Inference on 3D Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Stat.* **38** 424–446. MR2833839 https://doi.org/10.1111/j.1467-9469.2010.00724.x

Huckemann, S.F. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259. MR2981621 https://doi.org/10.1007/s10463-012-0352-2

Huckemann, S.F. and Eltzner, B. (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *Ann. Statist.* **46** 1994–2019. MR3845008 https://doi.org/10.1214/17-AOS1609

Hundrieser, S., Eltzner, B. and Huckemann, S.F. (2020). Finite sample smeariness of Fréchet means and application to climate. arXiv:2005.02321 [stat].

Jensen, M.H. and Sommer, S. (2022). Mean estimation on the diagonal of product manifolds. *Algorithms* **15** 92. https://doi.org/10.3390/a15030092

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541. MR0442975 https://doi.org/10.1002/cpa.3160300502

Kendall, W.S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. Lond. Math. Soc. (3)* **61** 371–406. MR1063050 https://doi.org/10.1112/plms/s3-61.2.371

Le, H. and Barden, D. (2014). On the measure of the cut locus of a Fréchet mean. *Bull. Lond. Math. Soc.* **46** 698–708. MR3239607 https://doi.org/10.1112/blms/bdu025

Le, H. and Kume, A. (2000). The Fréchet mean shape and the shape of the means. *Adv. in Appl. Probab.* **32** 101–113. MR1765168 https://doi.org/10.1239/aap/1013540025

Li, P. and Yau, S.-T. (1986). On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156** 153–201. MR0834612 https://doi.org/10.1007/BF02399203

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. Revised reprint of ıt Statistics of directional data by Mardia [MR0336854 (49 #1627)]. MR1828667

McCormack, A. and Hoff, P. (2022). The Stein effect for Fréchet means. *Ann. Statist.* **50** 3647–3676. MR4524511 https://doi.org/10.1214/22-aos2245

McElhinny, M.W. and Lock, J. (1996). IAGA paleomagnetic databases with access. *Surv. Geophys.* **17** 575–591.

Molchanov, I. (2005). *Theory of Random Sets*. *Probability and Its Applications (New York)*. London: Springer London, Ltd. MR2132405

Nye, T.M.W., Tang, X., Weyenberg, G. and Yoshida, R. (2017). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika* **104** 901–922. MR3737311 https://doi.org/10.1093/biomet/asx047

Papaspiliopoulos, O. and Roberts, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations*. *Monogr. Statist. Appl. Probab.* **124** 311–340. Boca Raton, FL: CRC Press. MR2976985 https://doi.org/10.1201/b12126-5

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vision* **25** 127–154. MR2254442 https://doi.org/10.1007/s10851-006-6228-4

Pennec, X. (2015). Barycentric subspaces and affine spans in manifolds. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **9389** 12–21. Cham: Springer. MR3442180 https://doi.org/10.1007/978-3-319-25040-3_2

Pennec, X. (2018). Barycentric subspace analysis on manifolds. *Ann. Statist.* **46** 2711–2746. MR3851753 https://doi.org/10.1214/17-AOS1636

Pennec, X. (2019). Curvature effects on the empirical mean in Riemannian and affine manifolds: A non-asymptotic high concentration expansion in the small-sample regime. arXiv:1906.07418 [math, stat].

Pennec, X. and Arsigny, V. (2013). Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In *Matrix Information Geometry* 123–166. Heidelberg: Springer. MR2964451 https://doi.org/10.1007/978-3-642-30232-9_7

Renlund, H. (2011). Limit theorems for stochastic approximation algorithms. arXiv:1102.4741.

Said, S. and Manton, J.H. (2012). Extrinsic mean of Brownian distributions on compact Lie groups. *IEEE Trans. Inf. Theory* **58** 3521–3535. MR2924382 https://doi.org/10.1109/TIT.2012.2185680

Schötz, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics* **56** 34–52. MR4392146 https://doi.org/10.1080/02331888.2022.2032063

Small, C.G. (1996). *The Statistical Theory of Shape*. *Springer Series in Statistics*. New York: Springer. MR1418639 https://doi.org/10.1007/978-1-4612-4032-7

Sommer, S. (2015). Anisotropic distributions on manifolds: Template estimation and most probable paths. In *Inf. Process. Med. Imag.* **24** 193–204.

Sommer, S. and Svane, A.M. (2017). Modelling anisotropic covariance using stochastic development and sub-Riemannian frame bundle geometry. *J. Geom. Mech.* **9** 391–410. MR3666688 https://doi.org/10.3934/jgm.2017015

Sommer, S., Arnaudon, A., Kühnel, L. and Joshi, S. (2017). Bridge simulation and metric estimation on landmark manifolds. In *MMBIA at MICCAI 2017*. *LNCS* 79–91. Springer. https://doi.org/10.1007/978-3-319-67675-3\protect \T1\textunderscore 8

Tran, D., Eltzner, B. and Huckemann, S. (2021). Smeariness Begets finite sample smeariness. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **12829** 29–36. Cham: Springer. MR4424306 https://doi.org/10.1007/978-3-030-80209-7_4

Yang, C.-H., Doss, H. and Vemuri, B.C. (2020). An empirical Bayes approach to shrinkage estimation on the manifold of symmetric positive-definite matrices. arXiv:2007.02153.

Zhang, M. and Fletcher, T. (2013). Probabilistic principal geodesic analysis. In *Adv. Neural Inf. Process. Sys.* **26** 1178–1186. Curran Associates.

Zhao, C. and Song, J.S. (2018). Exact heat kernel on a hypersphere and its applications in kernel SVM. *Front. Appl. Math. Stat.* **4**. https://doi.org/10.3389/fams.2018.00001

Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A* 591–602. Dordrecht: Reidel. MR0501230

Ziezold, H. (1994). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biom. J.* **36** 491–510. MR1292314 https://doi.org/10.1002/bimj.4710360409

Bernoulli **29**(4), 2023, 3171–3197
https://doi.org/10.3150/22-BEJ1579

Spiked eigenvalues of noncentral Fisher matrix with applications

ZHIQIANG HOU[1,a], XIAOZHUO ZHANG[2,b], ZHIDONG BAI[2,c] and JIANG HU[2,d]

[1] *School of Statistics, Shandong University of Finance and Economics,* [a] *houzq@sdufe.edu.cn*
[2] *KLASMOE and School of Mathematics and Statistics, Northeast Normal University,* [b] *zhangxz722@nenu.edu.cn,* [c] *baizd@nenu.edu.cn,* [d] *huj156@nenu.edu.cn*

In this paper, we investigate the asymptotic behavior of spiked eigenvalues of the noncentral Fisher matrix defined by $\mathbf{F}_p = \mathbf{C}_n(\mathbf{S}_N)^{-1}$, where $\mathbf{C}_n$ is a noncentral sample covariance matrix defined by $(\mathbf{\Xi}_n + \mathbf{X}_n)(\mathbf{\Xi}_n + \mathbf{X}_n)^*/n$ and $\mathbf{S}_N = \mathbf{Y}_N\mathbf{Y}_N^*/N$. The matrices $\mathbf{X}_n$ and $\mathbf{Y}_N$ are two independent random matrices with standard Gaussian entries, and $\mathbf{\Xi}_n$ is a nonrandom matrix. When the dimensions of $\mathbf{X}_n$ and $\mathbf{Y}_N$ grow to infinity proportionally, we establish a phase transition of the spiked eigenvalues of $\mathbf{F}_p$. Furthermore, we derive the central limit theorem (CLT) for these spiked eigenvalues. As a byproduct of the proof of the above results, we provide the fluctuations of the spiked eigenvalues of $\mathbf{C}_n$, which are also of interest. Furthermore, we develop the limits and CLT for the sample canonical correlation coefficients using the results of the spiked noncentral Fisher matrix and present consistent estimators of the population spiked eigenvalues and the population canonical correlation coefficients.

Keywords: Noncentral Fisher matrix; spiked eigenvalues; central limit theorem; canonical correlation analysis

References

[1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR1990662
[2] Bai, Z., Hou, Z., Hu, J., Jiang, D. and Zhang, X. (2021). Limiting canonical distribution of two large dimensional random vectors. In *Advances on Methodology and Applications of Statistics – a Volume in Honor of C.R. Rao on the Occasion of His 100th Birthday* (C.A. Coelho, N. Balakrishnan and B.C. Arnold, eds.).
[3] Bai, Z. and Ding, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011, 21. MR2934717 https://doi.org/10.1142/S2010326311500110
[4] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8
[5] Bai, Z. and Silverstein, J.W. (2012). No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. *Random Matrices Theory Appl.* **1** 1150004, 44. MR2930382 https://doi.org/10.1142/S2010326311500043
[6] 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
[7] 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
[8] 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
[9] Banna, M., Najim, J. and Yao, J. (2020). A CLT for linear spectral statistics of large random information-plus-noise matrices. *Stochastic Process. Appl.* **130** 2250–2281. MR4074699 https://doi.org/10.1016/j.spa.2019.06.017
[10] Bao, Z., Ding, X. and Wang, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

1350-7265

[11] Bao, Z., Hu, J., Pan, G. and Zhou, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 https://doi.org/10.1214/18-AOS1704

[12] Bodnar, T., Dette, H. and Parolya, N. (2019). Testing for independence of large dimensional vectors. *Ann. Statist.* **47** 2977–3008. MR3988779 https://doi.org/10.1214/18-AOS1771

[13] 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

[14] Capitaine, M. (2014). Exact separation phenomenon for the eigenvalues of large information-plus-noise type matrices, and an application to spiked models. *Indiana Univ. Math. J.* **63** 1875–1910. MR3298725 https://doi.org/10.1512/iumj.2014.63.5432

[15] Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129

[16] Ding, X. and Yang, F. (2021). Spiked separable covariance matrices and principal components. *Ann. Statist.* **49** 1113–1138. MR4255121 https://doi.org/10.1214/20-aos1995

[17] Dozier, R.B. and Silverstein, J.W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123 https://doi.org/10.1016/j.jmva.2006.09.006

[18] Dozier, R.B. and Silverstein, J.W. (2007). Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *J. Multivariate Anal.* **98** 1099–1122. MR2326242 https://doi.org/10.1016/j.jmva.2006.12.005

[19] Han, X., Pan, G. and Zhang, B. (2016). The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933 https://doi.org/10.1214/15-AOS1427

[20] Horn, R.A. and Johnson, C.R. (2013). *Matrix Analysis*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2978290

[21] Hou, Z., Zhang, X., Bai, Z. and Hu, J. (2023). Supplement to "Spiked eigenvalues of noncentral Fisher matrix with applications." https://doi.org/10.3150/22-BEJ1579SUPP

[22] Hu, J. and Bai, Z. (2014). Strong representation of weak convergence. *Sci. China Math.* **57** 2399–2406. MR3266500 https://doi.org/10.1007/s11425-014-4855-6

[23] James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. MR0181057 https://doi.org/10.1214/aoms/1177703550

[24] 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

[25] Jiang, D. Hou, Z. Bai, Z. and Li, R. (2022). Invariance principle and CLT for the spiked eigenvalues of large-dimensional Fisher matrices and applications. arXiv:2203.14248.

[26] Jiang, D., Hou, Z. and Hu, J. (2021). The limits of the sample spiked eigenvalues for a high-dimensional generalized Fisher matrix and its applications. *J. Statist. Plann. Inference* **215** 208–217. MR4244718 https://doi.org/10.1016/j.jspi.2021.03.004

[27] 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

[28] Johnstone, I.M. and Nadler, B. (2017). Roy's largest root test under rank-one alternatives. *Biometrika* **104** 181–193. MR3626473 https://doi.org/10.1093/biomet/asw060

[29] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. *Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. London-New York-Toronto, Ont.: Academic Press [Harcourt Brace Jovanovich, Publishers]. MR0560319

[30] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0652932

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

[32] Skorohod, A.V. (1956). Limit theorems for stochastic processes. *Teor. Veroyatn. Primen.* **1** 289–319. MR0084897

[33] Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[34] Wachter, K.W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* **8** 937–957. MR0585695

[35] Wang, Q. and Yao, J. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.* **45** 415–460. MR3611497 https://doi.org/10.1214/16-AOS1463

[36] Ma, Z. and Yang, F. (2021). Sample canonical correlation coefficients of high-dimensional random vectors with finite rank correlations. arXiv:2102.03297.

[37] Yang, F. (2022). Sample canonical correlation coefficients of high-dimensional random vectors: Local law and Tracy-Widom limit. *Random Matrices Theory Appl.* **11** Paper No. 2250007, 70. MR4391830 https://doi.org/10.1142/S2010326322500071

[38] Yang, F. (2022). Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors. *Electron. J. Probab.* **27** Paper No. 86, 71. MR4455877 https://doi.org/10.1214/22-ejp814

[39] Zheng, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 https://doi.org/10.1214/11-AIHP414

[40] Zheng, S., Bai, Z. and Yao, J. (2015). CLT for linear spectral statistics of a rescaled sample precision matrix. *Random Matrices Theory Appl.* **4** 1550014, 43. MR3418843 https://doi.org/10.1142/S2010326315500148

[41] Zheng, S., Bai, Z. and Yao, J. (2017). CLT for eigenvalue statistics of large-dimensional general Fisher matrices with applications. *Bernoulli* **23** 1130–1178. MR3606762 https://doi.org/10.3150/15-BEJ772

Bernoulli **29**(4), 2023, 3198–3220
https://doi.org/10.3150/22-BEJ1580

Sampling without replacement from a high-dimensional finite population

JIANG HU[1,a], SHAOCHEN WANG[2,b], YANGCHUN ZHANG[3,c] and WANG ZHOU[4,d]

[1]*KLASMOE and School of Mathematics & Statistics, Northeast Normal University, Changchun, 130024, P.R. China,* [a]*huj156@nenu.edu.cn*
[2]*School of Mathematics, South China University of Technology, Guangzhou, 510640, P.R. China,* [b]*mascwang@scut.edu.cn*
[3]*Department of Mathematics, Shanghai University, Shanghai, 200444, P.R. China,* [c]*zycstatis@shu.edu.cn*
[4]*Department of Statistics and Data Science, National University of Singapore, 117546, Singapore,* [d]*stazw@nus.edu.sg*

It is well known that most of the existing theoretical results in statistics are based on the assumption that the sample is generated with replacement from an infinite population. However, in practice, available samples are almost always collected without replacement. If the population is a finite set of real numbers, whether we can still safely use the results from samples drawn without replacement becomes an important problem. In this paper, we focus on the eigenvalues of high-dimensional sample covariance matrices generated without replacement from finite populations. Specifically, we derive the Tracy-Widom laws for their largest eigenvalues and apply these results to parallel analysis. We provide new insight into the permutation methods proposed by Buja and Eyuboglu in [Multivar Behav Res. 27(4) (1992) 509–540]. Simulation and real data studies are conducted to demonstrate our results.

Keywords: Largest eigenvalue; Tracy-Widom law; sample covariance matrix; finite population model; parallel analysis

References

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley.
Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8
Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613
Bao, Z. (2017). Tracy-Widom limit for Spearman's rho. *preprint*.
Bao, Z. (2019). Tracy-Widom limit for Kendall's tau. *Ann. Statist.* **47** 3504–3532. MR4025750 https://doi.org/10.1214/18-AOS1786
Bao, Z., Pan, G. and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 88. MR2988403 https://doi.org/10.1214/EJP.v17-1962
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
Bao, Z., Lin, L.-C., Pan, G. and Zhou, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Ann. Statist.* **43** 2588–2623. MR3405605 https://doi.org/10.1214/15-AOS1353
Benaych-Georges, F. and Knowles, A. (2017). Local semicircle law for Wigner matrices. In *Advanced Topics in Random Matrices*. *Panor. Synthèses* **53** 1–90. Paris: Soc. Math. France. MR3792624
Bloemendal, A., Knowles, A., Yau, H.-T. and Yin, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x
Buja, A. and Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivar. Behav. Res.* **27** 509–540.

Dobriban, E. (2020). Permutation methods for factor analysis and PCA. *Ann. Statist.* **48** 2824–2847. MR4152122 https://doi.org/10.1214/19-AOS1907

Dobriban, E. and Owen, A.B. (2019). Deterministic parallel analysis: An improved method for selecting factors and principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 163–183. MR3904784

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

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

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

Gao, J., Han, X., Pan, G. and Yang, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

Hu, J., Li, W., Liu, Z. and Zhou, W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Ann. Statist.* **47** 527–555. MR3909941 https://doi.org/10.1214/18-AOS1699

Hu, J., Wang, S.C., Zhang, Y.C. and Zhou, W. (2023). Supplement to "Sampling without replacement from a high-dimensional finite population." https://doi.org/10.3150/22-BEJ1580SUPP

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

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

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

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

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319** 1100–1104. https://doi.org/10.1126/science.1153717

Marcenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457.

Onatski, A. (2008). The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. MR2398763 https://doi.org/10.1214/07-AAP454

Pillai, N.S. and Yin, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042 https://doi.org/10.1214/12-AOS1022

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

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

Tian, B., Zhang, Y. and Zhou, W. (2021). Tracy-Widom law for the largest eigenvalue of sample covariance matrix generated by VARMA. *Random Matrices Theory Appl.* **10** 2150022. MR4260217 https://doi.org/10.1142/S2010326321500222

Wen, J., Xie, J.H., Yu, L. and Zhou, W. (2022). Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions. *Bernoulli* **28** 2941–2967. MR4474568 https://doi.org/10.3150/21-BEJ1443

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

Bernoulli **29**(4), 2023, 3221–3254
https://doi.org/10.3150/22-BEJ1581

Hypothesis testing for equality of latent positions in random graphs

XINJIE DU[a] and MINH TANG[b]

Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA, [a]*xdu22@ncsu.edu,* [b]*mtang8@ncsu.edu*

We consider the hypothesis testing problem that two vertices i and j of a generalized random dot product graph have the same latent positions, possibly up to scaling. Special cases of this hypothesis testing problem include testing whether two vertices in a stochastic block model or degree-corrected stochastic block model graph have the same block membership vectors, or testing whether two vertices in a popularity adjusted block model have the same community assignment. We propose several test statistics based on the empirical Mahalanobis distances between the ith and jth rows of either the adjacency or the normalized Laplacian spectral embedding of the graph. We show that, under mild conditions, these test statistics have limiting chi-square distributions under both the null and local alternative hypothesis, and we derive explicit expressions for the non-centrality parameters under the local alternative. Using these limiting results, we address the model selection problems including choosing between the standard stochastic block model and its degree-corrected variant, and choosing between the Erdős–Rényi model and stochastic block model. The effectiveness of our proposed tests is illustrated via both simulation studies and real data applications.

Keywords: Asymptotic normality; generalized random dot product graphs; model selection; spectral embedding; stochastic block models

References

Abbe, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86 pp. MR3827065

Adamic, L.A. and Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43.

Agterberg, J., Park, Y., Larson, J., White, C., Priebe, C.E. and Lyzinski, V. (2020). Vertex nomination, consistent estimation, and adversarial modification. *Electron. J. Stat.* **14** 3230–3267. MR4146361 https://doi.org/10.1214/20-EJS1744

Airoldi, E.M., Blei, D.M., Fienberg, S.E. and Xing, E.P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

Athreya, A., Priebe, C.E., Tang, M., Lyzinski, V., Marchette, D.J. and Sussman, D.L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. MR3494576 https://doi.org/10.1007/s13171-015-0071-x

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

Bickel, P.J. and Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

Cai, T.T. and Li, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. MR3346696 https://doi.org/10.1214/14-AOS1290

Chaudhuri, K., Chung, F. and Tsiatas, A. (2012). Spectral partitioning of graphs with general degrees and the extended planted partition model. In *Proceedings of the 25th Conference on Learning Theory*.

Chen, K. and Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 https://doi.org/10.1080/01621459.2016.1246365

1350-7265

Coifman, R.R. and Lafon, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30. MR2238665 https://doi.org/10.1016/j.acha.2006.04.006

Du, X. and Tang, M. (2023). Supplement to "Hypothesis testing for equality of latent positions in random graphs." https://doi.org/10.3150/22-BEJ1581SUPP

Fan, J., Fan, Y., Han, X. and Lv, J. (2022a). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505

Fan, J., Fan, Y., Han, X. and Lv, J. (2022b). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990

Fishkind, D.E., Lyzinski, V., Pao, H., Chen, L. and Priebe, C.E. (2015). Vertex nomination schemes for membership prediction. *Ann. Appl. Stat.* **9** 1510–1532. MR3418733 https://doi.org/10.1214/15-AOAS834

Gao, C., Ma, Z., Zhang, A.Y. and Zhou, H.H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45 pp. MR3687603

Ghoshdastidar, D., Gutzeit, M., Carpentier, A. and von Luxburg, U. (2020). Two-sample hypothesis testing for inhomogeneous random graphs. *Ann. Statist.* **48** 2208–2229. MR4134792 https://doi.org/10.1214/19-AOS1884

Gilpin, S., Eliassi-Rad, T. and Davidson, I. (2013). Guided learning for role discovery (GLRD): Framework, algorithms, and applications. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Ginestet, C.E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E.D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

Hu, J., Qin, H., Yan, T. and Zhao, Y. (2020). Corrected Bayesian information criterion for stochastic block models. *J. Amer. Statist. Assoc.* **115** 1771–1783. MR4189756 https://doi.org/10.1080/01621459.2019.1637744

Jin, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

Karrer, B. and Newman, M.E.J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)* **83** 016107, 10 pp. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

Koo, J., Tang, M. and Trosset, M.W. (2023). Popularity adjusted block models are generalized random dot product graphs. *J. Comput. Graph. Statist.* https://doi.org/10.1080/10618600.2022.2081576

Le, C.M. and Levina, E. (2022). Estimating the number of communities by spectral methods. *Electron. J. Stat.* **16** 3315–3342. MR4422967 https://doi.org/10.1214/21-ejs1971

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

Lei, J. (2021). Network representation using graph root distributions. *Ann. Statist.* **49** 745–768. MR4255106 https://doi.org/10.1214/20-aos1976

Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

Li, T., Levina, E. and Zhu, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. MR4108931 https://doi.org/10.1093/biomet/asaa006

Ma, S., Su, L. and Zhang, Y. (2021). Determining the number of communities in degree-corrected stochastic block models. *J. Mach. Learn. Res.* **22** Paper No. 69, 63 pp. MR4253762

Ng, A., Jordan, M. and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Adv. Neural Inf. Process. Syst.* **14**.

Nickel, C.L.M. (2008). Random dot product graphs a model for social networks. Ph.D. thesis, Johns Hopkins Univ. MR2710154

Noroozi, M., Rimal, R. and Pensky, M. (2021). Estimation and clustering in popularity adjusted block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 293–317. MR4250277 https://doi.org/10.1111/rssb.12410

Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

Rubin-Delanchy, P., Cape, J., Tang, M. and Priebe, C.E. (2022). A statistical interpretation of spectral embedding: The generalised random dot product graph. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1446–1473. MR4494166 https://doi.org/10.1111/rssb.12509

Sengupta, S. and Chen, Y. (2018). A block model for node popularity in networks with community structure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 365–386. MR3763696 https://doi.org/10.1111/rssb.12245

Serre, D. (2002). *Matrices: Theory and Applications*. *Graduate Texts in Mathematics* **216**. New York: Springer. MR1923507

Stewart, G.W. (1977). On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM Rev.* **19** 634–662. MR0461871 https://doi.org/10.1137/1019104

Sussman, D.L., Tang, M., Fishkind, D.E. and Priebe, C.E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899 https://doi.org/10.1080/01621459.2012.699795

Tang, M. and Priebe, C.E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.* **46** 2360–2415. MR3845021 https://doi.org/10.1214/17-AOS1623

Tang, M., Athreya, A., Sussman, D.L., Lyzinski, V., Park, Y. and Priebe, C.E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *J. Comput. Graph. Statist.* **26** 344–354. MR3640191 https://doi.org/10.1080/10618600.2016.1193505

von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807 https://doi.org/10.1214/009053607000000640

Wang, Y.X.R. and Bickel, P.J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528. MR3650391 https://doi.org/10.1214/16-AOS1457

Wang, B., Pourshafeie, A., Zitnik, M., Zhu, J., Bustamante, C.D., Batzoglou, S. and Leskovec, J. (2018). Network enhancement as a general method to denoise weighted biological networks. *Nat. Commun.* **9** 1–8.

Xie, F. and Xu, Y. (2023). Efficient estimation for random dot product graphs via a one-step procedure. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2021.1948419

Yan, X., Shalizi, C., Jensen, J.E., Krzakala, F., Moore, C., Zdeborová, L., Zhang, P. and Zhu, Y. (2014). Model selection for degree-corrected block models. *J. Stat. Mech. Theory Exp.* **2014** P05007.

Bernoulli **29**(4), 2023, 3255–3279
https://doi.org/10.3150/22-BEJ1582

Tail processes for stable-regenerative multiple-stable model

SHUYANG BAI[1,a] and YIZAO WANG[2,b]

[1]*Department of Statistics, University of Georgia, 310 Herty Drive, Athens, GA, 30602, USA,* [a]*bsy9142@uga.edu*
[2]*Department of Mathematical Sciences, University of Cincinnati, 2815 Commons Way, Cincinnati, OH, 45221-0025, USA,* [b]*yizao.wang@uc.edu*

We investigate a family of discrete-time stationary processes defined by multiple stable integrals and renewal processes with infinite means. The model may exhibit behaviors of short-range or long-range dependence, respectively, depending on the parameters. The main contribution is to establish a phase transition in terms of the tail processes that characterize local clustering of extremes. Moreover, in the short-range dependence regime, the model provides an example where the extremal index is different from the candidate extremal index.

Keywords: Extremal index; long-range dependence; multiple integral; phase transition; regular variation; renewal process; short-range dependence; tail process

References

Bai, S. (2022). Limit theorems for conservative flows on multiple stochastic integrals. *J. Theoret. Probab.* **35** 917–948. MR4414409 https://doi.org/10.1007/s10959-021-01090-9

Bai, S., Owada, T. and Wang, Y. (2020). A functional non-central limit theorem for multiple-stable processes with long-range dependence. *Stochastic Process. Appl.* **130** 5768–5801. MR4127346 https://doi.org/10.1016/j.spa.2020.04.007

Bai, S. and Taqqu, M.S. (2020). Limit theorems for long-memory flows on Wiener chaos. *Bernoulli* **26** 1473–1503. MR4058375 https://doi.org/10.3150/19-BEJ1168

Bai, S. and Wang, Y. (2021). Phase transition of extremes for a family of multiple-stable processes.

Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Heidelberg: Springer. MR3075595 https://doi.org/10.1007/978-3-642-35512-7

Bertoin, J. (1999). Subordinators: Examples and applications. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*. *Lecture Notes in Math.* **1717** 1–91. Berlin: Springer. MR1746300 https://doi.org/10.1007/978-3-540-48115-7_1

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

Davis, R.A. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23** 879–917. MR1334176

Davis, R.A. and Mikosch, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26** 2049–2080. MR1673289 https://doi.org/10.1214/aos/1024691368

de la Peña, V.H., Montgomery-Smith, S.J. and Szulga, J. (1994). Contraction and decoupling inequalities for multilinear forms and U-statistics. *Ann. Probab.* **22** 1745–1765. MR1331202

Doney, R.A. (1997). One-sided local large deviation and renewal theorems in the case of infinite mean. *Probab. Theory Related Fields* **107** 451–465. MR1440141 https://doi.org/10.1007/s004400050093

Giacomin, G. (2007). *Random Polymer Models*. London: Imperial College Press. MR2380992 https://doi.org/10.1142/9781860948299

Itô, K. (1951). Multiple Wiener integral. *J. Math. Soc. Japan* **3** 157–169. MR0044064 https://doi.org/10.2969/jmsj/00310157

Krakowiak, W. and Szulga, J. (1986). Random multilinear forms. *Ann. Probab.* **14** 955–973. MR0841596

Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

Kwapień, S. and Woyczyński, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*. *Probability and Its Applications*. Boston, MA: Birkhäuser, Inc. MR1167198 https://doi.org/10.1007/978-1-4612-0425-1

O'Brien, G.L., Torfs, P.J.J.F. and Vervaat, W. (1990). Stationary self-similar extremal processes. *Probab. Theory Related Fields* **87** 97–119. MR1076958 https://doi.org/10.1007/BF01217748

Owada, T. and Samorodnitsky, G. (2015). Functional central limit theorem for heavy tailed stationary infinitely divisible processes generated by conservative flows. *Ann. Probab.* **43** 240–285. MR3298473 https://doi.org/10.1214/13-AOP899

Pipiras, V. and Taqqu, M.S. (2017). *Long-Range Dependence and Self-Similarity*. *Cambridge Series in Statistical and Probabilistic Mathematics* **45**. Cambridge: Cambridge Univ. Press. MR3729426

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. New York: Springer. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424

Rosiński, J. and Samorodnitsky, G. (1996). Classes of mixing stable processes. *Bernoulli* **2** 365–377. MR1440274 https://doi.org/10.2307/3318419

Rosiński, J. and Samorodnitsky, G. (1999). Product formula, tails and independence of multiple stable integrals. In *Advances in Stochastic Inequalities (Atlanta, GA, 1997)*. *Contemp. Math.* **234** 169–194. Providence, RI: Amer. Math. Soc. MR1694771 https://doi.org/10.1090/conm/234/03453

Samorodnitsky, G. (2016). *Stochastic Processes and Long Range Dependence*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3561100 https://doi.org/10.1007/978-3-319-45575-4

Samorodnitsky, G. and Szulga, J. (1989). An asymptotic evaluation of the tail of a multiple symmetric α-stable integral. *Ann. Probab.* **17** 1503–1520. MR1048942

Samorodnitsky, G. and Wang, Y. (2019). Extremal theory for long range dependent infinitely divisible processes. *Ann. Probab.* **47** 2529–2562. MR3980927 https://doi.org/10.1214/18-AOP1318

Smith, R.L. (1988). A counterexample concerning the extremal index. *Adv. in Appl. Probab.* **20** 681–683. MR0955511 https://doi.org/10.2307/1427042

Szulga, J. (1991). Multiple stochastic integrals with respect to symmetric infinitely divisible random measures. *Ann. Probab.* **19** 1145–1156. MR1112410

Wang, Y. (2022). Choquet random sup-measures with aggregations. *Extremes* **25** 25–54. MR4376583 https://doi.org/10.1007/s10687-021-00425-3

Bernoulli **29**(4), 2023, 3280–3306
https://doi.org/10.3150/23-BEJ1583

Central limit theorems and asymptotic independence for local U-statistics on diverging halfspaces

ANDREW M. THOMAS[1,2,a]

[1]*Center for Applied Mathematics, Cornell University, Ithaca, NY, USA,* [a]*amt269@cornell.edu*
[2]*Department of Statistics and Data Science, Cornell University, Ithaca, NY, USA*

We consider the stochastic behavior of a class of local U-statistics of Poisson processes—which include subgraph and simplex counts as special cases, and amounts to quantifying clustering behavior—for point clouds lying in diverging halfspaces. We provide limit theorems for distributions with light and heavy tails. In particular, we prove finite-dimensional central limit theorems. In the light tail case we investigate tails that decay at least as slow as exponential and at least as fast as Gaussian. These results also furnish as a corollary that U-statistics for halfspaces diverging at different angles are asymptotically independent, and that there is no asymptotic independence for heavy-tailed densities. Using state-of-the-art bounds derived from recent breakthroughs combining Stein's method and Malliavin calculus, we quantify the rate of this convergence in terms of Kolmogorov distance. We also investigate the behavior of local U-statistics of a Poisson process conditioned to lie in a diverging halfspace and find that the upper bound on the Kolmogorov distance to a standard normal distribution is smaller the lighter the tail of the density is.

Keywords: Asymptotic independence; central limit theorems; conditional extremes; Stein's method; U-statistics

References

[1] Adler, R.J., Bobrowski, O. and Weinberger, S. (2014). Crackle: The homology of noise. *Discrete Comput. Geom.* **52** 680–704. MR3279544 https://doi.org/10.1007/s00454-014-9621-6
[2] Bachmann, S. and Reitzner, M. (2018). Concentration for Poisson U-statistics: Subgraph counts in random geometric graphs. *Stochastic Process. Appl.* **128** 3327–3352. MR3849811 https://doi.org/10.1016/j.spa.2017.11.001
[3] Balkema, G. and Embrechts, P. (2004). Multivariate excess distributions. ETHZ preprint, available at https://people.math.ethz.ch/~embrecht/ftp/guuspe08Jun04.pdf.
[4] Balkema, G. and Embrechts, P. (2007). *High Risk Scenarios and Extremes: A Geometric Approach*. *Zurich Lectures in Advanced Mathematics*. Zürich: European Mathematical Society (EMS). MR2372552 https://doi.org/10.4171/035
[5] Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Adv. in Appl. Probab.* **42** 411–432. MR2675110 https://doi.org/10.1239/aap/1275055236
[6] Durrett, R. (2010). *Probability: Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics* **31**. Cambridge: Cambridge Univ. Press. MR2722836 https://doi.org/10.1017/CBO9780511779398
[7] Eichelsbacher, P. and Thäle, C. (2014). New Berry-Esseen bounds for non-linear functionals of Poisson random measures. *Electron. J. Probab.* **19** no. 102, 25. MR3275854 https://doi.org/10.1214/EJP.v19-3061
[8] Hashorva, E. (2005). Asymptotics and bounds for multivariate Gaussian tails. *J. Theoret. Probab.* **18** 79–97. MR2132272 https://doi.org/10.1007/s10959-004-2577-3
[9] Hashorva, E. (2019). Approximation of some multivariate risk measures for Gaussian risks. *J. Multivariate Anal.* **169** 330–340. MR3875603 https://doi.org/10.1016/j.jmva.2018.10.006

1350-7265

[10] Hug, D. and Weil, W. (2020). *Lectures on Convex Geometry*. *Graduate Texts in Mathematics* **286**. Cham: Springer. MR4180684 https://doi.org/10.1007/978-3-030-50180-8

[11] Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. in Appl. Probab.* **34** 587–608. MR1929599 https://doi.org/10.1239/aap/1033662167

[12] Last, G. and Penrose, M. (2018). *Lectures on the Poisson Process*. *Institute of Mathematical Statistics Textbooks* **7**. Cambridge: Cambridge Univ. Press. MR3791470

[13] Owada, T. (2017). Functional central limit theorem for subgraph counting processes. *Electron. J. Probab.* **22** Paper No. 17, 38. MR3622887 https://doi.org/10.1214/17-EJP30

[14] Owada, T. and Adler, R.J. (2017). Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** 2004–2055. MR3650420 https://doi.org/10.1214/16-AOP1106

[15] Owada, T. and Wei, Z. (2022). Functional strong law of large numbers for Betti numbers in the tail. *Extremes* **25** 655–693. MR4505183 https://doi.org/10.1007/s10687-022-00441-x

[16] Peccati, G., Solé, J.L., Taqqu, M.S. and Utzet, F. (2010). Stein's method and normal approximation of Poisson functionals. *Ann. Probab.* **38** 443–478. MR2642882 https://doi.org/10.1214/09-AOP477

[17] 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

[18] Reitzner, M. and Schulte, M. (2013). Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.* **41** 3879–3909. MR3161465 https://doi.org/10.1214/12-AOP817

[19] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424

[20] Schneider, R. (2014). *Convex Bodies: The Brunn-Minkowski Theory*, expanded ed. *Encyclopedia of Mathematics and Its Applications* **151**. Cambridge: Cambridge Univ. Press. MR3155183

[21] Schulte, M. (2016). Normal approximation of Poisson functionals in Kolmogorov distance. *J. Theoret. Probab.* **29** 96–117. MR3463079 https://doi.org/10.1007/s10959-014-0576-6

[22] Schulte, M. and Thäle, C. (2012). The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process. Appl.* **122** 4096–4120. MR2971726 https://doi.org/10.1016/j.spa.2012.08.011

[23] Thomas, A.M. and Owada, T. (2021). Functional strong laws of large numbers for Euler characteristic processes of extreme sample clouds. *Extremes* **24** 699–724. MR4331054 https://doi.org/10.1007/s10687-021-00419-1

[24] Walschap, G. (2015). *Multivariable Calculus and Differential Geometry*. *De Gruyter Graduate*. Berlin: de Gruyter. MR3381889 https://doi.org/10.1515/9783110369540

Bernoulli **29**(4), 2023, 3307–3333
https://doi.org/10.3150/23-BEJ1584

Asymptotics for densities of exponential functionals of subordinators

MARTIN MINCHEV[1,a] and MLADEN SAVOV[1,2,b,c]

[1]*Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", 5, James Bourchier blvd., 1164 Sofia, Bulgaria,* [a]*mjminchev@fmi.uni-sofia.bg,* [b]*msavov@fmi.uni-sofia.bg*
[2]*Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Akad. Georgi Bonchev str., Block 8, Sofia 1113, Bulgaria,* [c]*mladensavov@math.bas.bg*

In this paper we derive non-classical Tauberian asymptotics at infinity for the tail, the density and its derivatives of a large class of exponential functionals of subordinators. More precisely, we consider the case for which the Lévy measure of the subordinator satisfies the well-known and mild condition of *positive increase*. This is achieved via a convoluted application of the saddle point method to the Mellin transform of these exponential functionals which is given in terms of Bernstein–gamma functions. To apply the saddle point method, we improved the Stirling type asymptotics for Bernstein–gamma functions in the complex plane. As an application we derive the asymptotics of the density and its derivatives for all exponential functionals of a broad class of non-decreasing, potentially killed compound Poisson processes, which turns out to be precisely as that of an exponentially distributed random variable. We show further that a large class of densities are even analytic in a cone of the complex plane.

Keywords: Bernstein functions; exponential functionals of Lévy processes; non-classical Tauberian asymptotics; special functions

References

[1] Alili, L., Jedidi, W. and Rivero, V. (2014). On exponential functionals, harmonic potential measures and undershoots of subordinators. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** 711–735. MR3323879

[2] Balkema, A.A., Klüppelberg, C. and Stadtmüller, U. (1995). Tauberian results for densities with Gaussian tails. *J. Lond. Math. Soc. (2)* **51** 383–400. MR1325581 https://doi.org/10.1112/jlms/51.2.383

[3] 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

[4] Behme, A., Lindner, A. and Maller, R. (2011). Stationary solutions of the stochastic differential equation $dV_t = V_{t-} dU_t + dL_t$ with Lévy noise. *Stochastic Process. Appl.* **121** 91–108. MR2739007 https://doi.org/10.1016/j.spa.2010.09.003

[5] Behme, A., Lindner, A. and Reker, J. (2021). On the law of killed exponential functionals. *Electron. J. Probab.* **26** Paper No. 60, 35. MR4254802 https://doi.org/10.1214/21-ejp616

[6] Békéssy, A. (1957). Eine Verallgemeinerung der Laplaceschen Methode. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **2** 105–125. MR0097667

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

[8] Bertoin, J. and Caballero, M.-E. (2002). Entrance from 0+ for increasing semi-stable Markov processes. *Bernoulli* **8** 195–205. MR1895890

[9] Bertoin, J., Curien, N. and Kortchemski, I. (2018). Random planar maps and growth-fragmentations. *Ann. Probab.* **46** 207–260. MR3758730 https://doi.org/10.1214/17-AOP1183

[10] Bertoin, J. and Kortchemski, I. (2016). Self-similar scaling limits of Markov chains on the positive integers. *Ann. Appl. Probab.* **26** 2556–2595. MR3543905 https://doi.org/10.1214/15-AAP1157

1350-7265

[11] Bertoin, J. and Yor, M. (2002). The entrance laws of self-similar Markov processes and exponential functionals of Lévy processes. *Potential Anal.* **17** 389–400. MR1918243 https://doi.org/10.1023/A:1016377720516
[12] Bertoin, J. and Yor, M. (2005). Exponential functionals of Lévy processes. *Probab. Surv.* **2** 191–212. MR2178044 https://doi.org/10.1214/154957805100000122
[13] 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
[14] Budd, T. (2018). The peeling process on random planar maps coupled to an $O(n)$ loop model (with an appendix by linxiao Chen). arXiv:1809.02012.
[15] Caballero, M.E. and Chaumont, L. (2006). Weak convergence of positive self-similar Markov processes and overshoots of Lévy processes. *Ann. Probab.* **34** 1012–1034. MR2243877 https://doi.org/10.1214/009117905000000611
[16] Feigin, P.D. and Yashchin, E. (1983). On a strong Tauberian result. *Z. Wahrsch. Verw. Gebiete* **65** 35–48. MR0717931 https://doi.org/10.1007/BF00534992
[17] Haas, B. (2022). Tail asymptotics for extinction times of self-similar fragmentations. arXiv:2111.07757.
[18] Haas, B. (2022). Precise asymptotics for the density and the upper tail of exponential functionals of subordinators. arXiv:2106.08691.
[19] Haas, B. and Rivero, V. (2012). Quasi-stationary distributions and Yaglom limits of self-similar Markov processes. *Stochastic Process. Appl.* **122** 4054–4095. MR2971725 https://doi.org/10.1016/j.spa.2012.08.006
[20] Hackmann, D. and Kuznetsov, A. (2014). Asian options and meromorphic Lévy processes. *Finance Stoch.* **18** 825–844. MR3255753 https://doi.org/10.1007/s00780-014-0237-8
[21] Hirsch, F. and Yor, M. (2013). On the Mellin transforms of the perpetuity and the remainder variables associated to a subordinator. *Bernoulli* **19** 1350–1377. MR3102555 https://doi.org/10.3150/12-BEJSP01
[22] Klüppelberg, C., Lindner, A. and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: Stationarity and second-order behaviour. *J. Appl. Probab.* **41** 601–622. MR2074811 https://doi.org/10.1017/s0021900200020428
[23] Kuznetsov, A. and Pardo, J.C. (2013). Fluctuations of stable processes and exponential functionals of hypergeometric Lévy processes. *Acta Appl. Math.* **123** 113–139. MR3010227 https://doi.org/10.1007/s10440-012-9718-y
[24] 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
[25] Loeffen, R., Patie, P. and Savov, M. (2019). Extinction time of non-Markovian self-similar processes, persistence, annihilation of jumps and the Fréchet distribution. *J. Stat. Phys.* **175** 1022–1041. MR3959988 https://doi.org/10.1007/s10955-019-02279-3
[26] Maulik, K. and Zwart, B. (2006). Tail asymptotics for exponential functionals of Lévy processes. *Stochastic Process. Appl.* **116** 156–177. MR2197972 https://doi.org/10.1016/j.spa.2005.09.002
[27] Minchev, M. and Savov, M. (2023). Supplement to "Asymptotics for densities of exponential functionals of subordinators." https://doi.org/10.3150/23-BEJ1584SUPP
[28] 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
[29] 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
[30] Pardo, J.C., Rivero, V. and van Schaik, K. (2013). On the density of exponential functionals of Lévy processes. *Bernoulli* **19** 1938–1964. MR3129040 https://doi.org/10.3150/12-BEJ436
[31] Patie, P. (2009). Exponential functional of a new family of Lévy processes and self-similar continuous state branching processes with immigration. *Bull. Sci. Math.* **133** 355–382. MR2532690 https://doi.org/10.1016/j.bulsci.2008.10.001
[32] Patie, P. (2013). Asian options under one-sided Lévy models. *J. Appl. Probab.* **50** 359–373. MR3102485 https://doi.org/10.1239/jap/1371648946
[33] Patie, P. and Savov, M. (2012). Extended factorizations of exponential functionals of Lévy processes. *Electron. J. Probab.* **17** no. 38, 22. MR2928721 https://doi.org/10.1214/EJP.v17-2057

[34] Patie, P. and Savov, M. (2013). Exponential functional of Lévy processes: Generalized Weierstrass products and Wiener-Hopf factorization. *C. R. Math. Acad. Sci. Paris* **351** 393–396. MR3072167 https://doi.org/10.1016/j.crma.2013.04.023
[35] Patie, P. and Savov, M. (2017). Cauchy problem of the non-self-adjoint Gauss-Laguerre semigroups and uniform bounds for generalized Laguerre polynomials. *J. Spectr. Theory* **7** 797–846. MR3713026 https://doi.org/10.4171/JST/178
[36] Patie, P. and Savov, M. (2018). Bernstein-gamma functions and exponential functionals of Lévy processes. *Electron. J. Probab.* **23** Paper No. 75, 101. MR3835481 https://doi.org/10.1214/18-EJP202
[37] Patie, P. and Savov, M. (2021). Spectral expansions of non-self-adjoint generalized Laguerre semigroups. *Mem. Amer. Math. Soc.* **272** vii+182. MR4320772 https://doi.org/10.1090/memo/1336
[38] Patie, P., Savov, M. and Zhao, Y. (2019). Intertwining, excursion theory and Krein theory of strings for non-self-adjoint Markov semigroups. *Ann. Probab.* **47** 3231–3277. MR4021250 https://doi.org/10.1214/19-AOP1338
[39] Patie, P. and Vaidyanathan, A. (2022). Non-classical Tauberian and Abelian type criteria for the moment problem. *Math. Nachr.* **295** 970–990. MR4453945
[40] Pierre, P. (2009). Infinite divisibility of solutions to some self-similar integro-differential equations and exponential functionals of Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 667–684. MR2548498 https://doi.org/10.1214/08-AIHP182
[41] Schilling, R.L., Song, R. and Vondraček, Z. (2012). *Bernstein Functions: Theory and Applications*, 2nd ed. *De Gruyter Studies in Mathematics* **37**. Berlin: de Gruyter. MR2978140 https://doi.org/10.1515/9783110269338
[42] Urbanik, K. (1995). Infinite divisibility of some functionals on stochastic processes. *Probab. Math. Statist.* **15** 493–513. Dedicated to the memory of Jerzy Neyman. MR1369817
[43] Xu, W. (2021). Asymptotic results for heavy-tailed Lévy processes and their exponential functionals. *Bernoulli* **27** 2766–2803. MR4303903 https://doi.org/10.3150/21-BEJ1330
[44] Yor, M. (2001). *Exponential Functionals of Brownian Motion and Related Processes*. *Springer Finance*. Berlin: Springer. MR1854494 https://doi.org/10.1007/978-3-642-56634-9

Bernoulli **29**(4), 2023, 3334–3368
https://doi.org/10.3150/23-BEJ1585

Minimax boundary estimation and estimation with boundary

EDDIE AAMARI[1,a], CATHERINE AARON[2,c] and CLÉMENT LEVRARD[1,b]

[1]*Laboratoire de Probabilités, Statistique et Modélisation, CNRS, Université Paris Cité, Sorbonne Université, Paris, France,* [a]*aamari@lpsm.paris,* [b]*levrard@lpsm.paris*
[2]*Laboratoire de Mathématiques Blaise Pascal, CNRS, Université Clermont Auvergne, Clermont-Ferrand, France,* [c]*catherine.aaron@uca.fr*

We derive non-asymptotic minimax bounds for the Hausdorff estimation of d-dimensional submanifolds $M \subset \mathbb{R}^D$ with (possibly) non-empty boundary ∂M. The model reunites and extends the most prevalent C^2-type set estimation models: manifolds without boundary, and full-dimensional domains. We consider both the estimation of the manifold M itself and that of its boundary ∂M if non-empty. Given n samples, the minimax rates are of order $O\big((\log n/n)^{2/d}\big)$ if $\partial M = \emptyset$ and $O\big((\log n/n)^{2/(d+1)}\big)$ if $\partial M \neq \emptyset$, up to logarithmic factors. In the process, we develop a Voronoi-based procedure that allows to identify enough points $O\big((\log n/n)^{2/(d+1)}\big)$-close to ∂M for reconstructing it. Explicit constant derivations are given, showing that these rates do not depend on the ambient dimension $D \gg d$.

Keywords: Boundary; geometric inference; manifold estimation; minimax risk

References

[1] Aamari, E., Aaron, C. and Levrard, C. (2023). Supplement to "Minimax boundary estimation and estimation with boundary". https://doi.org/10.3150/23-BEJ1585SUPP
[2] Aamari, E., Kim, J., Chazal, F., Michel, B., Rinaldo, A. and Wasserman, L. (2019). Estimating the reach of a manifold. *Electron. J. Stat.* **13** 1359–1399.
[3] Aamari, E. and Levrard, C. (2018). Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete Comput. Geom.* **59** 923–971. MR3802310 https://doi.org/10.1007/s00454-017-9962-z
[4] Aamari, E. and Levrard, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204. MR3909931 https://doi.org/10.1214/18-AOS1685
[5] Aaron, C. and Bodart, O. (2016). Local convex hull support and boundary estimation. *J. Multivariate Anal.* **147** 82–101. MR3484171 https://doi.org/10.1016/j.jmva.2016.01.003
[6] Aaron, C. and Cholaquidis, A. (2020). On boundary detection. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2028–2050. MR4116716 https://doi.org/10.1214/19-AIHP1027
[7] Aaron, C., Cholaquidis, A. and Fraiman, R. (2022). Estimation of surface area. *Electron. J. Stat.* **16** 3751–3788. MR4448877 https://doi.org/10.1214/22-ejs2031
[8] Aizenbud, Y. and Sober, B. (2021). Non-parametric estimation of manifolds from noisy data. ArXiv E-prints. Available at arXiv:2105.04754.
[9] Belkin, M., Niyogi, P. and Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **7** 2399–2434. MR2274444
[10] Berenfeld, C., Harvey, J., Hoffmann, M. and Shankar, K. (2022). Estimating the reach of a manifold via its convexity defect function. *Discrete Comput. Geom.* **67** 403–438. MR4376572 https://doi.org/10.1007/s00454-021-00290-8
[11] Berry, T. and Sauer, T. (2017). Density estimation on manifolds with boundary. *Comput. Statist. Data Anal.* **107** 1–17. MR3575055 https://doi.org/10.1016/j.csda.2016.09.011

[12] Boissonnat, J.-D. and Ghosh, A. (2014). Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.* **51** 221–267. MR3148657 https://doi.org/10.1007/s00454-013-9557-2

[13] Boissonnat, J.-D., Guibas, L.J. and Oudot, S.Y. (2009). Manifold reconstruction in arbitrary dimensions using witness complexes. *Discrete Comput. Geom.* **42** 37–70. MR2506737 https://doi.org/10.1007/s00454-009-9175-1

[14] Boissonnat, J.-D. and Wintraecken, M. (2020). The topological correctness of PL-approximations of isomanifolds. In *36th International Symposium on Computational Geometry (SoCG 2020)* (S. Cabello and D.Z. Chen, eds.). *Leibniz International Proceedings in Informatics (LIPIcs)* **164** 20:1–20:18. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[15] Bredon, G.E. (1993). *Topology and Geometry*. *Graduate Texts in Mathematics* **139**. New York: Springer. MR1224675 https://doi.org/10.1007/978-1-4757-6848-0

[16] Calder, J., Park, S. and Slepčev, D. (2022). Boundary estimation from point clouds: Algorithms, guarantees and applications. *J. Sci. Comput.* **92** 56. MR4448202 https://doi.org/10.1007/s10915-022-01894-9

[17] Chazal, F., Glisse, M., Labruère, C. and Michel, B. (2015). Convergence rates for persistence diagram estimation in topological data analysis. *J. Mach. Learn. Res.* **16** 3603–3635. MR3450548

[18] Chazal, F. and Michel, B. (2017). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. ArXiv E-prints. Available at arXiv:1710.04019.

[19] Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* **36** 340–354. MR2058139 https://doi.org/10.1239/aap/1086957575

[20] Dey, T.K., Li, K., Ramos, E.A. and Wenger, R. (2009). Isotopic reconstruction of surfaces with boundaries. In *Computer Graphics Forum* **28** 1371–1382. Wiley Online Library.

[21] Divol, V. (2021). Minimax adaptive estimation in manifold inference. *Electron. J. Stat.* **15** 5888–5932. MR4355700 https://doi.org/10.1214/21-ejs1934

[22] Divol, V. (2021). Reconstructing measures on manifolds: An optimal transport approach. ArXiv E-prints. Available at arXiv:2102.07595.

[23] Dümbgen, L. and Walther, G. (1996). Rates of convergence for random approximations of convex sets. *Adv. in Appl. Probab.* **28** 384–393. MR1387882 https://doi.org/10.2307/1428063

[24] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078 https://doi.org/10.2307/1993504

[25] Federer, H. (1969). *Geometric Measure Theory*. *Die Grundlehren der Mathematischen Wissenschaften, Band 153*. New York: Springer. MR0257325

[26] Fefferman, C., Ivanov, S., Lassas, M. and Narayanan, H. (2019). Fitting a manifold of large reach to noisy data. ArXiv E-prints. Available at arXiv:1910.05084.

[27] Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012). Minimax manifold estimation. *J. Mach. Learn. Res.* **13** 1263–1291. MR2930639

[28] Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939 https://doi.org/10.1214/12-AOS994

[29] Har-Peled, S. (2011). *Geometric Approximation Algorithms*. *Mathematical Surveys and Monographs* **173**. Providence, RI: Amer. Math. Soc. MR2760023 https://doi.org/10.1090/surv/173

[30] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed. *Springer Series in Statistics*. New York: Springer. Data mining, inference, and prediction. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[31] Hatcher, A. (2002). *Algebraic Topology*. Cambridge: Cambridge Univ. Press. MR1867354

[32] Hirsch, M.W. (1976). *Differential Topology*. *Graduate Texts in Mathematics*, **33**. New York: Springer. MR0448362

[33] Kim, A.K.H. and Zhou, H.H. (2015). Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.* **9** 1562–1582. MR3376117 https://doi.org/10.1214/15-EJS1039

[34] Lee, J.A. and Verleysen, M. (2007). *Nonlinear Dimensionality Reduction*. *Information Science and Statistics*. New York: Springer. MR2373983 https://doi.org/10.1007/978-0-387-39351-3

[35] Lee, J.M. (2011). *Introduction to Topological Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **202**. New York: Springer. MR2766102 https://doi.org/10.1007/978-1-4419-7940-7

[36] Maggioni, M., Minsker, S. and Strawn, N. (2016). Multiscale dictionary learning: Non-asymptotic bounds and robustness. *J. Mach. Learn. Res.* **17** 2. MR3482922

[37] Mammen, E. and Tsybakov, A.B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579 https://doi.org/10.1214/aos/1176324533

[38] Møller, J. (1989). Random tessellations in $\mathbf{R}^d$. *Adv. in Appl. Probab.* **21** 37–73. MR0980736 https://doi.org/10.2307/1427197

[39] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768 https://doi.org/10.1007/s00454-008-9053-2

[40] Puchkin, N. and Spokoiny, V. (2022). Structure-adaptive manifold estimation. *J. Mach. Learn. Res.* **23** 40. MR4420765

[41] Rineau, L. and Yvinec, M. (2008). Meshing 3D domains bounded by piecewise smooth surfaces*. In *Proceedings of the 16th International Meshing Roundtable* (M.L. Brewer and D. Marcum, eds.) 443–460. Berlin, Heidelberg: Springer.

[42] Rodríguez Casal, A. (2007). Set estimation under convexity type assumptions. *Ann. Inst. Henri Poincaré Probab. Stat.* **43** 763–774. MR3252430 https://doi.org/10.1016/j.anihpb.2006.11.001

[43] Sharma, A. and Paliwal, K.K. (2007). Fast principal component analysis using fixed-point algorithm. *Pattern Recogn. Lett.* **28** 1151–1155.

[44] Sheehy, D.R. (2015). An output-sensitive algorithm for computing weighted α-complexes. In *CCCG*.

[45] Wasserman, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.* **5** 501–535. MR3774757 https://doi.org/10.1146/annurev-statistics-031017-100045

Bernoulli **29**(4), 2023, 3369–3395
https://doi.org/10.3150/23-BEJ1587

Necessary and sufficient conditions for the asymptotic normality of higher order Turing estimators

JIE CHANG[a] and MICHAEL GRABCHAK[b]

Department of Mathematics and Statistics, UNC Charlotte, Charlotte, NC, USA, [a]*jchang35@uncc.edu,* [b]*mgrabcha@uncc.edu*

This paper establishes necessary and sufficient conditions for the asymptotic normality of higher order Turing estimators. It further gives several easy to verify sufficient conditions. These conditions are then used to show that the assumptions hold for large classes of distributions with regularly varying tails. This includes classes for which asymptotic normality had not been previously verified.

Keywords: Missing mass; occupancy probabilities; Turing estimators

References

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*. New York: Dover. MR0208797

Ayed, F., Battiston, M., Camerlenghi, F. and Favaro, S. (2021). On consistent and rate optimal estimation of the missing mass. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1476–1494. MR4291456 https://doi.org/10.1214/20-aihp1126

Barbour, A.D. and Gnedin, A.V. (2009). Small counts in the infinite occupancy scheme. *Electron. J. Probab.* **14** 365–384. MR2480545 https://doi.org/10.1214/EJP.v14-608

Ben-Hamou, A., Boucheron, S. and Ohannessian, M.I. (2017). Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli* **23** 249–287. MR3556773 https://doi.org/10.3150/15-BEJ743

Billingsley, P. (1995). *Probability and Measure*. 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York-Chichester-Brisbane: Wiley. MR0534323

Chang, J. and Grabchak, M. (2023). Supplement to "Necessary and sufficient conditions for the asymptotic normality of higher order Turing estimators." https://doi.org/10.3150/23-BEJ1587SUPP

Chao, A. (1981). On estimating the probability of discovering a new species. *Ann. Statist.* **9** 1339–1342. MR0630117

Chao, A., Lee, S.-M. and Chen, T.-C. (1988). A generalized Good's nonparametric coverage estimator. *Chinese J. Math.* **16** 189–199. MR0993611

Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. and Gotelli, N.J. (2015). Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology* **96** 1189–1201. https://doi.org/10.1890/14-0550.1

Chebunin, M. and Zuyev, S. (2022). Functional central limit theorems for occupancies and missing mass process in infinite urn models. *J. Theoret. Probab.* **35** 1–19. MR4379457 https://doi.org/10.1007/s10959-020-01053-6

Chen, S.F. and Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Comput. Speech Lang.* **13** 359–394.

Decrouez, G., Grabchak, M. and Paris, Q. (2018). Finite sample properties of the mean occupancy counts and probabilities. *Bernoulli* **24** 1910–1941. MR3757518 https://doi.org/10.3150/16-BEJ915

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

1350-7265

Esty, W.W. (1982). Confidence intervals for the coverage of low coverage samples. *Ann. Statist.* **10** 190–196. MR0642730

Esty, W.W. (1983). A normal limit law for a nonparametric estimator of the coverage of a random sample. *Ann. Statist.* **11** 905–912. MR0707940 https://doi.org/10.1214/aos/1176346256

Gao, F. (2013). Moderate deviations for a nonparametric estimator of sample coverage. *Ann. Statist.* **41** 641–669. MR3099116 https://doi.org/10.1214/13-AOS1091

Gnedin, A., Hansen, B. and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Probab. Surv.* **4** 146–171. MR2318403 https://doi.org/10.1214/07-PS092

Gnedin, A. and Iksanov, A. (2020). On nested infinite occupancy scheme in random environment. *Probab. Theory Related Fields* **177** 855–890. MR4126933 https://doi.org/10.1007/s00440-020-00963-0

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264. MR0061330 https://doi.org/10.1093/biomet/40.3-4.237

Good, I.J. and Toulmin, G.H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63. MR0077039 https://doi.org/10.1093/biomet/43.1-2.45

Grabchak, M. and Cosme, V. (2017). On the performance of Turing's formula: A simulation study. *Comm. Statist. Simulation Comput.* **46** 4199–4209. MR3672540 https://doi.org/10.1080/03610918.2015.1109658

Grabchak, M., Kelbert, M. and Paris, Q. (2020). On the occupancy problem for a regime-switching model. *J. Appl. Probab.* **57** 53–77. MR4094383 https://doi.org/10.1017/jpr.2020.33

Grabchak, M. and Zhang, Z. (2017). Asymptotic properties of Turing's formula in relative error. *Mach. Learn.* **106** 1771–1785. MR3708277 https://doi.org/10.1007/s10994-016-5620-6

Gut, A. (2012). Anscombe's theorem 60 years later. *Sequential Anal.* **31** 368–396. MR2958641

Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17** 373–401. MR0216548 https://doi.org/10.1512/iumj.1968.17.17020

Mao, C.X. and Lindsay, B.G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika* **89** 669–681. MR1929171 https://doi.org/10.1093/biomet/89.3.669

McAllester, D. and Ortiz, L. (2003). Concentration inequalities for the missing mass and for histogram rule error. *J. Mach. Learn. Res.* **4** 895–911. MR2076001 https://doi.org/10.1162/1532443041424292

McAllester, D.A. and Schapire, R.E. (2000). On the convergence rate of Good-Turing estimators. In *COLT'00: Proceedings of the Thirteenth Annual Conference on Computational Learning Theory* 1–6.

Mossel, E. and Ohannessian, M.I. (2019). On the impossibility of learning the missing mass. *Entropy* **21**. https://doi.org/10.3390/e21010028

Ohannessian, M.I. and Dahleh, M.A. (2012). Rare probability estimation under regularly varying heavy tails. In *Proceedings of the 25th Annual Conference on Learning Theory*. *Proceedings of Machine Learning Research* **23** 21.1–21.24.

Robbins, H.E. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Stat.* **39** 256–257. MR0221695 https://doi.org/10.1214/aoms/1177698526

Thisted, R. and Efron, B. (1987). Did Shakespeare write a newly-discovered poem? *Biometrika* **74** 445–455. MR0909350 https://doi.org/10.1093/biomet/74.3.445

Wagner, A.B., Viswanath, P. and Kulkarni, S.R. (2011). Probability estimation in the rare-events regime. *IEEE Trans. Inf. Theory* **57** 3207–3229. MR2817014 https://doi.org/10.1109/TIT.2011.2137210

Zhang, C.-H. (2005). Estimation of sums of random variables: Examples and information bounds. *Ann. Statist.* **33** 2022–2041. MR2211078 https://doi.org/10.1214/009053605000000390

Zhang, Z. (2013). A multivariate normal law for Turing's formulae. *Sankhya A* **75** 51–73. MR3061846 https://doi.org/10.1007/s13171-012-0019-3

Zhang, Z. (2017). *Statistical Implications of Turing's Formula*. Hoboken, NJ: Wiley. MR3642750

Zhang, Z. (2018). On normal law conditions for Turing's formula. In *Wiley StatsRef: Statistics Reference Online* (N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J.L. Teugels, eds.). https://doi.org/10.1002/9781118445112.stat08084

Zhang, Z. and Huang, H. (2007). Turing's formula revisited. *J. Quant. Ling.* **14** 222–241.

Zhang, Z. and Huang, H. (2008). A sufficient normality condition for Turing's formula. *J. Nonparametr. Stat.* **20** 431–446. MR2424251 https://doi.org/10.1080/10485250802172126

Zhang, C.-H. and Zhang, Z. (2009). Asymptotic normality of a nonparametric estimator of sample coverage. *Ann. Statist.* **37** 2582–2595. MR2543704 https://doi.org/10.1214/08-AOS658

Bernoulli **29**(4), 2023, 3396–3422
https://doi.org/10.3150/23-BEJ1588

Characterization of the second order random fields subject to linear distributional PDE constraints

IAIN HENDERSON[a], PASCAL NOBLE[b] and OLIVIER ROUSTANT[c]

Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, INSA, F-31077 Toulouse, France, [a]*henderso@insa-toulouse.fr,* [b]*noble@insa-toulouse.fr,* [c]*roustant@insa-toulouse.fr*

Let L be a linear differential operator acting on functions defined over an open set $\mathcal{D} \subset \mathbb{R}^d$. In this article, we characterize the measurable second order random fields $U = (U(x))_{x\in\mathcal{D}}$ whose sample paths all verify the partial differential equation (PDE) $L(u) = 0$, solely in terms of their first two moments. When compared to previous similar results, the novelty lies in that the equality $L(u) = 0$ is understood in the *sense of distributions*, which is a powerful functional analysis framework mostly designed to study linear PDEs. This framework enables to reduce to the minimum the required differentiability assumptions over the first two moments of $(U(x))_{x\in\mathcal{D}}$ as well as over its sample paths in order to make sense of the PDE $L(U_\omega) = 0$. In view of Gaussian process regression (GPR) applications, we show that when $(U(x))_{x\in\mathcal{D}}$ is a Gaussian process (GP), the sample paths of $(U(x))_{x\in\mathcal{D}}$ conditioned on pointwise observations still verify the constraint $L(u) = 0$ in the distributional sense. We finish by deriving a simple but instructive example, a GP model for the 3D linear wave equation, for which our theorem is applicable and where the previous results from the literature do not apply in general.

Keywords: Generalized functions; linear constraints; linear partial differential equations; second order random fields

References

[1] Adler, R.J. (2010). *The Geometry of Random Fields*. *Classics in Applied Mathematics* **62**. Philadelphia, PA: SIAM. MR3396215 https://doi.org/10.1137/1.9780898718980

[2] Adler, R.J. and Taylor, J.E. (2007). *Random Fields and Geometry*. *Springer Monographs in Mathematics*. New York: Springer. MR2319516 https://doi.org/10.1007/978-0-387-48116-6

[3] Albert, C.G. and Rath, K. (2020). Gaussian process regression for data fulfilling linear differential equations with localized sources. *Entropy* **22** Paper No. 152, 16 pp. MR4075751 https://doi.org/10.3390/e22020152

[4] Álvarez, M., Luengo, D. and Lawrence, N. (2013). Linear latent force models using Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Learn. Intell.* **35** 2693–2705. https://doi.org/10.1109/TPAMI.2013.86

[5] Azaïs, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Hoboken, NJ: Wiley. MR2478201 https://doi.org/10.1002/9780470434642

[6] Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston, MA: Kluwer Academic. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

[7] Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. *Universitext*. New York: Springer. MR2759829 https://doi.org/10.1007/978-0-387-70914-7

[8] Carrizo Vergara, R., Allard, D. and Desassis, N. (2022). A general framework for SPDE-based stationary random fields. *Bernoulli* **28** 1–32. MR4337696 https://doi.org/10.3150/20-bej1317

[9] Chen, Y., Hosseini, B., Owhadi, H. and Stuart, A.M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *J. Comput. Phys.* **447** Paper No. 110668, 29 pp. MR4311012 https://doi.org/10.1016/j.jcp.2021.110668

1350-7265

[10] Da Prato, G. and Zabczyk, J. (1992). *Stochastic Equations in Infinite Dimensions*. *Encyclopedia of Mathematics and Its Applications* **44**. Cambridge: Cambridge Univ. Press. MR1207136 https://doi.org/10.1017/CBO9780511666223

[11] Dieudonné, J. (1954). Sur les espaces de Montel métrisables. *C. R. Acad. Sci. Paris* **238** 194–195. MR0058854

[12] Doob, J.L. (1937). Stochastic processes depending on a continuous parameter. *Trans. Amer. Math. Soc.* **42** 107–140. MR1501916 https://doi.org/10.2307/1989677

[13] Doob, J.L. (1990). *Stochastic Processes*. *Wiley Classics Library*. New York: Wiley. MR1038526

[14] Duistermaat, J.J. and Kolk, J.A.C. (2010). *Distributions: Theory and Applications*. *Cornerstones*. Boston, MA: Birkhäuser. MR2680692 https://doi.org/10.1007/978-0-8176-4675-2

[15] Estrade, A. and Fournier, J. (2020). Anisotropic Gaussian wave models. *ALEA Lat. Am. J. Probab. Math. Stat.* **17** 329–353. MR4105921 https://doi.org/10.30757/alea.v17-13

[16] Evans, L.C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Providence, RI: Amer. Math. Soc. MR2597943 https://doi.org/10.1090/gsm/019

[17] Evans, L.C. and Gariepy, R.F. (2015). *Measure Theory and Fine Properties of Functions*, Revised ed. *Textbooks in Mathematics*. Boca Raton, FL: CRC Press. MR3409135 https://doi.org/10.1201/b18333

[18] Wahlström, N. and Kok, M. and Schön, T.B. and Gustafsson, F. (2013). Modeling magnetic fields using Gaussian processes. *IEEE Trans. Acoust. Speech Signal Process.* 3522–3526. https://doi.org/10.1109/ICASSP.2013.6638313

[19] Fan, M., Paul, D., Lee, T.C.M. and Matsuo, T. (2018). Modeling tangential vector fields on a sphere. *J. Amer. Statist. Assoc.* **113** 1625–1636. MR3902234 https://doi.org/10.1080/01621459.2017.1356322

[20] Fasshauer, G.E. (2007). *Meshfree Approximation Methods with MATLAB*. *Interdisciplinary Mathematical Sciences* **6**. Hackensack, NJ: World Scientific. MR2357267 https://doi.org/10.1142/6437

[21] Fiedler, J. (2016). Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields Ph.D. thesis. https://doi.org/10.11588/heidok.00022193

[22] Fuselier, E.J. Jr (2006). *Refined Error Estimates for Matrix-Valued Radial Basis Functions*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)–Texas A&M University. MR2710397

[23] Gapaillard, J. and Michaux, J. (1986/87). Sur les processus linéaires définis sur un espace nucléaire. *Ann. Fac. Sci. Toulouse Math. (5)* **8** 75–92. MR0899598

[24] Geist, A. and Trimpe, S. (2020). Learning Constrained Dynamics with Gauss' Principle adhering Gaussian Processes. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*. *Proc. Mach. Learn. Res.* **120** 225–234. PMLR.

[25] Gel'fand, I.M. and Vilenkin, N.Ya. (1964). *Generalized Functions. Vol. 4: Applications of Harmonic Analysis*. New York: Academic Press. MR0173945 https://doi.org/10.1016/C2013-0-12221-0

[26] Ginsbourger, D., Roustant, O. and Durrande, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *J. Statist. Plann. Inference* **170** 117–128. MR3427864 https://doi.org/10.1016/j.jspi.2015.10.002

[27] Graepel, T. (2003). Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML), Washington, DC, USA* 234–241. AAAI Press.

[28] Gulian, M., Frankel, A. and Swiler, L. (2022). Gaussian process regression constrained by boundary value problems. *Comput. Methods Appl. Mech. Engrg.* **388** Paper No. 114117, 18 pp. MR4330866 https://doi.org/10.1016/j.cma.2021.114117

[29] Janson, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge: Cambridge Univ. Press. MR1474726 https://doi.org/10.1017/CBO9780511526169

[30] Jidling, C., Hendriks, J., Wahlstrom, N., Gregg, A., Schon, T., Wensrich, C. and Wills, A. (2018). Probabilistic modelling and reconstruction of strain. *Nucl. Instrum. Methods Phys. Res., Sect. B, Beam Interact. Mater. Atoms* **436** 141–155. https://doi.org/10.1016/j.nimb.2018.08.051

[31] Jidling, C., Wahlström, N., Wills, A. and Schön, T.B. (2017). Linearly constrained Gaussian processes. In *Adv. Neural Inf. Process Syst.* **30**. Curran Associates.

[32] Kuchment, P. and Kunyansky, L. (2015). Mathematics of photoacoustic and thermoacoustic tomography. In *Handbook of Mathematical Methods in Imaging. Vol. 1, 2, 3* 1117–1167. New York: Springer. MR3560086

[33] Lang, S. (1993). *Real and Functional Analysis*, 3rd ed. *Graduate Texts in Mathematics* **142**. New York: Springer. MR1216137 https://doi.org/10.1007/978-1-4612-0897-6

[34] Lange-Hegermann, M. (2018). Algorithmic linearly constrained Gaussian processes. In *Adv. Neural Inf. Process Syst.* **31**. Curran Associates.

[35] Lange-Hegermann, M. (2021). Linearly constrained Gaussian processes with boundary conditions. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. *Proc. Mach. Learn. Res.* **130** 1090–1098. PMLR.

[36] Le Gall, J.-F. (2013). *Mouvement Brownien, Martingales et Calcul Stochastique*. *Mathématiques & Applications* **71**. Heidelberg: Springer. MR3184878 https://doi.org/10.1007/978-3-642-31898-6

[37] Lindgren, F., Bolin, D. and Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spat. Stat.* **50** Paper No. 100599, 29 pp. MR4439328 https://doi.org/10.1016/j.spasta.2022.100599

[38] Mendes, F.M. and da Costa Júnior, E.A. (2012). Bayesian inference in the numerical solution of Laplace's equation. *AIP Conf. Proc.* **1443** 72–79. https://doi.org/10.1063/1.3703622

[39] Narcowich, F.J. and Ward, J.D. (1994). Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Math. Comp.* **63** 661–687. MR1254147 https://doi.org/10.2307/2153288

[40] Nguyen, N.C. and Peraire, J. (2015). Gaussian functional regression for linear partial differential equations. *Comput. Methods Appl. Mech. Engrg.* **287** 69–89. MR3318661 https://doi.org/10.1016/j.cma.2015.01.008

[41] Owhadi, H. (2015). Bayesian numerical homogenization. *Multiscale Model. Simul.* **13** 812–828. MR3369060 https://doi.org/10.1137/140974596

[42] Pazy, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. *Applied Mathematical Sciences* **44**. New York: Springer. MR0710486 https://doi.org/10.1007/978-1-4612-5561-1

[43] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2017). Machine learning of linear differential equations using Gaussian processes. *J. Comput. Phys.* **348** 683–693. MR3689653 https://doi.org/10.1016/j.jcp.2017.07.050

[44] Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2514435 https://doi.org/10.7551/mitpress/3206.001.0001

[45] Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Berlin: Springer. MR1763973 https://doi.org/10.1007/BFb0103934

[46] Roques, L., Allard, D. and Soubeyrand, S. (2022). Spatial statistics and stochastic partial differential equations: A mechanistic viewpoint. *Spat. Stat.* **50** Paper No. 100591, 25 pp. MR4439325 https://doi.org/10.1016/j.spasta.2022.100591

[47] Rudin, W. (1991). *Functional Analysis*, 2nd ed. *International Series in Pure and Applied Mathematics*. New York: McGraw-Hill, Inc. MR1157815

[48] Schaback, R. (2009). Solving the Laplace equation by meshless collocation using harmonic kernels. *Adv. Comput. Math.* **31** 457–470. MR2558262 https://doi.org/10.1007/s10444-008-9078-3

[49] Schaefer, H.H. and Wolff, M.P. (1999). *Topological Vector Spaces*, 2nd ed. *Graduate Texts in Mathematics* **3**. New York: Springer. MR1741419 https://doi.org/10.1007/978-1-4612-1468-7

[50] Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Process. Appl.* **120** 1879–1897. MR2673978 https://doi.org/10.1016/j.spa.2010.05.009

[51] Scheuerer, M. and Schlather, M. (2012). Covariance models for divergence-free and curl-free random vector fields. *Stoch. Models* **28** 433–451. MR2959449 https://doi.org/10.1080/15326349.2012.699756

[52] Schwartz, L. (1964). Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *J. Anal. Math.* **13** 115–256. MR0179587 https://doi.org/10.1007/BF02786620

[53] Serre, D. (1999). *Systems of Conservation Laws. Vol 1.* Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511612374

[54] Steinwart, I. (2019). Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *Potential Anal.* **51** 361–395. MR4023468 https://doi.org/10.1007/s11118-018-9715-5

[55] Trèves, F. (1967). *Topological Vector Spaces, Distributions and Kernels*. New York: Academic Press. MR0225131

[56] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450 https://doi.org/10.1093/biomet/41.3-4.434

Bernoulli **29**(4), 2023, 3423–3442
https://doi.org/10.3150/23-BEJ1589

Near-optimal estimation of the unseen under regularly varying tail populations

STEFANO FAVARO[1,a] and ZACHARIE NAULET[2,b]

[1]*Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto, Torino, 10134, Italy,* [a]*stefano.favaro@unito.it*
[2]*Laboratoire de mathématiques d'Orsay, CNRS, Université Paris-Saclay, 91405 Orsay, France,* [b]*zacharie.naulet@universite-paris-saclay.fr*

Given n samples from a population of individuals belonging to different species, what is the number U of hitherto unseen species that would be observed if λn new samples were collected? This is the celebrated unseen-species problem, which has been the subject of recent breakthrough studies introducing non-parametric estimators of U that are minimax near-optimal and consistent all the way up to $\lambda \asymp \log n$. These works do not rely on assumptions on the underlying unknown distribution p of the population, and therefore, while providing a theory in its greatest generality, worst-case distributions may hamper the estimation of U in concrete settings. In this paper, we strengthen the non-parametric framework for estimating U, making use of suitable assumptions on p. Inspired by the estimation of rare probabilities in extreme value theory, and motivated by the ubiquitous power-law type distributions in many natural and social phenomena, we make use of a semi-parametric assumption of regular variation of index $\alpha \in (0,1)$ for the tail behaviour of p. Under this assumption, we introduce an estimator of U that is simple, linear in the sampling information, computationally efficient, and scalable to massive datasets. Then, uniformly over our class of regularly varying tail distributions, we show that the proposed estimator has provable guarantees: i) it is minimax near-optimal, up to a power of $\log n$ factor; ii) it is consistent all of the way up to $\log \lambda \asymp n^{\alpha/2}/\sqrt{\log n}$, and this range is the best possible. This is the first study on the estimation of the unseen under regularly varying tail distributions p. Our results rely on a novel approach, of independent interest, which combines the renowned method of the two fuzzy hypotheses for minimax estimation of discrete functionals, with Bayesian arguments under Poisson-Kingman priors for p. An illustration of our method is presented for synthetic and real data.

Keywords: Multinomial model; optimal minimax estimation; Poisson-Kingman prior; power-law data; regularly varying tails; tail-index; useen-species problem

References

[1] Anevski, D., Gill, R.D. and Zohren, S. (2017). Estimating a probability mass function with unknown labels. *Ann. Statist.* **45** 2708–2735. MR3737907 https://doi.org/10.1214/17-AOS1542
[2] Ayed, F., Battiston, M., Camerlenghi, F. and Favaro, S. (2021). On consistent and rate optimal estimation of the missing mass. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1476–1494. MR4291456 https://doi.org/10.1214/20-aihp1126
[3] Balabdaoui, F. and Kulagina, Y. (2020). Completely monotone distributions: Mixing, approximation and estimation of number of species. *Comput. Statist. Data Anal.* **150** 107014. MR4101998 https://doi.org/10.1016/j.csda.2020.107014
[4] Balocchi, C., Favaro, S. and Naulet, Z. (2021). Bayesian nonparametric inference for "species-sampling problems". Preprint. Available at arXiv:2203.06076.
[5] Barabási, A.L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature* **435** 227.
[6] Ben-Hamou, A., Boucheron, S. and Ohannessian, M.I. (2017). Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli* **23** 249–287. MR3556773 https://doi.org/10.3150/15-BEJ743

[7] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1989). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR1015093
[8] Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: A review. *J. Amer. Statist. Assoc.* **88** 364–373.
[9] Camerlenghi, F., Favaro, S., Naulet, Z. and Panero, F. (2021). Optimal disclosure risk assessment. *Ann. Statist.* **49** 723–744. MR4255105 https://doi.org/10.1214/20-aos1975
[10] Cancho, R.F. and Solé, R.V. (2003). Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. USA* **100** 788–791. MR1951090 https://doi.org/10.1073/pnas.0335980100
[11] Chao, A. and Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Amer. Statist. Assoc.* **87** 210–217. MR1158639
[12] Chee, C.-S. and Wang, Y. (2016). Nonparametric estimation of species richness using discrete k-monotone distributions. *Comput. Statist. Data Anal.* **93** 107–118. MR3406199 https://doi.org/10.1016/j.csda.2014.10.021
[13] Clauset, A., Shalizi, C.R. and Newman, M.E.J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51** 661–703. MR2563829 https://doi.org/10.1137/070710111
[14] Daley, T. and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. *Nat. Methods* **10** 325–327. https://doi.org/10.1038/nmeth.2375
[15] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3
[16] Drees, H. (1998). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26** 434–448. MR1608148 https://doi.org/10.1214/aos/1030563992
[17] Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.
[18] Favaro, S. and Naulet, Z. (2023). Supplement to "Near-optimal estimation of the unseen under regularly varying tail populations." https://doi.org/10.3150/23-BEJ1589SUPP
[19] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. New York: Wiley. MR0270403
[20] Fisher, R.A., Corbet, A.S. and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12** 42–58.
[21] Formentin, M., Lovison, A., Maritan, A. and Zanzotto, G. (2014). Hidden scaling patterns and universality in written communication. *Phys. Rev. E* **90** 012817.
[22] Gao, Z., Tseng, C.H., Pei, Z. and Blaser, M.J. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proc. Natl. Acad. Sci. USA* **104** 2927–2932.
[23] Giguelay, J. and Huet, S. (2018). Testing k-monotonicity of a discrete distribution. Application to the estimation of the number of classes in a population. *Comput. Statist. Data Anal.* **127** 96–115. MR3820312 https://doi.org/10.1016/j.csda.2018.02.006
[24] Gnedin, A., Hansen, B. and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Probab. Surv.* **4** 146–171. MR2318403 https://doi.org/10.1214/07-PS092
[25] Good, I.J. and Toulmin, G.H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63. MR0077039 https://doi.org/10.1093/biomet/43.1-2.45
[26] Hall, P. and Welsh, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084. MR0751294 https://doi.org/10.1214/aos/1176346723
[27] Hall, P. and Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331–341. MR0773171 https://doi.org/10.1214/aos/1176346596
[28] Hao, Y. and Li, P. (2020). Optimal prediction of the number of unseen species with multiplicity. In *Advances in Neural Information Processing Systems*.
[29] Harald, B.R. (2001). *Word Frequency Distributions*. Berlin: Springer.
[30] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204
[31] Huberman, B.A. and Adamic, L.A. (1999). Internet: Growth dynamics of the World-Wide Web. *Nature* **401** 131.
[32] Ionita-Laza, I., Lange, C. and Laird, N.M. (2009). Estimating the number of unseen variants in the human genome. *Proc. Natl. Acad. Sci. USA* **106** 5008–5013. MR2496537 https://doi.org/10.1073/pnas.0807815106

[33] Jana, S., Polyanskiy, Y. and Wu, Y. (2020). Extrapolating the profile of a finite population. In *Conference on Learning Theory*.
[34] Jiao, J., Venkat, K., Han, Y. and Weissman, T. (2015). Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inf. Theory* **61** 2835–2885. MR3342309 https://doi.org/10.1109/TIT.2015.2412945
[35] Kingman, J.F.C. (1993). *Poisson Processes*. *Oxford Studies in Probability* **3**. Oxford University Press, New York: The Clarendon Press. MR1207584
[36] Kingman, J.F.C., Taylor, S.J., Hawkes, A.G., Walker, A.M., Cox, D.R., Smith, A.F.M., Hill, B.M., Burville, P.J. and Leonard, T. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** 1–22. MR0368264
[37] Kroes, I., Lepp, P.W. and Relman, D.A. (1999). Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci. USA* **96** 14547–14552.
[38] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N.L. Hjort, C.C. Holmes, P. Müller and S.G. Walker, eds.). *Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge: Cambridge Univ. Press. MR2730661
[39] Monechi, B., Ruiz-Serrano, Ã., Tria, F. and Loreto, V. (2017). Waves of novelties in the expansion into the adjacent possible. *PLoS ONE* **12** e0179303. https://doi.org/10.1371/journal.pone.0179303
[40] Mossel, E. and Ohannessian, M.I. (2019). On the impossibility of learning the missing mass. *Entropy* **21**. https://doi.org/10.3390/e21010028
[41] Muchnik, L., Pei, S., Parra, L.C., Reis, S.D.S., Andrade, J.S., Havlin, S. and Makse, H.A. (2013). Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Nature Scientific Reports* **3** 1783.
[42] Ohannessian, M.I. and Dahleh, M.A. (2012). Rare probability estimation under regularly varying heavy tails. *Journal of Machine Learning Reseach* **23** 1–24.
[43] Orlitsky, A., Suresh, A.T. and Wu, Y. (2016). Optimal prediction of the number of unseen species. *Proc. Natl. Acad. Sci. USA* **113** 13283–13288. MR3582444 https://doi.org/10.1073/pnas.1607774113
[44] Pitman, J. (2003). Poisson-Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed* (D.R. Goldstein, ed.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. Beachwood, OH: IMS. MR2004330 https://doi.org/10.1214/lnms/1215091133
[45] Pitman, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Berlin: Springer. MR2245368
[46] Polyanskiy, Y. and We, Y. (2020). Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens. Preprint. Available at arXiv:1902.05616.
[47] Rybski, D., Buldyrev, S.V., Havlin, S., Liljeros, F. and Makse, H.A. (2016). Scaling laws of human interaction activity. *Proc. Natl. Acad. Sci. USA* **106** 12640.
[48] Tria, F., Loreto, V., Servedio, V.D.P. and Strogatz, S.H. (2014). The dynamics of correlated novelties. *Nature Scientific Reports* **4** 5890.
[49] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359 https://doi.org/10.1007/b13794
[50] Valiant, G. and Valiant, P. (2013). Estimating the unseen: Improved estimators for entropy and other properties. In *Advances in Neural Information Processing Systems* **27** 2157–2165.
[51] Wu, Y. and Yang, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory* **62** 3702–3720. MR3506758 https://doi.org/10.1109/TIT.2016.2548468
[52] Wu, Y. and Yang, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann. Statist.* **47** 857–883. MR3909953 https://doi.org/10.1214/17-AOS1665
[53] Zipf, G.K. (1949). *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley.

Bernoulli **29**(4), 2023, 3443–3469
https://doi.org/10.3150/23-BEJ1590

Joint density of the stable process and its supremum: Regularity and upper bounds

JORGE IGNACIO GONZÁLEZ CÁZARES[1,3,a], ARTURO KOHATSU-HIGA[2,c] and ALEKSANDAR MIJATOVIĆ[1,3,b]

[1]*Department of Statistics, University of Warwick, Coventry, UK,* [a]*jorge.i.gonzalez-cazares@warwick.ac.uk,* [b]*a.mijatovic@warwick.ac.uk*
[2]*Ritsumeikan University, Nojihigashi, Kusatsu, Shiga, Japan,* [c]*khts00@fc.ritsumei.ac.jp*
[3]*The Alan Turing Institute, London, UK*

This article uses a combination of three ideas from simulation to establish a nearly optimal polynomial upper bound for the joint density of the stable process and its associated supremum at a fixed time on the entire support of the joint law. The representation of the concave majorant of the stable process and the Chambers-Mallows-Stuck representation for stable laws are used to define an approximation of the random vector of interest. An interpolation technique using multilevel Monte Carlo is applied to accelerate the approximation, allowing us to establish the infinite differentiability of the joint density as well as nearly optimal polynomial upper bounds for the joint mixed derivatives of any order.

Keywords: Joint density bounds; stable supremum

References

[1] Bally, V. and Caramellino, L. (2016). *Stochastic Integration by Parts*. *Advanced Courses in Mathematics–CRM Barcelona*. Springer International Publishing.
[2] Bernyk, V., Dalang, R.C. and Peskir, G. (2011). Predicting the ultimate supremum of a stable Lévy process with no negative jumps. *Ann. Probab.* **39** 2385–2423. MR2932671 https://doi.org/10.1214/10-AOP598
[3] Bhattacharya, R.N. and Ranga Rao, R. (1986). *Normal Approximation and Asymptotic Expansions*. Melbourne, FL: Robert E. Krieger Publishing Co., Inc. Reprint of the 1976 original. MR0855460
[4] Bichteler, K., Gravereaux, J.-B. and Jacod, J. (1987). *Malliavin Calculus for Processes with Jumps*. *Stochastics Monographs* **2**. New York: Gordon and Breach Science Publishers. MR1008471
[5] Bingham, N.H. (1973). Maxima of sums of random variables and suprema of stable processes. *Z. Wahrsch. Verw. Gebiete* **26** 273–296. MR0415780 https://doi.org/10.1007/BF00534892
[6] Bouleau, N. and Denis, L. (2009). Energy image density property and the lent particle method for Poisson measures. *J. Funct. Anal.* **257** 1144–1174. MR2535466 https://doi.org/10.1016/j.jfa.2009.03.004
[7] Chaumont, L. (2013). On the law of the supremum of Lévy processes. *Ann. Probab.* **41** 1191–1217. MR3098676 https://doi.org/10.1214/11-AOP708
[8] Chaumont, L. and Małecki, J. (2016). On the asymptotic behavior of the density of the supremum of Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1178–1195. MR3531705 https://doi.org/10.1214/15-AIHP674
[9] Chaumont, L. and Małecki, J. (2021). Density behaviour related to Lévy processes. *Trans. Amer. Math. Soc.* **374** 1919–1945. MR4216728 https://doi.org/10.1090/tran/8268
[10] Darling, D.A. (1956). The maximum of sums of stable random variables. *Trans. Amer. Math. Soc.* **83** 164–169. MR0080393 https://doi.org/10.2307/1992908
[11] Dębicki, K. and Mandjes, M. (2015). *Queues and Lévy Fluctuation Theory*. *Universitext*. Cham: Springer. MR3379923 https://doi.org/10.1007/978-3-319-20693-6
[12] Doney, R.A. (2008). A note on the supremum of a stable process. *Stochastics* **80** 151–155. MR2402160 https://doi.org/10.1080/17442500701830399

[13] Doney, R.A. and Savov, M.S. (2010). The asymptotic behavior of densities related to the supremum of a stable process. *Ann. Probab.* **38** 316–326. MR2599201 https://doi.org/10.1214/09-AOP479
[14] Giles, M. (2023). Multilevel Monte Carlo research.
[15] Giles, M.B. (2008). Multilevel Monte Carlo path simulation. *Oper. Res.* **56** 607–617. MR2436856 https://doi.org/10.1287/opre.1070.0496
[16] González Cázares, J., Kohatsu-Higa, A. and Mijatović, A. (2023). Supplement to "Joint density of the stable process and its supremum: Regularity and upper bounds." https://doi.org/10.3150/23-BEJ1590SUPP
[17] González Cázares, J.I. and Mijatović, A. (2022). Convex minorants and the fluctuation theory of Lévy processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **19** 983–999. MR4448688 https://doi.org/10.30757/alea.v19-39
[18] González Cázares, J.I., Mijatović, A. and Uribe Bravo, G. (2019). Exact simulation of the extrema of stable processes. *Adv. in Appl. Probab.* **51** 967–993. MR4032169 https://doi.org/10.1017/apr.2019.39
[19] González Cázares, J.I., Mijatović, A. and Uribe Bravo, G. (2022). Geometrically convergent simulation of the extrema of Lévy processes. *Math. Oper. Res.* **47** 1141–1168. MR4435010 https://doi.org/10.1287/moor.2021.1163
[20] González Cázares, J.I., Kohatsu Higa, A. and Mijatović, A. (2022). Presentation on "Joint density of the stable process and its supremum: Regularity and upper bounds". https://youtu.be/x0n3Up9CxCA. YouTube video.
[21] Heyde, C.C. (1969). On the maximum of sums of random variables and the supremum functional for stable processes. *J. Appl. Probab.* **6** 419–429. MR0251766 https://doi.org/10.1017/s0021900200032927
[22] Kuznetsov, A. (2011). On extrema of stable processes. *Ann. Probab.* **39** 1027–1060. MR2789582 https://doi.org/10.1214/10-AOP577
[23] Kuznetsov, A. (2013). On the density of the supremum of a stable process. *Stochastic Process. Appl.* **123** 986–1003. MR3005012 https://doi.org/10.1016/j.spa.2012.11.001
[24] Kuznetsov, A. and Kwaśnicki, M. (2018). Spectral analysis of stable processes on the positive half-line. *Electron. J. Probab.* **23** Paper No. 10, 29. MR3771747 https://doi.org/10.1214/18-EJP134
[25] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. *Probability and Its Applications (New York)*. Berlin: Springer. MR2200233
[26] Nualart, D. and Nualart, E. (2018). *Introduction to Malliavin Calculus*. *Institute of Mathematical Statistics Textbooks* **9**. Cambridge: Cambridge Univ. Press. MR3838464 https://doi.org/10.1017/9781139856485
[27] Patie, P. and Savov, M. (2018). Bernstein-gamma functions and exponential functionals of Lévy processes. *Electron. J. Probab.* **23** Paper No. 75, 101. MR3835481 https://doi.org/10.1214/18-EJP202
[28] Pitman, J. and Uribe Bravo, G. (2012). The convex minorant of a Lévy process. *Ann. Probab.* **40** 1636–1674. MR2978134 https://doi.org/10.1214/11-AOP658
[29] 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
[30] Savov, M. (2020). private communication.
[31] Weron, R. (1996). On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statist. Probab. Lett.* **28** 165–171. MR1394670 https://doi.org/10.1016/0167-7152(95)00113-1
[32] Zolotarev, V.M. (1986). *One-Dimensional Stable Distributions*. *Translations of Mathematical Monographs* **65**. Providence, RI: Amer. Math. Soc. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver. MR0854867 https://doi.org/10.1090/mmono/065