BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty Two Number One February 2026 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty Two Number One February 2026 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 Volume 32 Number 1 February 2026 Pages 1–847 ISI/BS

Volume 32 Number 1 February 2026 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 2026 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **32**(1), 2026, 1–4
https://doi.org/10.3150/25-BEJ1910

Ole Eiler Barndorff-Nielsen

1350-7265

Bernoulli **32**(1), 2026, 5–23
https://doi.org/10.3150/25-BEJ1896

On Ole E. Barndorff-Nielsen's contributions to classical and free infinite divisibility and Lévy processes

VICTOR PÉREZ-ABREU[1,a] and STEEN THORBJØRNSEN[2,b]

[1]*Universidad Virtual del Estado de Guanajuato, Guanajuato, Mexico,* [a]*pabreuv@gmail.com*
[2]*Department of Mathematics, Aarhus University, Aarhus, Denmark,* [b]*steenth@math.au.dk*

This article reviews Ole E. Barndorff-Nielsen's contributions and perspectives on theoretical aspects of classical infinite divisibility and its interrelation with free infinite divisibility. Emphasis is placed on his understanding of such a relation and how this perspective was fertile for finding new concrete examples and families of infinitely divisible distributions in both senses.

Keywords: O.E. Barndorff-Nielsen; infinite divisibility; Lévy processes; free probability

References

Arizmendi, O., Barndorff-Nielsen, O.E. and Pérez-Abreu, V. (2010). On free and classical type G distributions. *Braz. J. Probab. Stat.* **24** 106–127. MR2643561 https://doi.org/10.1214/09-BJPS039

Arizmendi, O. and Hasebe, T. (2013). On a class of explicit Cauchy-Stieltjes transforms related to monotone stable and free Poisson laws. *Bernoulli* **19** 2750–2767. MR3160570 https://doi.org/10.3150/12-BEJ473

Arizmendi, O. and Pérez-Abreu, V. (2009). The *S*-transform of symmetric probability measures with unbounded supports. *Proc. Amer. Math. Soc.* **137** 3057–3066. MR2506464 https://doi.org/10.1090/S0002-9939-09-09841-4

Barndorff-Nielsen, O.E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2** 41–68. MR1804664 https://doi.org/10.1007/s007800050032

Barndorff-Nielsen, O.E., Gill, R. and Jupp, P. (2003). On quantum statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 775–816. MR2017871 https://doi.org/10.1111/1467-9868.00415

Barndorff-Nielsen, O.E. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrsch. Verw. Gebiete* **38** 309–311. MR0436260 https://doi.org/10.1007/BF00533162

Barndorff-Nielsen, O.E., Maejima, M. and Sato, K.-I. (2004). Some classes of multivariate infinitely divisible distributions admitting stochastic integral representation. *Bernoulli* **12** 1–33. MR2202318

Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S. (2001). *Lévy Processes, Theory and Applications*. Boston: Birkhäuser. MR1833689 https://doi.org/10.1007/978-1-4612-0197-7

Barndorff-Nielsen, O.E., Pérez-Abreu, V. and Thorbjørnsen, S. (2013). Lévy mixing. *ALEA Lat. Amer. J. Probab. Math. Stat.* **10** 1013–1062. MR3151748

Barndorff-Nielsen, O.E., Rosiński, J. and Thorbjørnsen, S. (2008). General Υ-transformations. *ALEA Lat. Amer. J. Probab. Math. Stat.* **4** 131–165. MR2421179

Barndorff-Nielsen, O.E. and Thorbjørnsen, S. (2002a). Selfdecomposability and Lévy processes in free probability. *Bernoulli* **8** 323–366. MR1913111

Barndorff-Nielsen, O.E. and Thorbjørnsen, S. (2002b). Lévy laws in free probability. *Proc. Natl. Acad. Sci. USA* **99** 16568–16575. MR1947756 https://doi.org/10.1073/pnas.232052399

Barndorff-Nielsen, O.E. and Thorbjørnsen, S. (2005). The Lévy-Itô decomposition in free probability. *Probab. Theory Related Fields* **131** 197–228. MR2117952 https://doi.org/10.1007/s00440-003-0322-y

Barndorff-Nielsen, O.E. and Thorbjørnsen, S. (2006). Regularising mappings of Lévy measures. *Stochastic Process. Appl.* **116** 423–446. MR2199557 https://doi.org/10.1016/j.spa.2005.09.011

Barndorff-Nielsen, O.E. and Thorbjørnsen, S. (2007). A connection between free and classical infinite divisibility. *Infin. Dimens. Anal. Quantum Probab.* **7** 573–590. MR2105912 https://doi.org/10.1142/S0219025704001773

Belinschi, S.T., Bożejko, M., Lehner, F. and Speicher, R. (2011). The normal distribution is ⊞-infinitely divisible. *Adv. Math.* **226** 3677–3698. MR2764902 https://doi.org/10.1016/j.aim.2010.10.025

Benth, F.E. and Veraart, A.E.D. (n.d.). Research frontiers in ambit stochastics: In memory of Ole E. Barndorff-Nielsen. In this volume. MR3839270 https://doi.org/10.1007/978-3-319-94129-5

Bercovici, H. and Pata, V. (1999). Stable laws and domains of attraction in free probability theory. *Ann. Math.* **149** 1023–1060. MR1709310 https://doi.org/10.2307/121080

Bercovici, H. and Pata, V. (2000). A free analogue of Hincin's characterization of infinite divisibility. *Proc. Amer. Math. Soc.* **128** 1011–1015. MR1636930 https://doi.org/10.1090/S0002-9939-99-05087-X

Bercovici, H. and Voiculescu, D.V. (1993). Free convolution of measures with unbounded support. *Indiana Univ. Math. J.* **42** 733–773. MR1254116 https://doi.org/10.1512/iumj.1993.42.42033

Bercovici, H. and Voiculescu, D.V. (1995). Superconvergence to the central limit and failure of the Cramér theorem for free random variables. *Probab. Theory Related Fields* **103** 215–222. MR1355057 https://doi.org/10.1007/BF01204215

Biane, P. (1998). Processes with free increments. *Math. Z.* **227** 143–174. MR1605393 https://doi.org/10.1007/PL00004363

Bożejko, M. and Speicher, R. (1991). An example of a generalized Brownian motion. *Comm. Math. Phys.* **137** 519–531. MR1105428

Chistyakov, G.P. and Götze, F. (2008). Limit theorems in free probability theory I. *Ann. Probab.* **36** 54–90. MR2370598 https://doi.org/10.1214/009117907000000051

Collet, F., Leisen, F. and Thorbjørnsen, S. (2021). Completely random measures and Lévy bases in free probability. *Electron. J. Probab.* **26**. MR4247974 https://doi.org/10.1214/21-ejp620

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* **II**. Wiley & Sons. MR0270403

Gnedenko, B.V. and Kolmogorov, A.N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley Publishing Company, Inc. MR0233400

Haagerup, U. and Thorbjørnsen, S. (2014). On the free gamma distributions. *Indiana Univ. Math. J.* **63** 1159–1194. MR3263926 https://doi.org/10.1512/iumj.2014.63.5288

Halgreen, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrsch. Verw. Gebiete* **47** 13–17. MR0521527 https://doi.org/10.1007/BF00533246

Hasebe, T. (2014). Free infinite divisibility for beta distributions and related ones. *Electron. J. Probab.* **19** 1–33. MR3256881 https://doi.org/10.1214/EJP.v19-3448

Hasebe, T. and Sakuma, N. (2017). Unimodality for free Lévy processes. *Ann. Inst. Henri Poincaré* **53** 916–936. MR3634280 https://doi.org/10.1214/16-AIHP742

Hasebe, T., Sakuma, N. and Thorbjørnsen, S. (2019). The normal distribution is freely selfdecomposable. *Int. Math. Res. Not.* **2019** 1758–1787. MR3932595 https://doi.org/10.1093/imrn/rnx171

Hasebe, T. and Thorbjørnsen, S. (2016). Unimodality of the freely selfdecomposable probability laws. *J. Theoret. Probab.* **29** 922–940. MR3540484 https://doi.org/10.1007/s10959-015-0595-y

Itô, K. (2004). *Stochastic Processes: Lectures Given at Aarhus University*. Berlin: Springer. Edited and with a foreword by Ole E. Barndorff-Nielsen and Ken-iti Sato. MR2053326 https://doi.org/10.1007/978-3-662-10065-3

Jurek, Z.J. (1985). Relations between the s-self-decomposable and self-decomposable measures. *Ann. Probab.* **13** 592–608. MR0781426

Jurek, Z.J. (1990). On Lévy (spectral) measures of integral form on Banach spaces. *Probab. Math. Statist.* **11** 139–148. MR1096945

Jurek, Z.J. and Mason, J.D. (1993). *Operator-Limit Distributions in Probability Theory*. Wiley. MR1243181

Jurek, Z.J. and Verwaat, W. (1983). An integral representation for selfdecomposable Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **62** 247–262. MR0688989 https://doi.org/10.1007/BF00538800

Lukacs, E. (1975). *Stochastic Convergence*, 2nd ed. Academic Press. MR0375405

Maassen, H. (1992). Addition of freely independent random variables. *J. Funct. Anal.* **106** 409–438. MR1165862 https://doi.org/10.1016/0022-1236(92)90055-N

Maejima, M. (2015). Classes of infinitely divisible distributions and examples. In *Lévy Matters V. Functionals of Lévy Processes* 1–65. MR3467344 https://doi.org/10.1007/978-3-319-23138-9_1

Nica, A. (1995). A one-parameter family of transforms, linearizing convolution laws for probability distributions. *Comm. Math. Phys.* **168** 187–207. MR1324395

Pérez-Abreu, V. and Sakuma, N. (2008). Free generalized gamma convolutions. *Electron. Commun. Probab.* **13** 526–539. MR2447839 https://doi.org/10.1214/ECP.v13-1413

Pérez-Abreu, V. and Sakuma, N. (2012). Free infinite divisibility of free multiplicative mixtures of the Wigner distribution. *J. Theor. Probab.* **25** 100–121. MR2886381 https://doi.org/10.1007/s10959-010-0288-5

Rajput, B.S. and Rosiński, J. (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **112** 451–487. MR1001524 https://doi.org/10.1007/BF00339998

Reid, N. and Battey, H.S. (n.d.), The Magical Mystery Formula. In this volume.

Sakuma, N. (2009). Characterizations of the class of free self-decomposable distributions and its subclasses. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **12** 51–65. MR2509984 https://doi.org/10.1142/S0219025709003525

Sato, K. (2006a). Two families of improper stochastic integrals with respect to Lévy processes. *ALEA Lat. Amer. J. Probab. Math. Stat.* **1** 47–87. MR2235174

Sato, K. (2006b). Additive processes and stochastic integrals. *Illinois J. Math.* **50** 825–851. MR2247848

Sato, K. (2007). Transformations of infinitely divisible distributions via improper stochastic integrals. *ALEA Lat. Amer. J. Probab. Math. Stat.* **3** 67–110. MR2349803

Shephard, N. (n.d.). Ole Eiler Barndorff-Nielsen and financial econometrics. In this volume.

Sørensen, M. (n.d.). Ole Barndorff-Nielsen: Sand, Wind and inference. In this volume.

Voiculescu, D.V. (1985). Symmetries of some reduced free product C*-algebras. In *Operator Algebras and Their Connections with Topology and Ergodic Theory*. *Springer Lecture Notes in Mathematics* **1132** 556–588. MR0799593 https://doi.org/10.1007/BFb0074909

Voiculescu, D.V. (1986). Addition of certain non-commuting random variables. *J. Funct. Anal.* **66** 323–346. MR0839105 https://doi.org/10.1016/0022-1236(86)90062-5

Voiculescu, D.V. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

Wang, J. (2010). Local limit theorems in free probability theory. *Ann. Probab.* **38** 1492–1506. MR2663634 https://doi.org/10.1214/09-AOP505

Williams, J. (2012). Uniform convergence and the free central limit theorem. *Complex Anal. Oper. Theory* **6** 23–31. MR4456097

Wolfe, S.J. (1982). On a continuous analogue of the stochastic difference equation $X_n = \rho X_{n-1} + B_n$. *Stochastic Process. Appl.* **12** 301–312. MR0656279 https://doi.org/10.1016/0304-4149(82)90050-3

Yamazato, Y. (1978). Unimodality of infinitely divisible distribution functions of class L. *Ann. Probab.* **6** 523–531. MR0482941

Bernoulli **32**(1), 2026, 24–48
https://doi.org/10.3150/25-BEJ1902

Research frontiers in ambit stochastics: In memory of Ole E. Barndorff-Nielsen

FRED ESPEN BENTH[1,a] and ALMUT E.D. VERAART[2,b]

[1] *BI Norwegian Business School, Department of Data Science and Analytics, N-0484 Oslo, Norway,*
[a] *fred.e.benth@bi.no*
[2] *Department of Mathematics, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK,*
[b] *a.veraart@imperial.ac.uk*

This article surveys key aspects of ambit stochastics and remembers Ole E. Barndorff-Nielsen's important contributions to the foundation and advancement of this new research field over the last two decades. It also highlights some of the emerging trends in ambit stochastics.

Keywords: Ambit fields; trawl processes; Lévy semistationary processes; stochastic volatility; limit theory; turbulence; stochastic growth model

References

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton, New Jersey: Princeton University Press.

Applebaum, D. (2015). Infinite dimensional Ornstein-Uhlenbeck processes driven by Lévy processes. *Probab. Surv.* **12** 33–54.

Applebaum, D. and Riedle, M. (2010). Cylindrical Lévy processes in Banach spaces. *Proc. Lond. Math. Soc.* **101** 697–726.

Barndorff-Nielsen, O.E., Bennedsen, M. and Schmiegel, J. (2022a). Universal statistical description of the main velocity component in a turbulent boundary layer. Work in progress.

Barndorff-Nielsen, O.E., Bennedsen, M. and Schmiegel, J. (2022b). A tempo-spatial Brownian Semistationary model of the main velocity component in a turbulent boundary layer. Work in progress.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2011). Ambit processes and stochastic partial differential equations. In *Advanced Mathematical Methods for Finance* (G. Di Nunno and B. Øksendal, eds.) 35–74. Springer.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2013). Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli* **19** 803–845.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2015). Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. *Banach Center Publ.* **104** 25–60.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2018). Ambit stochastics. In *Probability Theory and Stochastic Modelling* **88**. Cham: Springer. https://doi.org/10.1007/978-3-319-94129-5

Barndorff-Nielsen, O.E., Blæsild, P. and Schmiegel, J. (2004). A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41** 345–363.

Barndorff-Nielsen, O.E., Corcuera, J.M. and Podolskij, M. (2011). Multipower variation for Brownian semistationary processes. *Bernoulli* **17** 1159–1194.

Barndorff-Nielsen, O.E., Corcuera, J.M. and Podolskij, M. (2013). Limit theorems for functionals of higher order differences of Brownian semistationary processes. In *Prokhorov and Contemporary Probability* (A.E. Shiryaev, S.R.S. Varadhan and E. Presman, eds.). *Springer Proceedings in Mathematics and Statistics* **33** 69–96.

Barndorff-Nielsen, O.E. and Graversen, S.E. (2011). Volatility determination in an ambit process setting. *J. Appl. Probab.* **48A** 263–275.

Barndorff-Nielsen, O.E., Pakkanen, M. and Schmiegel, J. (2014). Assessing relative volatility/intermittency/energy dissipation. *Electron. J. Stat.* **8** 1996–2021.

Barndorff-Nielsen, O.E. and Pedersen, J. (2012). Meta-times and extended subordination. *Theory Probab. Appl.* **56** 319–327. https://doi.org/10.1137/S0040585X97985467

Barndorff-Nielsen, O.E., Pedersen, J. and Sato, K.-I. (2001). Multivariate subordination, self-decomposability and stability. *Adv. Appl. Probab.* **33** 160–187. https://doi.org/10.1017/S0001867800010685

Barndorff-Nielsen, O.E. and Pérez-Abreu, V. (2008). Matrix subordinators and related upsilon transformations. *Theory Probab. Appl.* **52** 1–23. https://doi.org/10.1137/S0040585X97982839

Barndorff-Nielsen, O.E. and Schmiegel, J. (2004). Lévy-based tempo-spatial modelling; with applications to turbulence. *Uspekhi Mat. Nauk* **59** 63–90.

Barndorff-Nielsen, O.E. and Schmiegel, J. (2007). Ambit processes: With applications to turbulence and tumour growth. In *Stochastic Analysis and Applications: The Abel Symposium 2005* (F.E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal and T. Zhang, eds.) 93–124. Heidelberg: Springer.

Barndorff-Nielsen, O.E. and Schmiegel, J. (2009). Brownian semistationary processes and volatility/intermittency. In *Advanced Financial Modelling* (H. Albrecher, W. Rungaldier and W. Schachermeyer, eds.). *Radon Series on Computational and Applied Mathematics* **8** 1–26. Berlin: W. de Gruyter.

Barndorff-Nielsen, O.E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 167–241.

Barndorff-Nielsen, O.E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280.

Barndorff-Nielsen, O.E. and Shephard, N. (2003). Realized power variation and stochastic volatility models. *Bernoulli* **9** 243–265.

Barndorff-Nielsen, O.E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *J. Financ. Econom.* **2** 1–37.

Barndorff-Nielsen, O.E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl.* **116** 796–806.

Barndorff-Nielsen, O.E. and Shiryaev, A. (2015). *Change of Time and Change of Measure*, second ed. ed. *Advanced Series on Statistical Science and Applied Probability* **21**. World Scientific. https://doi.org/10.1142/9609

Barndorff-Nielsen, O.E. and Stelzer, R. (2011). Multivariate supOU processes. *Ann. Appl. Probab.* **21** 140–182.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M. and Shephard, N. (2006a). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From Stochastic Calculus to Mathematical Finance* 33–68. Berlin: Springer.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J. and Shephard, N. (2006b). Limit theorems for bipower variation in financial econometrics. *Econometric Theory* **22** 677–719.

Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M. and Woerner, J.H.C. (2009). Bipower variation for Gaussian processes with stationary increments. *J. Appl. Probab.* **46** 132–150.

Barndorff-Nielsen, O.E., Hedevang, E., Schmiegel, J. and Szozda, B. (2016). Some recent developments in ambit stochastics. In *Stochastics of Environmental and Financial Economics*. *Springer Proceedings in Mathematics & Statistics* **138** 3–25. Netherlands: Springer. https://doi.org/10.1007/978-3-319-23425-0_1

Basse-O'Connor, A., Heinrich, C. and Podolskij, M. (2018). On limit theory for Lévy semi-stationary processes. *Bernoulli* **24** 3117–3146.

Basse-O'Connor, A., Lachièze-Rey, R. and Podolskij, M. (2017). Power variation for a class of stationary increments Lévy driven moving averages. *Ann. Probab.* **45** 4477–4528. https://doi.org/10.1214/16-AOP1170

Bennedsen, M. (2017). A rough multi-factor model of electricity spot prices. *Energy Econ.* **63** 301–313.

Bennedsen, M., Lunde, A. and Pakkanen, M.S. (2021). Decoupling the short- and long-term behavior of stochastic volatility. *J. Financ. Econom.* **20** 961–1006. https://doi.org/10.1093/jjfinec/nbaa049

Bennedsen, M., Lunde, A., Shephard, N. and Veraart, A.E.D. (2023). Inference and forecasting for continuous-time integer-valued trawl processes. *J. Econometrics* **236** 105476.

Benth, F.E., Detering, N. and Krühner, P. (2022). Stochastic Volterra integral equations and a class of first order stochastic partial differential equations. *Stochastics* **94** 1054–1076.

Benth, F.E. and Eyjolfsson, H. (2017). Representation and approximation of ambit fields in Hilbert space. *Stochastics* **89** 311–347.

Benth, F.E. and Krühner, P. (2023). *Stochastic Models for Prices Dynamics in Energy and Commodity Markets – an Infinite-Dimensional Perspective*. Cham: Springer Finance. Springer Verlag.

Benth, F.E. and Šaltytė Benth, J. (2009). Dynamic pricing of wind futures. *Energy Econ.* **31** 16–24.
Benth, F.E. and Šaltytė Benth, J. (2011). Weather derivatives and stochastic modelling of temperature. *Int. J. Stoch. Anal.* **21** 576791.
Benth, F.E. and Šaltytė Benth, J. (2012). Modeling and pricing in financial markets for weather derivatives. In *Advanced Series on Statistical Science and Applied Probability* **17**. World Scientific.
Benth, F.E., Schroers, D. and Veraart, A.E.D. (2022). A weak law of large numbers for realised covariation in a Hilbert space setting. *Stochastic Process. Appl.* **145** 241–268.
Benth, F.E., Schroers, D. and Veraart, A.E.D. (2024). A feasible central limit theorem for realised covariation of SPDEs in the context of functional data. *Ann. Appl. Probab.* **34** 2208–2242. https://doi.org/10.1214/23-AAP2019
Benth, F.E. and Simonsen, I. (2021). Metatimes, random measures and cylindrical random variables. *Mod. Stoch. Theory Appl.* **8** 349–371.
Benth, F.E. and Süss, A. (2018). Continuous-time autoregressive moving-average processes in Hilbert space. In *Computation and Combinatorics in Dynamics, Stochastics and Control* (E. Celledoni, G. Di Nunno, K. Ebrahimi-Fard and H.Z. Munthe-Kaas, eds.) 297–320. Cham: Springer International Publishing.
Birnir, B. (2013). *The Kolmogorov-Obukhov Theory of Turbulence*. New York: Springer.
Chong, C. (2017). Lévy-driven Volterra equations in space and time. *J. Theor. Probab.* **30** 1014–1058.
Ciołek, G., Marushkevych, D. and Podolskij, M. (2020). On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.* **14** 4395–4420. https://doi.org/10.1214/20-EJS1775
Ciolek, G., Marushkevych, D. and Podolskij, M. (2025). On Lasso estimator for the drift function in diffusion models. *Bernoulli* **31** 1811–1833.
Corcuera, J.-M., Farkas, G., Schoutens, W. and Valkeila, E. (2013b). A short rate model using ambit processes. In *Malliavin Calculus and Stochastic Analysis* (F. Viens, J. Feng, Y. Hu and E. Nualart, eds.). *Springer Proceedings in Mathematics & Statistics* **34** 525–553. Springer US.
Corcuera, J.M., Hedevang, E., Pakkanen, M.S. and Podolskij, M. (2013a). Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stochastic Process. Appl.* **123** 2552–2574. A Special Issue on the Occasion of the 2013 International Year of Statistics.
Courgeau, V. and Veraart, A.E.D. (2022a). Likelihood theory for the graph Ornstein-Uhlenbeck process. *Stat. Inference Stoch. Process.* **25** 227–260.
Courgeau, V. and Veraart, A.E.D. (2022b). High-frequency estimation of the Lévy-driven graph Ornstein-Uhlenbeck process. *Electron. J. Stat.* **16** 4863–4925. https://doi.org/10.1214/22-ejs2052
Dunford, N. and Schwartz, J.T. (1958). *Linear Operators. Part I: General Theory*. New York: Interscience Publishers.
Frisch, U. (1995). *Turbulence – the Legacy of A. N. Kolmogorov*. Cambridge University Press.
Gärtner, K. and Podolskij, M. (2015). On non-standard limits of Brownian semi-stationary processes. *Stochastic Process. Appl.* **125** 653–677.
Granelli, A. and Veraart, A.E.D. (2019). A central limit theorem for the realised covariation of a bivariate Brownian semistationary process. *Bernoulli* **25** 2245–2278. https://doi.org/10.3150/18-BEJ1052
Hedevang, E. and Schmiegel, J. (2013). A causal continuous time stochastic model for the turbulent energy dissipation in a helium jet flow. *J. Turbul.* **14** 1–26.
Hedevang, E. and Schmiegel, J. (2014). A Lévy based approach to random vector fields: With a view towards turbulence. *Int. J. Nonlinear Sci. Numer. Simul.* **15** 411–435.
Heiny, J. and Podolskij, M. (2021). On estimation of quadratic variation for multivariate pure jump semimartingales. *Stochastic Process. Appl.* **138**. https://doi.org/10.1016/j.spa.2021.04.016
Holden, H., Øksendal, B., Ubøe, J. and Zhang, T. (2010). *Stochastic Partial Differential Equations – a Modeling, White Noise Functional Approach*, second edition ed. *Universitext*. New York: Springer-Verlag.
Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.* **118** 517–559.
Jacod, J. and Protter, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer.
Jónsdóttir, K.Y., Schmiegel, J. and Vedel Jensen, E.B. (2008). Lévy-based growth models. *Bernoulli* **14** 62–90.
Jónsdóttir, K.Y. and Vedel Jensen, E.B. (2005). Gaussian radial growth. *Image Anal. Stereol.* **24** 111–126.

Kolmogorov, A.N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds number. *Dokl. Akad. Nauk SSSR* **30** 9–13.

Leonte, D. (2023). Ambit-Stochastics Python library.

Lucchese, L., Pakkanen, M.S. and Veraart, A.E.D. (2025). Estimation and inference for multivariate continuous-time autoregressive processes. *Ann. Appl. Probab.* To appear.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *G. Ist. Ital. Attuari* **64** 19–47.

Mancini, C. (2004). Estimating the integrated volatility in stochastic volatility models with Lévy jumps Technical report, Università di Firenze.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.* **36** 270–296.

Mehta, S. and Veraart, A.E.D. (2025). Estimation and inference for the graph supOU process. arXiv preprint arXiv:2502.08838.

Nguyen, M. and Veraart, A.E.D. (2017). Spatio-temporal Ornstein-Uhlenbeck processes: Theory, simulation and statistical inference. *Scand. J. Stat.* **44** 46–80.

Norlyk, A. and Veraart, A.E.D. (2025). The Stochastic Block Ornstein-Uhlenbeck Process. Work in progress.

Nualart, D. and Peccati, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. https://doi.org/10.1214/009117904000000621

Obukhov, A.M. (1941). On the distribution of energy in the spectrum of turbulent flow. *Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz.* **5** 453–466.

Onsager, L. (1949). Statistical hydrodynamics. *Nuovo Cimento* **6** 279–287.

Pakkanen, M.S. (2014). Limit theorems for power variations of ambit fields driven by white noise. *Stochastic Process. Appl.* **124** 1942–1973.

Passeggeri, R. and Veraart, A.E.D. (2019). Limit theorems for multivariate Brownian semistationary processes and feasible results. *Adv. Appl. Probab.* **51** 667–716. https://doi.org/10.1017/apr.2019.30

Peszat, S. and Zabczyk, J. (2007). *Stochastic Partial Differential Equations with Lévy Noise*. *Encyclopedia of Mathematics and Its Applications* **113**. Cambridge: Cambridge University Press.

Podolskij, M. (2015). Ambit fields: Survey and new challenges. In *XI Symposium of Probability and Stochastic Processes: CIMAT, Mexico, November 18-22, 2013*, (R.H. Mena, J.C. Pardo, V. Rivero and G.U. Bravo, eds.). *Progress in Probability* **69** 241–279. Springer.

Protter, P.E. (2005). *Stochastic Integration and Differential Equations*, Second edition ed. *Stochastic Modelling and Applied Probability* **21**. Berlin: Springer-Verlag. Version 2.1, Corrected third printing.

Rényi, A. (1963). On stable sequences of events. *Sankhyā Ser. A* **25** 293–302.

Sauri, O. (2020). On the divergence and vorticity of vector ambit fields. *Stochastic Process. Appl.* **130** 6184–6225. https://doi.org/10.1016/j.spa.2020.05.007

Sauri, O. and Veraart, A.E.D. (2024). Nonparametric estimation of trawl processes: Theory and Applications. arXiv preprint arXiv:2209.05894.

Schmiegel, J. (2005). Self-scaling of turbulent energy dissipation correlators. *Phys. Lett. A* **337** 342–353.

Schmiegel, J. (2006). Self-scaling tumour growth. *Phys. A* **367** 509–524.

Schmiegel, J. (2018). Ambit fields: A stochastic modelling approach. *Math. Comput. Model. Dyn. Syst.* **24** 383–400. https://doi.org/10.1080/13873954.2018.1488738

Schmiegel, J., Barndorff-Nielsen, O.E. and Eggers, H.C. (2005). A class of spatio-temporal and causal stochastic processes with application to multiscaling and multifractality. *South Afr. J. Sci.* **101** 513–519.

Shephard, N. (2025). Ole Eiler Barndorff-Nielsen and financial econometrics. Bernoulli. To appear.

Taylor, G.I. (1938). The spectrum of turbulence. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **164** 476–490.

Veraart, A.E.D. (2021). trawl: Estimation and Simulation of Trawl Processes R package version 0.2.2.

Veraart, A.E.D. (2022). ambit: Simulation and Estimation of Ambit Processes R package version 0.1.2.

Walsh, J.B. (1986). An introduction to stochastic partial differential equations. In *Ecole d'Eté de Probabilités de Saint-Flour XIV* (R. Carmona, H. Kesten and J.B. Walsh, eds.). *Lecture Notes in Mathematics* **1180**. Springer.

Bernoulli **32**(1), 2026, 49–67
https://doi.org/10.3150/25-BEJ1903

Ole E. Barndorff-Nielsen: Sand, wind and inference

MICHAEL SØRENSEN[a]

Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, [a]*michael@math.ku.dk*

This paper reviews Ole Eiler Barndorff-Nielsen's research in the first decades of his career. The focus is on topics that he kept returning to throughout his scientific life, and on papers that he built on in later important contributions. First his early contributions to the foundations of statistical inference are reviewed with focus on conditional inference and exponential families, two topics in which he had a lifelong interest. The second half of the paper reviews his research on wind blown sand and hyperbolic distributions and processes, including his early contributions to modelling of turbulent wind fields. This research laid the foundations for his later work on financial econometrics and ambit processes.

Keywords: Aeolian sand transport; conditional inference; differential geometry and inference; exponential families; hyperbolic distributions; hyperbolic processes; turbulence

References

Anderson, R.S., Sørensen, M. and Willetts, B.B. (1991). A review of recent progress in our understanding of aeolian sediment transport. *Acta Mech., Suppl.* **1** 1–9.

Bagnold, R.A. (1941). *The Physics of Blown Sand and Desert Dunes*. London: Chapman and Hall.

Bagnold, R.A. and Barndorff-Nielsen, O.E. (1980). The pattern of natural size distributions. *Sedimentology* **27** 199–207.

Barnard, G.A. (1963). Some logical aspects of the fiducial argument. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **25** 111–114. MR0172436

Barndorff-Nielsen, O.E. (1973). On M-ancillarity. *Biometrika* **60** 447–455. MR0345255 https://doi.org/10.1093/biomet/60.3.447

Barndorff-Nielsen, O.E. (1976a). Nonformation. *Biometrika* **63** 567–571. MR0494594 https://doi.org/10.1093/biomet/63.3.567

Barndorff-Nielsen, O.E. (1976b). Plausibility inference (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **38** 103–131. MR0428525

Barndorff-Nielsen, O.E. (1976c). Factorization of likelihood functions for full exponential families. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **38** 37–44. MR0438539

Barndorff-Nielsen, O.E. (1977). Exponentially decreasing log-size distributions. *Proc. R. Soc. Lond. A* **353** 401–419.

Barndorff-Nielsen, O.E. (1978a). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley. MR0489333

Barndorff-Nielsen, O.E. (1978b). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Stat.* **5** 151–157. MR0509451

Barndorff-Nielsen, O.E. (1979). Models for non-Gaussian variation with applications to turbulence. *Proc. R. Soc. Lond. A* **368** 501–520. MR0554672 https://doi.org/10.1098/rspa.1979.0144

Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika* **67** 293–310. MR0581727 https://doi.org/10.1093/biomet/67.2.293

Barndorff-Nielsen, O.E. (1982). Hyperbolic distributions in statistical physics. *Scand. J. Stat.* **9** 43–46. MR0651860

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

Barndorff-Nielsen, O.E. (1985). Differential and integral geometry in statistical inference. In *Differential Geometry in Statistical Inference 95–161*. Hayward, California: IMS.

Barndorff-Nielsen, O.E. (1986). Likelihood and observed geometries. *Ann. Statist.* **14** 856–873. MR0856794 https://doi.org/10.1214/aos/1176350038

Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **24** 1–13. MR1436619 https://doi.org/10.1111/1467-9469.00045

Barndorff-Nielsen, O.E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2** 41–68. MR1804664 https://doi.org/10.1007/s007800050032

Barndorff-Nielsen, O.E. (1999). L-nonformation, L-ancillarity and L-sufficiency. *Theory Probab. Appl.* **44** 225–229. MR1751203 https://doi.org/10.1137/S0040585X97977495

Barndorff-Nielsen, O.E. (2001). Superposition of Ornstein-Uhlenbeck type processes. *Theory Probab. Appl.* **45** 175–194. MR1967758 https://doi.org/10.1137/S0040585X97978166

Barndorff-Nielsen, O.E. (2018). *Stochastic in Science – Some Autobiographical Notes*. Højbjerg: Forlaget Morgenstjerne.

Barndorff-Nielsen, O.E. and Blæsild, P. (1975). S-ancillarity in exponential families. *Sankhyā Ser. A* **37** 354–385. MR0461732

Barndorff-Nielsen, O.E. and Blæsild, P. (1983a). Reproductive exponential families. *Ann. Statist.* **11** 770–782. MR0707928 https://doi.org/10.1214/aos/1176346244

Barndorff-Nielsen, O.E. and Blæsild, P. (1983b). Hyperbolic distributions. In *Encyclopedia of Statistical Science, 3* (S. Kotz and N.S. Johnson, eds.) 700–707. Wiley: New York. MR0651860

Barndorff-Nielsen, O.E. and Blæsild, P. (1987). Strings: Mathematical theory and statistical examples. *Proc. R. Soc. Lond. A* **411** 155–176. MR0895355

Barndorff-Nielsen, O.E., Blæsild, P. and Schmiegel, J. (2004). A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41** 345–363.

Barndorff-Nielsen, O.E. and Christiansen, C. (1988). Erosion, deposition and size distributions of sand. *Proc. R. Soc. Lond. A* **417** 335–352. MR0952340

Barndorff-Nielsen, O.E. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrsch. Verw. Gebiete* **38** 309–311. MR0436260 https://doi.org/10.1007/BF00533162

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1982). On the relation between size and distance travelled for winddriven grains – results and discussion of a pilot experiment using coloured sand. In *Proceedings of Euromech 156, Mechanics of Sediment Transport* (B.M. Sumer and A. Müller, eds.) 55–64. Balkema: Amsterdam.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1989). Wind shear and hyperbolic distributions. *Bound.-Layer Meteorol.* **46** 417–431.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1990). Parametric modelling of turbulence. *Philos. Trans. R. Soc. Lond. A* **332** 439–455. MR0345255 https://doi.org/10.1093/biomet/60.3.447

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1993). A statistical model for the streamwise component of a turbulent velocity field. *Ann. Geophys.* **11** 99–103.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1998). Some stationary processes in discrete and continuous time. *Adv. Appl. Probab.* **30** 499–424. MR1671092 https://doi.org/10.1239/aap/1035228204

Barndorff-Nielsen, O.E. and Jupp, P.E. (1988). Differential geometry, profile likelihood, L-sufficiency and composite transformation models. *Ann. Statist.* **16** 1009–1043. MR0959187 https://doi.org/10.1214/aos/1176350946

Barndorff-Nielsen, O.E., Kent, J. and Sørensen, M. (1982). Normal variance-mean mixtures and z-distributions. *Int. Stat. Rev.* **50** 145–159. MR0678296 https://doi.org/10.2307/1402598

Barndorff-Nielsen, O.E. and Koudou, A.E. (1995). Cuts in natural exponential families. *Theory Probab. Appl.* **40** 361–372. MR1346472 https://doi.org/10.1137/1140025

Barndorff-Nielsen, O.E. and Schmiegel, J. (2004). Lévy-based spatial-temporal modelling, with application to turbulence. *Uspekhi Mat. Nauk* **59** 63–90.

Barndorff-Nielsen, O.E. and Schmiegel, J. (2009). Brownian semistationary processes and volatility/intermittency. In *Advanced Financial Modelling* (H. Albrecher, W. Rungaldier and W. Schachermayer, eds.). *Radon Series Comp. Appl. Math* **8** 1–25. MR2648456 https://doi.org/10.1515/9783110213140.1

Barndorff-Nielsen, O.E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. Roy. Statist. Soc. Ser. B, Methodol.* **63** 167–241. MR1841412 https://doi.org/10.1111/1467-9868.00282

Barndorff-Nielsen, O.E. and Sørensen, M. (1991). On the temporal-spatial variation of sediment size distributions. *Acta Mech., Suppl.* **2** 23–35.

Barndorff-Nielsen, O.E., Blæsild, P., Jensen, J.L. and Jørgensen, B. (1982a). Exponential transformation models. *Proc. R. Soc. Lond. A* **379** 41–65. MR0643215 https://doi.org/10.1098/rspa.1982.0004

Barndorff-Nielsen, O.E., Dalsgaard, K., Halgreen, C., Kuhlmann, H., Møller, J.T. and Schou, G. (1982b). Variation in particle size distribution over a small dune. *Sedimentology* **29** 53–65.

Barndorff-Nielsen, O.E., Blæsild, P., Jensen, J.L. and Sørensen, M. (1985a). The fascination of sand. In *A Celebration of Statistics – the ISI Centenary Volume* (A.C. Atkinson and S.E. Fienberg, eds.) 57–87. New York: Springer.

Barndorff-Nielsen, O.E., Jensen, J.L., Nielsen, H.L., Rasmussen, K.R. and Sørensen, M. (1985b). Wind tunnel tracer studies of grain progress. In *Proceedings of International Workshop on the Physics of Blown Sand, Memoirs No. 8* (O.E. Barndorff-Nielsen, J.T. Møller, K.R. Rasmussen and B.B. Willetts, eds.) 243–251. Dept. Theor. Statist., Aarhus University.

Barndorff-Nielsen, O.E., Eggers, H.C., Greiner, M. and Schmiegel, J. (2005). A class of spatio-temporal and causal processes with application to multiscaling and multifractality. *South Afr. J. Sci.* **101** 1–7. MR2397785 https://doi.org/10.1007/978-3-540-70847-6\protect \relax \T1\textunderscore 5

Battey, H. and Reid, N. (2025). The Magical Mystery Formula. *Bernoulli* **32** 81–95. This issue.

Benth, F.E. and Veraart, A.E.D. (2025). Research frontiers in ambit stochastics: In memory of O. E. Barndorff-Nielsen. *Bernoulli* **32** 24–48. This issue.

Bibby, B.M. and Sørensen, M. (2003). Hyperbolic processes in finance. In *Handbook of Heavy Tailed Distributions in Finance* (S.T. Rachev, ed.) 211–248. Amsterdam: Elsevier.

Birnir, B. (2013). The Kolmogorov-Obukhov theory of turbulence. *J. Nonlinear Sci.* **23** 657–688. MR3079672 https://doi.org/10.1007/s00332-012-9164-z

Fraser, D.A.S. (1956). Sufficient statistics with nuisance parameters. *Ann. Math. Statist.* **27** 838–842. MR0080411 https://doi.org/10.1214/aoms/1177728191

Good, I.J. (1953). The population frequences of species and the estimation of population parameters. *Biometrika* **40** 237–264. MR0061330 https://doi.org/10.1093/biomet/40.3-4.237

Halgreen, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrsch. verw. Gebiete* **47** 13–17. MR0521527 https://doi.org/10.1007/BF00533246

Jensen, J.L. (1981). On the hyperboloid distribution. *Scand. J. Stat.* **8** 193–206. MR0642800

Madan, D.B. and Seneta, E. (1990). The Variance Gamma (V.G.) model for share market returns. *J. Bus.* **63** 511–524.

Møller, C. (1968). Gibbs' statistical mechanics in the theory of relativity. *Kgl. Dan. Vid. Selsk., Mat.-Fys. Medd.* **36** 1–45.

Monin, A.S. and Yaglom, A.M. (1971). *Statistical Fluid Mechanics: Mechanics of Turbulence* **1**. Cambridge, Massachusetts: MIT Press. MR2406667

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32. MR0025113 https://doi.org/10.2307/1914288

Pérez-Abreu, V. and Thorbjørnsen, S. (2025). On Ole E. Barndorff-Nielsen's contributions to classical and free infinite divisibility and Lévy processes. *Bernoulli* **32** 5–23. This issue.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Pædagogiske Institut.

Rémon, M. (1984). On a concept of partial sufficiency: L-sufficiency. *Int. Stat. Rev.* **52** 127–135. MR0967206 https://doi.org/10.2307/1403095

Rydberg, T. (1997). The normal inverse Gaussian Lévy process: Simulation and approximation. *Commun. Stat., Stoch. Models* **13** 887–910. MR1482297 https://doi.org/10.1080/15326349708807456

Sandved, E. (1967). A principle for conditioning on an ancillary. *Skand. Aktuarietidskr.* **50** 39–47. MR0224210 https://doi.org/10.1080/03461238.1966.10405695

Seshadri, V. (1997). Halphen's laws. In *In Encyclopedia of Statistical Sciences, Update* (S. Kotz, C.B. Read and D.L. Banks, eds.) **1** 302–306. New York: Wiley. MR1605063

Shephard, N. (2025). Ole Eiler Barndorff-Nielsen and financial econometrics. *Bernoulli* **32** 68–80. This issue.

Sichel, H.S. (1973). Statistical evaluation of diamondiferous deposits. *J. S. Afr. Inst. Min. Metall.* **76** 235–243.

Sørensen, M. (2004). On the rate of aeolian sand transport. *Geomorphology* **59** 53–62.

Sørensen, M. (2016). On the size distribution of sand. In *The Fascination of Probability, Statistics and Their Applications* (M. Podolskij, R. Stelzer, S. Thorbjørnsen and A.E.D. Veraart, eds.) 1–13. Berlin: Springer. MR3495677

Sverdrup, E. (1966). The present state of the decision theory and the Neyman-Pearson theory. *Rev. Inst. Int. Stat.* **34** 309–333. MR0207140

Bernoulli **32**(1), 2026, 68–80
https://doi.org/10.3150/25-BEJ1905

Ole Eiler Barndorff-Nielsen and financial econometrics

NEIL SHEPHARD[a]

Department of Economics and Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A., [a]*shephard@fas.harvard.edu*

This note reviews some of the contributions Ole Eiler Barndorff-Nielsen made to financial econometrics. He was particularly active in that area from the mid-nineties for around a dozen years. His innovations include the NIG Lévy process, the Barndorff-Nielsen-Shephard model, the supOU process, the formalization of realized volatility and realised beta, the introduction of bipower variation, realized semivariance, realized kernels and gradual jumps.

Keywords: High frequency financial econometrics; jumps; in-fill asymptotics; Lévy process; stochastic volatility

References

Ait-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econometrics* **201** 384–399.

Andersen, T.G. and Benzoni, L. (2009). Realized volatility. In *Handbook of Financial Time Series* (T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch, eds.) 555–575. Springer-Verlag.

Andersen, T.G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *J. Empir. Finance* **4** 115–158.

Andersen, T.G. and Bollerslev, T. (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.* **39** 885–905.

Andersen, T.G. and Bollerslev, T. (1998b). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *J. Finance* **53** 219–265.

Andersen, T.G. and Bollerslev, T. (1998c). Towards a unifoed framework for high and low frequency return volaitlity modeling. *Stat. Neerl.* **52** 273–302.

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2010). Parametric and nonparametric measurement of volatility. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L.P. Hansen, eds.) 67–138. Amsterdam: North Holland.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001). The distribution of exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Wu, G. (2006). Realized beta: Persistence and predictability. *Adv. Econom.* **20** 1–39.

Andersen, T., Li, Y., Todorov, V. and Zhou, B. (2023). Volatility measurement with pockets of extreme return persistence. *J. Econometrics* **237** 105048.

Ang, A., Chen, J. and Xing, Y. (2006). Downside risk. *Rev. Financ. Stud.* **19** 1191–1239.

Bacry, E., Mastromatteo, I. and Muzy, J. (2015). Hawkes processes in finance. *Mark. Microstruct. Liq.* **1**.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **74** 3–30.

Bandi, F.M. and Russell, J.R. (2008). Microstructure noise, realized variance, and optimal sampling. *Rev. Econ. Stud.* **75** 339–369.

Barndorff-Nielsen, O.E. (1979). Models for non-Gaussian variation, with applications to turbulence. *Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci.* **368** 501–520.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–65.

Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **24** 1–14.

Barndorff-Nielsen, O.E. (1998a). Probability and statistics: Selfdecomposability, finance and turbulence. In *Probability Towards 2000. Proceedings of a Symposium Held 2-5 October 1995 at Columbia University* (L. Accardi and C.C. Heyde, eds.) 47–57. New York: Springer-Verlag.

Barndorff-Nielsen, O.E. (1998b). Processes of normal inverse Gaussian type. *Finance Stoch.* **2** 41–68.

Barndorff-Nielsen, O.E. (2001). Superposition of Ornstein-Uhlenbeck type processes. *Theory Probab. Appl.* **46** 175–194.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2010). Modelling electricity forward markets by ambit fields. CREATES Research Paper 2010–41.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A. (2013). Modelling electricity futures by ambit fields. *Adv. Appl. Probab.* **46** 719–745.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A. (2015a). Cross-commodity modelling by multivariate ambit fields. In *Commodities, Energy and Environmental Finance* (R. Aid, M. Ludkovski and R. Sircar, eds.) **74** 109–148. Springer.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2015b). Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. *Banach Center Publ.* **104** 25–60.

Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A. (2018). *Ambit Stochastics*. Heidelberg: Springer.

Barndorff-Nielsen, O.E., Corcuera, J.M. and Podolskij, M. (2009). Power variation for Gaussian processes with stationary increments. *Stochastic Process. Appl.* **119** 1845–1865.

Barndorff-Nielsen, O.E., Corcuera, J.M. and Podolskij, M. (2011). Multipower variation for Brownian semistationary processes. *Bernoulli* **17** 1159–1194.

Barndorff-Nielsen, O.E., Corcuera, J.M. and Podolskij, M. (2013). Limit theorems for functionals of higher order differences of Brownian semistationary processes. In *Prokhorov and Contemporary Probability Theory* (A. Shiryaev, S.R.S. Varadhan and E. Presman, eds.) **33**. Berlin: Springer.

Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications. *J. R. Stat. Soc. Ser. B* **41** 279–299.

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

Barndorff-Nielsen, O.E., Graversen, S.E. and Shephard, N. (2004). Power variation and stochastic volatility: A review and some new results. *J. Appl. Probab.* **41A** 133–143.

Barndorff-Nielsen, O.E. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalised inverse Gaussian distributions. *Z. Wahrsch. Verw. Gebiete* **38** 309–311.

Barndorff-Nielsen, O.E., Kinnebrouck, S. and Shephard, N. (2010). Measuring downside risk: Realised semivariance. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (T. Bollerslev, J. Russell and M. Watson, eds.) 117–136. Oxford University Press.

Barndorff-Nielsen, O.E., Nicolato, E. and Shephard, N. (2002). Some recent developments in stochastic volatility modelling. *Quant. Finance* **2** 11–23.

Barndorff-Nielsen, O.E., Pollard, D.G. and Shephard, N. (2012). Integer-valued Lévy processes and low latency financial econometrics. *Quant. Finance* **12** 587–605.

Barndorff-Nielsen, O.E. and Prause, K. (2001). Apparent scaling. *Finance Stoch.* **5** 103–113.

Barndorff-Nielsen, O.E. and Shephard, N. (1998). Aggregation and model construction for volatility models. Unpublished discussion paper. Oxford: Nuffield College.

Barndorff-Nielsen, O.E. and Shephard, N. (2001a). Modelling by Lévy processes for financial econometrics. In *Lévy Processes – Theory and Applications* (O.E. Barndorff-Nielsen, T. Mikosch and S. Resnick, eds.) 283–318. Boston: Birkhäuser.

Barndorff-Nielsen, O.E. and Shephard, N. (2001b). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Stat. Soc. Ser. B* **63** 167–241.

Barndorff-Nielsen, O.E. and Shephard, N. (2001c). Normal modified stable processes. *Theory Probab. Math. Statist.* **65** 1–19.

Barndorff-Nielsen, O.E. and Shephard, N. (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B* **64** 253–280.

Barndorff-Nielsen, O.E. and Shephard, N. (2002b). Estimating quadratic variation using realised variance. *J. Appl. Econometrics* **17** 457–477.

Barndorff-Nielsen, O.E. and Shephard, N. (2003a). Integrated OU processes and non-Gaussian OU-based stochastic volatility. *Scand. J. Stat.* **30** 277–295.

Barndorff-Nielsen, O.E. and Shephard, N. (2003b). Realised power variation and stochastic volatility. *Bernoulli* **9** 243–265. Correction published in pages 1109–1111.

Barndorff-Nielsen, O.E. and Shephard, N. (2004a). Econometric analysis of realised covariation: High frequency covariance, regression and correlation in financial economics. *Econometrica* **72** 885–925.

Barndorff-Nielsen, O.E. and Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps (with discussion). *J. Financ. Econom.* **2** 1–48.

Barndorff-Nielsen, O.E. and Shephard, N. (2005a). How accurate is the asymptotic approximation to the distribution of realised volatility? In *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg* (D.W.K. Andrews and J.H. Stock, eds.) 306–331. Cambridge: Cambridge University Press.

Barndorff-Nielsen, O.E. and Shephard, N. (2005b). Multipower variation and stochastic volatility. In *Stochastic Finance* (A. Shiryaev, M. Grossinho, P. Oliveira and M. Esquivel, eds.) 73–82. Springer.

Barndorff-Nielsen, O.E. and Shephard, N. (2005c). Power variation and time change. *Theory Probab. Appl.* **50** 1–15.

Barndorff-Nielsen, O.E. and Shephard, N. (2006a). Econometrics of testing for jumps in financial economics using bipower variation. *J. Financ. Econom.* **4** 1–30.

Barndorff-Nielsen, O.E. and Shephard, N. (2006b). Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes. *J. Econometrics* **131** 217–252.

Barndorff-Nielsen, O.E. and Shephard, N. (2007). Variation, jumps and high frequency data in financial econometrics. In *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress* (R. Blundell, T. Persson and W.K. Newey, eds.). *Econometric Society Monographs* 328–372. Cambridge University Press.

Barndorff-Nielsen, O.E. and Shephard, N. (2010). Volatility. In *Encyclopedia of Quantitative Finance* (R. Cont, ed.) **IV** 1898–1901. Chichester: Wiley.

Barndorff-Nielsen, O.E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl.* **116** 796–806.

Barndorff-Nielsen, O.E. and Shiryaev, A.N. (2010). *Change of Time and Change of Measure*. Singapore: World Scientific.

Barndorff-Nielsen, O.E. and Stelzer, R. (2013). The multivariate supOU stochastic volatility model. *Math. Finance* **23** 275–296.

Barndorff-Nielsen, O.E. and Veraart, A. (2013). Stochastic volatility of volatility and variance risk premia. *J. Financ. Econom.* **11** 1–46.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M. and Shephard, N. (2006a). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev* (Y. Kabanov, R. Lipster and J. Stoyanov, eds.) 33–68. Springer.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J. and Shephard, N. (2006b). Limit theorems for realised bipower variation in econometrics. *Econometric Theory* **22** 677–719.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536.

Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M. and Worner, J. (2009a). Bipower variation for Gaussian processes with stationary increments. *J. Appl. Probab.* **46** 132–150.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2009b). Realised kernels in practice: Trades and quotes. *Econom. J.* **12** C1–C32.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2011a). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162** 149–169.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2011b). Subsampling realised kernels. *J. Econometrics* **160** 204–219.

Barndorff-Nielsen, O.E., Lunde, A., Shephard, N. and Veraart, A.E.D. (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scand. J. Stat.* **41** 693–724.

Bauwens, L. and Hautsch, N. (2009). Modelling financial high frequency data using point processes. In *Handbook of Financial Time Series* (T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch, eds.) 953–979. Springer-Verlag.

Benth, F.E. and Veraat, A.E.D. (2025). Research frontiers in ambit stochastics: In memory of Ole E. Barndorff-Nielsen. *Bernoulli*. Submitted.

Black, F. (1976). Studies of stock price volatility changes. 177–181.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics* **51** 307–327.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). ARCH models. In *The Handbook of Econometrics* (R.F. Engle and D. McFadden, eds.) **4** 2959–3038. Amsterdam: North-Holland.

Bollerslev, T., Patton, A.J. and Quaedvleig, R. (2022). Realized semibetas: Disentangling good and bad downside risks. *J. Financ. Econ.* **144** 227–246.

Bollerslev, T., Patton, A., Li, J. and Quaedvlieg, R. (2020). Realized semicovariances. *Econometrica* **88** 1515–1551.

Bouchaud, J.-P., Mezard, M. and Potters, M. (2002). Statistical properties of stock order books: Empirical results and models. *Quant. Finance* **2** 251–256.

Bowsher, C.G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *J. Econometrics* **141** 876–912.

Campbell, J.Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *J. Financ. Econ.* **31** 281–318.

Christensen, K., Nielsen, M.S. and Podolskij, M. (2023). High-dimensional estimation of quadratic variation based on penalized realized variance. *Stat. Inference Stoch. Process.* **26** 331–359.

Christensen, K., Oomen, R.C. and Reno, R. (2022). The drift burst hypothesis. *J. Econometrics* **227** 461–497.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. London: Chapman and Hall.

Corsi, F. (2009). A simple long memory model of realized volatility. *J. Financ. Econom.* **7** 174–196.

Dacorogna, M.M., Müller, U.A., Olsen, R.B. and Pictet, O.V. (1998). Modelling short term volatility with GARCH and HARCH. In *Nonlinear Modelling of High Frequency Financial Time Series* (C. Dunis and B. Zhou, eds.) 161–176. Chichester: Wiley.

Drost, F.C. and Nijman, T.E. (1993). Temporal aggregation of GARCH processes. *Econometrica* **61** 909–927.

Duffie, D., Filipovic, D. and Schachermayer, W. (2003). Affine processes and applications in finance. *Ann. Appl. Probab.* **13** 984–1053.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50** 987–1007.

Engle, R.F. and Russell, J.R. (1998). Forecasting transaction rates: The autoregressive conditional duration model. *Econometrica* **66** 1127–1162.

Fan, J. and Kim, D. (2018). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *J. Amer. Statist. Assoc.* **113** 1268–1283.

Goodhart, C.A.E. and Payne, R. (2000). *The Foreign Exchange Market: Empirical Studies with High-Frequency Data*. Palgrave Macmillan.

Grahovac, D., Leonenko, N.N. and Taqqu, M.S. (2018). Intermittency of trawl processes. *Statist. Probab. Lett.* **137** 235–242.

Grahovac, D., Leonenko, N.N. and Taqqu, M.S. (2019). Limit theorems, scaling of moments and intermittency for integrated finite variance supOU processes. *Stochastic Process. Appl.* **129** 5113–5150.

Grahovac, D., Leonenko, N.N. and Taqqu, M.S. (2020). The multifaceted behavior of integrated supOU processes: The infinite variance case. *J. Theor. Probab.* **33** 1801–1831.

Grahovac, D., Leonenko, N.N. and Taqqu, M.S. (2021). Intermittency and infinite variance: The case of integrated supOU processes. *Electron. J. Probab.* **26** 1–31.

Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14** 227–238.

Hansen, P.R. and Lunde, A. (2006). Realized variance and market microstructure noise (with discussion). *J. Bus. Econom. Statist.* **24** 127–218.

Hasbrouck, J. (1991). The summary informativeness of stock trades: An econometric analysis. *Rev. Financ. Stud.* **4** 571–595.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Springer.

Hawkes, A.G. (2018). Hawkes processes and their applications to finance: A review. *Quant. Finance* **18** 193–198.

Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379.

Jacod, J. (2018). Limit of random measures associated with the increments of a Brownian semimartingale. *Econometric Theory* **16** 526–569. Preprint number 120, Laboratoire de Probabilitiés, Université Pierre et Marie Curie, Paris, 1994.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049 https://doi.org/10.1214/aop/1022855419

Jacod, J. and Protter, P. (2012). *Discretization of Processes*. Heidelberg: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276.

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

Lépingle, D. (1976). La variation d'ordre p des semi-martingales. *Z. Wahrsch. Verw. Gebiete* **36** 295–316. MR0420837 https://doi.org/10.1007/BF00532696

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric Brownian motion. *G. Ist. Ital. Attuari* **LXIV** 19–47.

Mancini, C. (2004). Estimation of the characteristics of jump of a general Poisson-diffusion process. *Scand. Actuar. J.* **1** 42–52. MR2045358 https://doi.org/10.1080/034612303100170091

Mancini, C. (2023). Drift burst test statistic in the presence of infinite variation jumps. *Stochastic Process. Appl.* **163** 535–591. MR4616048 https://doi.org/10.1016/j.spa.2023.06.010

Merton, R.C. (1980). On estimating the expected return on the market: An exploratory investigation. *J. Financ. Econ.* **8** 323–361.

Mykland, P.A. and Zhang, L. (2012). The econometrics of high frequency data. In *Statistical Methods for Stochastic Differential Equations* (M. Kessler, A. Lindner and M. Sørensen, eds.) 109–190. Chapman & Hall/CRC Press. MR2976983 https://doi.org/10.1201/b12126-3

Mykland, P. and Zhang, L. (2016). Between data cleaning and inference: Pre-averaging and robust estimators of the efficient price. *J. Econometrics* **194** 242–262. MR3536979 https://doi.org/10.1016/j.jeconom.2016.05.005

Nelson, D.B. (1991). Conditional heteroskedasticity in asset pricing: A new approach. *Econometrica* **59** 347–370. MR1097532 https://doi.org/10.2307/2938260

Nelson, D.B. and Foster, D.P. (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica* **62** 1–41. MR1258664 https://doi.org/10.2307/2951474

Nicolato, E. and Venardos, E. (2003). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Math. Finance* **13** 445–466. MR2003131 https://doi.org/10.1111/1467-9965.t01-1-00175

Officer, R.R. (1973). The variability of the market factor of the New York stock exchange. *J. Bus.* **46** 434–453.

Patton, A.J. and Sheppard, K.K. (2015). Good volatility, bad volatility: Signed jumps and persistence of volatility. *Rev. Econ. Stat.* **97** 683–697.

Phillips, P.C.B. and Yu, J. (2023). Information loss in volatility measurement with flat price trading. *Empir. Econ.* **64** 2957–2999.

Potters, M. and Bouchaud, J.-P. (2003). More statistical properties of order books and price impact. *Phys. A, Stat. Mech. Appl.* **324** 133–140.

Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security prices. Working paper 11, Graduate School of Business Administration, University of California, Berkeley. Reprinted in Shephard (2005).

Roy, A.D. (1952). Safety first and the holding of assets. *Econometrica* **20** 341–449.

Russell, J.R. and Engle, R.F. (2010). Analysis of high-frequency data. In *Handbook of Financial Econometrics: Tools and Techniques* (Y. Ait-Sahalia and L.P. Hansen, eds.) 383–426. Elsevier.

Rydberg, T.H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *J. Financ. Econom.* **1** 2–25.

Scholes, M. and Williams, J. (1977). Estimating betas from nonsynchronous trading. *J. Financ. Econ.* **5** 309–27.

Schwert, G.W. (1989). Why does stock market volatility change over time? *J. Finance* **44** 1115–1153.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds.) 1–67. London: Chapman & Hall. MR1317097 https://doi.org/10.1007/978-1-4899-3210-5

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press. MR2203295

Sortino, F. and Satchell, S.E. (2001). *Managing Downside Risk in Financial Markets*. Butterworth-Heinemann.

Sortino, F.A. and van der Meer, R. (1991). Downside risk. *J. Portf. Manag.* **17** 27–31.

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450 https://doi.org/10.1198/016214505000000169

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *J. Bus. Econom. Statist.* **14** 45–52.

Zhou, B. (1998). Parametric and nonparametric volatility measurement. In *Nonlinear Modelling of High Frequency Financial Time Series* (C.L. Dunis and B. Zhou, eds.) 109–123. New York: John Wiley Sons Ltd.

Bernoulli **32**(1), 2026, 81–95
https://doi.org/10.3150/25-BEJ1906

O. E. Barndorff-Nielsen's approximate conditional inference

NANCY REID[1,a] and HEATHER BATTEY[2,b]

[1]*Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [a]*nancym.reid@utoronto.ca*
[2]*Department of Mathematics, Imperial College London, London, UK,* [b]*h.battey@imperial.ac.uk*

We review three interrelated contributions by Barndorff-Nielsen to statistical theory: his p^* approximation, modified profile likelihood function, and r^* approximation. The development was important for the foundations of inference, and the r^* approximation in particular has also proved useful in a range of applications.

Keywords: Ancillary statistic; asymptotic expansion; conditional inference; modified profile likelihood; large deviation

References

Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika* **67** 293–310. MR0581727 https://doi.org/10.1093/biomet/67.2.293

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

Barndorff-Nielsen, O.E. (1984). On conditionality resolution and the likelihood ratio for curved exponential models. *Scand. J. Stat.* **11** 157–170. MR0767240

Barndorff-Nielsen, O.E. (1986a). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. MR0855891 https://doi.org/10.2307/2336207

Barndorff-Nielsen, O.E. (1986b). Likelihood and observed geometries. *Ann. Statist.* **14** 856–873. MR0856794 https://doi.org/10.1214/aos/1176350038

Barndorff-Nielsen, O.E. (1987). Contribution to discussion of "Parameter orthogonality and approximate conditional inference". *J. Roy. Statist. Soc. Ser. B, Methodol.* **49** 18–20.

Barndorff-Nielsen, O.E. (1990). Approximate interval probabilities. *J. Roy. Statist. Soc. Ser. B, Methodol.* **52** 485–496. MR1086800

Barndorff-Nielsen, O.E. (1991). Modified signed log-likelihood ratio. *Biometrika* **78** 557–563. MR1130923 https://doi.org/10.1093/biomet/78.3.557

Barndorff-Nielsen, O.E. (2018). *Stochastics in Science: Some Autobiographical Notes*. Højbjerg: Forlaget Morgenstjerne.

Barndorff-Nielsen, O.E. and Blaesild, P. (1993). Orthogeodesic models. *Ann. Statist.* **21** 1018–1039. MR1232530 https://doi.org/10.1214/aos/1176349162

Barndorff-Nielsen, O.E. and Chamberlin, S.R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika* **81** 485–499. MR1311092 https://doi.org/10.1093/biomet/81.3.485

Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **41** 279–312. MR0557595

Barndorff-Nielsen, O.E. and Cox, D.R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B, Methodol.* **46** 483–495. MR0790634

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall. MR1010226 https://doi.org/10.1007/978-1-4899-3424-6

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall. MR1317097 https://doi.org/10.1007/978-1-4899-3210-5

1350-7265

Barndorff-Nielsen, O.E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75** 374–378. MR0946056 https://doi.org/10.1093/biomet/75.2.374

Barndorff-Nielsen, O.E. and Jupp, P.E. (1997). Statistics, yokes and symplectic geometry. *Ann. Fac. Sci. Toulouse Math.* **6** 389–427. MR1610891

Barndorff-Nielsen, O.E. and McCullagh, P. (1993). A note on the relation between modified profile likelihood and the Cox-Reid adjusted profile likelihood. *Biometrika* **80** 321–328. MR1243507 https://doi.org/10.1093/biomet/80.2.321

Barndorff-Nielsen, O.E. and Wood, A.T.A. (1998). On large deviations and choice of ancillary for p^* and r^*. *Bernoulli* **4** 35–63. MR1611871 https://doi.org/10.2307/3318531

Bartlett, M.S. (1936). Statistical information and properties of sufficiency. *Proc. R. Soc. Lond. A* **154** 124–137. MR0024103

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A* **160** 268–282. MR0024103

Battey, H.S. (2024). Maximal co-ancillarity and maximal co-sufficiency. *Inf. Geom.* **7** 355–369.

Bickel, P.J. and Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction—a Bayesian argument. *Ann. Statist.* **18** 1070–1090. MR1062699 https://doi.org/10.1214/aos/1176347740

Blaesild, P. (1994). Maximum likelihood estimation in exponential orthogeodesic models. *Ann. Inst. Statist. Math.* **46** 43–55. MR1272746 https://doi.org/10.1007/BF00773591

Brazzale, A.R., Davison, A.C. and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge: Cambridge University Press. MR2342742 https://doi.org/10.1017/CBO9780511611131

Cox, D.R. (1958a). The regression analysis of binary sequences. *J. Roy. Statist. Soc. Ser. B, Methodol.* **20** 215–242. MR0099097

Cox, D.R. (1958b). Two further applications of a model for binary regression. *Biometrika* **45** 562–565.

Cox, D.R. (1958c). Some problems connected with statistical inference. *Ann. Math. Stat.* **29** 357–372. MR0094890 https://doi.org/10.1214/aoms/1177706618

Cox, D.R. (1968). Notes on some aspects of regression analysis (with discussion). *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* **131** 265–279.

Cox, D.R. (1970). *The Analysis of Binary Data*. London: Methuen. MR0282453

Cox, D.R. (1980). Local ancillarity. *Biometrika* **67** 279–286. MR0581725 https://doi.org/10.1093/biomet/67.2.279

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **49** 1–39. MR0893334

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

Daniels, H.E. (1987). Tail probability approximations. *Int. Stat. Rev.* **55** 37–48. MR0962940 https://doi.org/10.2307/1403269

Davison, A.C. and Reid, N. (2024). The tangent exponential model. In *Handbook of Bayesian, Fiducial, and Frequentist Inference* (J.O. Berger, X.L. Meng, N. Reid and M. Xie, eds.) 210–237. Boca Raton: Chapman and Hall/CRC.

Davison, A.C., Fraser, D.A.S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *J. Amer. Statist. Assoc.* **109** 302–314. MR3180565 https://doi.org/10.1080/01621459.2013.839451

DiCaterina, C., Reid, N. and Sartori, N. (2025). Directional tests in Gaussian graphical models. *Statist. Sinica* **35** 361–387. https://doi.org/10.5705/ss.202022.0394. MR4853618 https://doi.org/10.5705/ss.202022.0394

DiCiccio, T.J. and Efron, B. (1996). Bootstrap confidence intervals. *Statist. Sci.* **11** 189–228. MR1436647 https://doi.org/10.1214/ss/1032280214

DiCiccio, T.J., Kuffner, T.A. and Young, G.A. (2017). The formal relationship between analytic and bootstrap approaches to parametric inference. *J. Statist. Plann. Inference* **191** 81–87. MR3679110 https://doi.org/10.1016/j.jspi.2017.05.007

Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–482. MR0521817 https://doi.org/10.1093/biomet/65.3.457

Elkantassi, S. and Davison, A.C. (2022). Space oddity? A statistical formulation of conjunction assessment. *J. Guid. Control Dynamics* **45** 2258–2274.

Elkantassi, S., Bellio, R., Brazzale, A.R. and Davison, A.C. (2023). Improved inference for a boundary parameter. *Canad. J. Statist.* **51** 780–799. MR4635242 https://doi.org/10.1002/cjs.11791

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc. Lond. A* **144** 285–307.
Fraser, D.A.S. (1964). Local conditional sufficiency. *J. Roy. Statist. Soc. Ser. B, Methodol.* **26** 52–62. MR0169327
Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods. *Biometrika* **77** 65–76. MR1049409 https://doi.org/10.1093/biomet/77.1.65
Fraser, D.A.S., Reid, N. and Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103** 625–639. MR3551788 https://doi.org/10.1093/biomet/asw022
Hinkley, D.V. (1980). Likelihood as approximate pivotal. *Biometrika* **67** 287–292. MR0581726 https://doi.org/10.1093/biomet/67.2.287
Jensen, J.L. (1987). On asymptotic expansions in non-ergodic models. *Scand. J. Stat.* **14** 305–318. MR0943291
Jensen, J.L. (1995). *Saddlepoint Approximations*. Oxford: Oxford University Press. MR1354837
Jensen, J.L. (2021). On the use of saddlepoint approximations in high dimensional inference. *Sankhyā* **83-A** 379–392. MR4227217 https://doi.org/10.1007/s13171-019-00188-x
Jupp, P.E. (2024). A parameterisation-invariant modification of the score test. *Inf. Geom.* **7** 429–439. MR4682880 https://doi.org/10.1007/s41884-023-00101-4
Lee, S.M.S. and Young, G.A. (2005). Parametric bootstrapping with nuisance parameters. *Statist. Probab. Lett.* **71** 143–153. MR2126770 https://doi.org/10.1016/j.spl.2004.10.026
Lee, S.M.S. and Young, G.A. (2016). Distribution of likelihood-based p-values under a local alternative hypothesis. *Biometrika* **103** 641–652. MR3551789 https://doi.org/10.1093/biomet/asw021
Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12** 475–490. MR0569438 https://doi.org/10.2307/1426607
McCormack, A., Reid, N., Sartori, N. and Theivendran, S.A. (2019). A directional look at F-tests. *Canad. J. Statist.* **47** 619–627. MR4035792 https://doi.org/10.1002/cjs.11515
McCullagh, P. (1984). Local sufficiency. *Biometrika* **71** 233–244. MR0767151 https://doi.org/10.1093/biomet/71.2.233
McCullagh, P. (1987). *Tensor Methods in Statistics* (2nd Edition Dover 2018). London: Chapman & Hall. MR0907286
Pierce, D.A. and Bellio, R. (2017). Modern likelihood-frequentist inference. *Int. Stat. Rev.* **85** 519–541. MR3723615 https://doi.org/10.1111/insr.12232
Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **54** 701–738. MR1185218
Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31** 1695–1731. MR2036388 https://doi.org/10.1214/aos/1074290325
Reid, N. and Fraser, D.A.S. (2010). Mean loglikelihood and higher-order approximations. *Biometrika* **97** 159–170. MR2594424 https://doi.org/10.1093/biomet/asq001
Ritz, C. and Skovgaard, I.M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika* **92** 507–517. MR2202642 https://doi.org/10.1093/biomet/92.3.507
Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90** 533–549. MR2006833 https://doi.org/10.1093/biomet/90.3.533
Severini, T.A. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika* **86** 235–347. MR1705371 https://doi.org/10.1093/biomet/86.2.235
Skovgaard, I.M. (1985). A second-order investigation of asymptotic ancillarity. *Ann. Statist.* **13** 534–551. MR0790555 https://doi.org/10.1214/aos/1176349537
Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24** 875–887. MR0913828 https://doi.org/10.2307/3214212
Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *J. Roy. Statist. Soc. Ser. B, Methodol.* **50** 269–280. MR0964180
Skovgaard, I.M. (1990a). On the density of minimum contrast estimators. *Ann. Statist.* **18** 779–789. MR1056336 https://doi.org/10.1214/aos/1176347625
Skovgaard, I.M. (1990b). *Analytic Statistical Models*. *IMS Monograph Series*. Hayward, CA. MR1218396
Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2** 145–165. MR1410135 https://doi.org/10.2307/3318548

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scand. J. Stat.* **28** 3–32. MR1844348 https://doi.org/10.1111/1467-9469.00223

Sørensen, M. (2026). Ole Barndorff-Nielsen: Sand, wind, and inference. Bernoulli.

Sweeting, T.J. (1992). Asymptotic ancillarity and conditional inference for stochastic processes. *Ann. Statist.* **20** 580–589. MR1150364 https://doi.org/10.1214/aos/1176348542

Bernoulli **32**(1), 2026, 96–126
https://doi.org/10.3150/24-BEJ1851

On the impossibility of detecting a late change-point in the preferential attachment random graph model

IBRAHIM KADDOURI[1,a], ZACHARIE NAULET[2,c] and ÉLISABETH GASSIAT[1,b]

[1] *Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France,*
[a] *ibrahim.kaddouri@universite-paris-saclay.fr,* [b] *elisabeth.gassiat@universite-paris-saclay.fr*
[2] *Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France,* [c] *zacharie.naulet@inrae.fr*

We consider the problem of late change-point detection under the preferential attachment random graph model with time dependent attachment function. This can be formulated as a hypothesis testing problem where the null hypothesis corresponds to a preferential attachment model with a constant affine attachment parameter δ_0 and the alternative corresponds to a preferential attachment model where the affine attachment parameter changes from δ_0 to δ_1 at a time $\tau_n = n - \Delta_n$ where $0 \le \Delta_n \le n$ and n is the size of the graph. It was conjectured in (Bet et al. (2023)) that when observing only the unlabeled graph, detection of the change is not possible for $\Delta_n = o(n^{1/2})$. In this work, we make a step towards proving the conjecture by proving the impossibility of detecting the change when $\Delta_n = o(n^{1/3})$. We also study change-point detection in the case where the labeled graph is observed and show that change-point detection is possible if and only if $\Delta_n \to \infty$, thereby exhibiting a strong difference between the two settings.

Keywords: Change-point detection; contiguity; preferential attachment; random graphs

References

[1] Adamic, L.A. and Huberman, B.A. (2000). Power-law distribution of the World Wide Web. *Science* **287** 2115–2115. https://doi.org/10.1126/science.287.5461.2115a

[2] Banerjee, S., Bhamidi, S. and Carmichael, I. (2023). Fluctuation bounds for continuous time branching processes and evolution of growing trees with a change point. *Ann. Appl. Probab.* **33** 2919–2980. MR4612658 https://doi.org/10.1214/22-aap1881

[3] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

[4] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Phys. A, Stat. Mech. Appl.* **311** 590–614. MR1943379 https://doi.org/10.1016/S0378-4371(02)00736-7

[5] Bet, G., Bogerd, K., Castro, R.M. and van der Hofstad, R. (2023). Detecting a late changepoint in the preferential attachment model. arXiv preprint. arXiv:2310.02603.

[6] Bhamidi, S., Jin, J. and Nobel, A. (2018). Change point detection in network models: Preferential attachment and long range dependence. *Ann. Appl. Probab.* **28** 35–78. MR3770872 https://doi.org/10.1214/17-AAP1297

[7] Bollobás, B., Borgs, C., Chayes, J. and Riordan, O. (2003). Directed scale-free graphs. In *Proc. Annu. ACM-SIAM Symp. Discrete Algorithms. SODA '03* 132–139. USA: Society for Industrial and Applied Mathematics. MR1974912

[8] Briend, S., Giraud, C., Lugosi, G. and Sulem, D. (2024). Estimating the history of a random recursive tree. *Bernoulli.* To appear.

[9] Bubeck, S., Mossel, E. and Rácz, M.Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Netw. Sci. Eng.* **2** 30–39. MR3361606 https://doi.org/10.1109/TNSE.2015.2397592

[10] Cirkovic, D., Wang, T. and Resnick, S.I. (2023). Preferential attachment with reciprocity: Properties and estimation. *J. Complex Netw.* **11**. MR4636652 https://doi.org/10.1093/comnet/cnad031
[11] Cirkovic, D., Wang, T. and Zhang, X. (2023). Likelihood-based inference for random networks with change-points. arXiv preprint. arXiv:2206.01076.
[12] Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On power-law relationships of the Internet topology. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. SIGCOMM '99* 251–262. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/316188.316229
[13] Gao, F. and van der Vaart, A. (2017). On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Process. Appl.* **127** 3754–3775. MR3707244 https://doi.org/10.1016/j.spa.2017.03.008
[14] Gao, F., van der Vaart, A., Castro, R. and van der Hofstad, R. (2017). Consistent estimation in general sublinear preferential attachment trees. *Electron. J. Stat.* **11** 3979–3999. MR3714305 https://doi.org/10.1214/17-EJS1356
[15] Holme, P. and Saramäki, J. (2019). *Temporal Network Theory 2*. Springer. MR3971532 https://doi.org/10.1007/978-3-030-23495-9
[16] Kaddouri, I., Naulet, Z. and Gassiat, É. (2026). Supplement to "On the impossibility of detecting a late change-point in the preferential attachment random graph model." https://doi.org/10.3150/24-BEJ1851SUPP
[17] Medo, M., Cimini, G. and Gualdi, S. (2011). Temporal effects in the growth of networks. *Phys. Rev. Lett.* **107** 238701. https://doi.org/10.1103/PhysRevLett.107.238701
[18] Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409. MR1812610 https://doi.org/10.1073/pnas.021544898
[19] Temme, N.M. (1996). *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. Wiley. MR1376370 https://doi.org/10.1002/9781118032572
[20] van der Hofstad, R. (2016). *Random Graphs and Complex Networks*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. MR3617364 https://doi.org/10.1017/9781316779422
[21] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. MR1652247 https://doi.org/10.1017/CBO9780511802256
[22] Wang, D., Yu, Y. and Rinaldo, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *Ann. Statist.* **49** 203–232. MR4206675 https://doi.org/10.1214/20-AOS1953
[23] Watts, D.J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. USA: Princeton University Press. MR1716136
[24] Watts, D.J. and Strogatz, S.H. (2006). Collective dynamics of 'small-world' networks. In *The Structure and Dynamics of Networks* 301–303. Princeton: Princeton University Press. MR1716136 https://doi.org/10.1515/9781400841356.301
[25] Zhao, Z., Chen, L. and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. arXiv preprint. arXiv:1908.01823.

Bernoulli **32**(1), 2026, 127–152
https://doi.org/10.3150/25-BEJ1852

Rates of convergence and normal approximations for estimators of local dependence random graph models

JONATHAN R. STEWART[1,a]

[1]*Department of Statistics, Florida State University, Tallahassee, FL, USA,* [a]*jrstewart@fsu.edu*

Local dependence random graph models are a class of block models for network data which allow for dependence among edges under a local dependence assumption defined around the block structure of the network. Since being introduced by Schweinberger and Handcock (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** (2015) 647–676), research in the statistical network analysis and network science literatures have demonstrated the potential and utility of this class of models. In this work, we provide the first theory for estimation and inference which ensures consistent and valid inference of parameter vectors of local dependence random graph models. This is accomplished by deriving convergence rates of estimation and inference procedures for local dependence random graph models based on a single observation of the graph, allowing both the number of model parameters and the sizes of blocks to tend to infinity. First, we derive non-asymptotic bounds on the ℓ_2-error of maximum likelihood estimators with convergence rates, outlining conditions under which these rates are minimax optimal. Second, and more importantly, we derive non-asymptotic bounds on the error of the multivariate normal approximation. These theoretical results are the first to achieve both optimal rates of convergence and non-asymptotic bounds on the error of the multivariate normal approximation for parameter vectors of local dependence random graph models.

Keywords: Local dependence random graph model; minimax bounds; multivariate normal approximation; network data; statistical network analysis

References

Agneessens, F., Trincado-Munoz, F.J. and Koskinen, J. (2024). Network formation in organizational settings: Exploring the importance of local social processes and team-level contextual variables in small groups using Bayesian hierarchical ERGMs. *Soc. Netw.* **77** 104–117. https://doi.org/10.1016/j.socnet.2022.07.001

Babkin, S., Stewart, J.R., Long, X. and Schweinberger, M. (2020). Large-scale estimation of random graph models with local dependence. *Comput. Statist. Data Anal.* **152** 107029. MR4130894 https://doi.org/10.1016/j.csda.2020.107029

Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*. Institute of Mathematical Statistics. MR0882001 https://doi.org/10.1214/lnms/1215466757

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

Dahbura, J.N.M., Komatsu, S., Nishida, T. and Mele, A. (2021). A structural model of business card exchange networks. arXiv preprint, arXiv:2105.12704. https://doi.org/10.48550/arXiv.2105.12704

Dahbura, J.N.M., Komatsu, S., Nishida, T. and Mele, A. (2023). Homophily and community structure at scale: An application to a large professional network. *AEA Pap. Proc.* **113** 156–160. https://doi.org/10.1257/pandp.20231094

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 657–683. MR1185217 https://doi.org/10.1111/j.2517-6161.1992.tb01443.x

1350-7265

Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613 https://doi.org/10.1006/jmva.1999.1874

Häggström, O. and Jonasson, J. (1999). Phase transition in the random triangle model. *J. Appl. Probab.* **36** 1101–1115. MR1742153 https://doi.org/10.1239/jap/1032374758

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

Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–50. MR0608176 https://doi.org/10.1080/01621459.1981.10477598

Hunter, D.R. and Handcock, M.S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069

Janková, J. and van de Geer, S. (2018). Inference in high-dimensional graphical models. In *Handbook of Graphical Models* (M. Maathuis, M. Drton, S. Lauritzen and M. Wainwright, eds.) 325–350. CRC Press. MR3888003 https://doi.org/10.1080/01621459.2020.1801279

Jonasson, J. (1999). The random triangle model. *J. Appl. Probab.* **36** 852–867. MR1737058 https://doi.org/10.1239/jap/1032374639

Krivitsky, P.N., Hunter, D.R., Morris, M. and Klumb, C. (2023). ergm 4: New features for analyzing exponential-family random graph models. *J. Stat. Softw.* **105** 1–44. https://doi.org/10.18637/jss.v105.i06

Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential Random Graph Models for Social Networks*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894701

Mele, A. (2022). A structural model of homophily and clustering in social networks. *J. Bus. Econom. Statist.* **40** 1377–1389. MR4439296 https://doi.org/10.1080/07350015.2021.1930013

Newman, M.E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69** 026113. MR1975193 https://doi.org/10.1103/PhysRevE.69.026113

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

Ravikumar, P., Wainwright, M.J. and Lafferty, J.D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691

Schubert, D. and Brand, A. (2022). "Whom should I talk to?": Role prescription and hierarchy building in supervised living groups. *Soc. Incl.* **10** 295–306. https://doi.org/10.17645/si.v10i3.5406

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841 https://doi.org/10.1198/jasa.2011.tm10747

Schweinberger, M. (2020). Consistent structure estimation of exponential-family random graph models with block structure. *Bernoulli* **26** 1205–1233. MR4058365 https://doi.org/10.3150/19-BEJ1153

Schweinberger, M. and Handcock, M.S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 https://doi.org/10.1111/rssb.12081

Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* **48** 374–396. MR4065166 https://doi.org/10.1214/19-AOS1810

Schweinberger, M., Krivitsky, P.N., Butts, C.T. and Stewart, J.R. (2020). Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios. *Statist. Sci.* **35** 627–662. MR4175389 https://doi.org/10.1214/19-STS743

Shao, M., Zhang, Y., Wang, Q., Zhang, Y., Luo, J. and Yan, T. (2021). L-2 Regularized maximum likelihood for β-model in large and sparse networks. arXiv preprint, arXiv:2110.11856. https://doi.org/10.48550/arXiv.2110.11856

Snijders, T.A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *J. Soc. Struct.* **3** 1–40.

Stewart, J.R. (2026). Supplement to "Rates of convergence and normal approximations for estimators of local dependence random graph models." https://doi.org/10.3150/25-BEJ1852SUPP

Stewart, J.R. and Schweinberger, M. (2019). R package mlergm. Multilevel Exponential-Family Random Graph Models. The Comprehensive R Archive Network.

Stewart, J.R. and Schweinberger, M. (2020). Pseudo-likelihood-based M-estimation of random graphs with dependent edges and parameter vectors of increasing dimension. arXiv preprint, arXiv:2012.07167. https://doi.org/10.48550/arXiv.2012.07167

Stewart, J., Schweinberger, M., Bojanowski, M. and Morris, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Soc. Netw.* **59** 98–119. https://doi.org/10.1016/j.socnet.2018.11.003

Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press. MR3969949 https://doi.org/10.1017/9781108604574

Tolochko, P. and Boomgaarden, H.G. (2024). Same but different: A comparison of estimation approaches for exponential random graph models for multiple networks. *Soc. Netw.* **76** 1–11. https://doi.org/10.1016/j.socnet.2023.05.003

Whetsell, T.A., Kroll, A. and Dehart-Davis, L. (2021). Formal hierarchies and informal networks: How organizational structure shapes information search in local government. *J. Public Adm. Res. Theory* **31** 653–669. https://doi.org/10.1093/jopart/muab003

Bernoulli **32**(1), 2026, 153–178
https://doi.org/10.3150/25-BEJ1853

Markov properties of Gaussian random fields on compact metric graphs

DAVID BOLIN[1,a], ALEXANDRE B. SIMAS[1,b] and JONAS WALLIN[2,c]

[1]*Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia,* [a]*david.bolin@kaust.edu.sa,* [b]*alexandre.simas@kaust.edu.sa*
[2]*Department of Statistics, Lund University, Lund, Sweden,* [c]*jonas.wallin@stat.lu.se*

There has recently been much interest in Gaussian fields on linear networks and, more generally, on compact metric graphs. One proposed strategy for defining such fields on a metric graph Γ is through a covariance function that is isotropic in a metric on the graph. Another is through a fractional-order differential equation $L^{\alpha/2}(\tau u) = \mathcal{W}$ on Γ, where $L = \kappa^2 - \nabla(a\nabla)$ for (sufficiently nice) functions κ, a, and $\mathcal{W}$ is Gaussian white noise. We study Markov properties of these two types of fields. First, we show that no Gaussian random fields exist on general metric graphs that are both isotropic and Markov. Then, we show that the second type of fields, the generalized Whittle–Matérn fields, are Markov if $\alpha \in \mathbb{N}$, and conversely, if a and κ are constant and u is Markov, then $\alpha \in \mathbb{N}$. Further, if $\alpha \in \mathbb{N}$, a generalized Whittle–Matérn field u is Markov of order α, which means that the field u in one region $S \subset \Gamma$ is conditionally independent of u in $\Gamma \setminus S$ given the values of u and its $\alpha - 1$ derivatives on ∂S. Finally, we provide two results as consequences of the theory developed: first we prove that the Markov property implies an explicit characterization of u on a fixed edge e, revealing that the conditional distribution of u on e given the values at the two vertices connected to e is independent of the geometry of Γ; second, we show that the solution to $L^{1/2}(\tau u) = \mathcal{W}$ on Γ can obtained by conditioning independent generalized Whittle–Matérn processes on the edges, with $\alpha = 1$ and Neumann boundary conditions, on being continuous at the vertices.

Keywords: Gaussian processes; networks; quantum graphs; stochastic partial differential equations

References

[1] Anderes, E., Møller, J. and Rasmussen, J.G. (2020). Isotropic covariance functions on graphs and their edges. *Ann. Statist.* **48** 2478–2503. MR4134803 https://doi.org/10.1214/19-AOS1896
[2] Ash, R.B. and Gardner, M.F. (1975). *Topics in Stochastic Processes*. *Probability and Mathematical Statistics* **27**. New York-London: Academic Press [Harcourt Brace Jovanovich, Publishers]. MR0448463 https://doi.org/10.1016/C2013-0-06165-8 MR448463
[3] Balakrishnan, A.V. (1960). Fractional powers of closed operators and the semigroups generated by them. *Pacific J. Math.* **10** 419–437. MR0115096
[4] Berestycki, N., Powell, E. and Ray, G. (2020). A characterisation of the Gaussian free field. *Probab. Theory Related Fields* **176** 1259–1301. MR4087493 https://doi.org/10.1007/s00440-019-00939-9
[5] Bolin, D., Kirchner, K. and Kovács, M. (2020). Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. *IMA J. Numer. Anal.* **40** 1051–1073. MR4092278 https://doi.org/10.1093/imanum/dry091
[6] Bolin, D., Kovács, M., Kumar, V. and Simas, A.B. (2024). Regularity and numerical approximation of fractional elliptic differential equations on compact metric graphs. *Math. Comp.* **93** 2439–2472. MR4759380 https://doi.org/10.1090/mcom/3929
[7] Bolin, D., Simas, A.B. and Wallin, J. (2023). Statistical inference for Gaussian Whittle–Matérn fields on metric graphs. Preprint. Available at arXiv:2304.10372. https://doi.org/10.48550/arXiv.2304.10372
[8] Bolin, D., Simas, A.B. and Wallin, J. (2024). Gaussian Whittle-Matérn fields on metric graphs. *Bernoulli* **30** 1611–1639. MR4699566 https://doi.org/10.3150/23-bej1647

1350-7265

[9] Bolin, D., Simas, A.B. and Wallin, J. (2026). Supplement to "Markov properties of Gaussian random fields on compact metric graphs." https://doi.org/10.3150/25-BEJ1853SUPP

[10] Borovitskiy, V., Azangulov, I., Terenin, A., Mostowsky, P., Deisenroth, M. and Durrande, N. (2021). Matérn Gaussian processes on graphs. In *International Conference on Artificial Intelligence and Statistics* 2593–2601. PMLR.

[11] Cai, Z. and Ding, J. (2024). One-arm exponent of critical level-set for metric graph Gaussian free field in high dimensions. *Probab. Theory Related Fields* 1–86. MR4898098 https://doi.org/10.1007/s00440-024-01295-z

[12] Chandler-Wilde, S.N., Hewett, D.P. and Moiola, A. (2015). Interpolation of Hilbert and Sobolev spaces: Quantitative estimates and counterexamples. *Mathematika* **61** 414–443. MR3343061 https://doi.org/10.1112/S0025579314000278

[13] Cressie, N., Frey, J., Harch, B. and Smith, M. (2006). Spatial prediction on a river network. *J. Agric. Biol. Environ. Stat.* **11** 127. https://doi.org/10.1198/108571106X110649

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

[15] Ding, J. and Wirth, M. (2020). Percolation for level-sets of Gaussian free fields on metric graphs. *Ann. Probab.* **48** 1411–1435. MR4112719 https://doi.org/10.1214/19-AOP1397

[16] Drewitz, A., Prévost, A. and Rodriguez, P.-F. (2022). Cluster capacity functionals and isomorphism theorems for Gaussian free fields. *Probab. Theory Related Fields* **183** 255–313. MR4421175 https://doi.org/10.1007/s00440-021-01090-0

[17] Fuglstad, G.-A., Lindgren, F., Simpson, D. and Rue, H. (2015). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statist. Sinica* **25** 115–133. MR3328806 https://doi.org/10.5705/ss.2013.106w

[18] Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spat. Stat.* **14** 505–531. MR3431054 https://doi.org/10.1016/j.spasta.2015.10.001

[19] Hildeman, A., Bolin, D. and Rychlik, I. (2021). Deformed SPDE models with an application to spatial modeling of significant wave height. *Spat. Stat.* **42**. Paper No. 100449. MR4233262 https://doi.org/10.1016/j.spasta.2020.100449

[20] Isaak, D.J., Ver Hoef, J.M., Peterson, E.E., Horan, D.L. and Nagel, D.E. (2017). Scalable population estimates using spatial-stream-network (SSN) models, fish density surveys, and national geospatial database frameworks for streams. *Can. J. Fish. Aquat. Sci.* **74** 147–156. https://doi.org/10.1139/cjfas-2016-0247

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

[22] Künsch, H. (1979). Gaussian Markov random fields. *J. Fac. Sci., Univ. Tokyo, Sect. 1A, Math.* **26** 53–73. MR0539773

[23] Lawler, G.F. (2007). Schramm-Loewner evolution. Preprint. Available at arXiv:0712.3256. MR2523461 https://doi.org/10.1090/pcms/016/05

[24] Lévy, P. (1956). A special problem of Brownian motion, and a general theory of Gaussian random functions. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955* **II** 133–175. Berkeley-Los Angeles, CA: Univ. California Press. MR0090934

[25] 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. MR4439328 https://doi.org/10.1016/j.spasta.2022.100599

[26] Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

[27] Lupu, T. (2016). From loop clusters and random interlacements to the free field. *Ann. Probab.* **44** 2117–2146. MR3502602 https://doi.org/10.1214/15-AOP1019

[28] Mandrekar, V. (1976). Germ-field Markov property for multiparameter processes. In *Séminaire de Probabilités, X. Lecture Notes in Math.* **511** 78–85. Berlin-New York: Springer. (Première partie, Univ Strasbourg, Strasbourg, année universitaire 1974/1975). MR0440683 https://doi.org/10.1007/BFb0101097

[29] Matérn, B. (1960). Spatial variation. Meddelanden från statens skogsforskningsinstitut 49. MR0169346
[30] McGill, L.M., Brooks, J.R. and Steel, E.A. (2021). Spatiotemporal dynamics of water sources in a mountain river basin inferred through δ^2H and δ^{18}O of water. *Hydrol. Process.* **35** e14063. https://doi.org/10.1002/hyp.14063
[31] McKean, H.P. Jr. (1963). Brownian motion with a several-dimensional time. *Teor. Veroyatn. Primen.* **8** 357–378. MR0157407 https://doi.org/10.1137/1108042 MR157407
[32] Møller, J. and Rasmussen, J.G. (2024). Cox processes driven by transformed Gaussian processes on linear networks—a review and new contributions. *Scand. J. Stat.* **51** 1288–1322. MR4812906 https://doi.org/10.1111/sjos.12720
[33] Pitt, L.D. (1971). A Markov property for Gaussian processes with a multidimensional parameter. *Arch. Ration. Mech. Anal.* **43** 367–391. MR0336798 https://doi.org/10.1007/BF00252003 MR336798
[34] Pitt, L.D. and Robeva, R.S. (2003). On the sharp Markov property for Gaussian random fields and spectral synthesis in spaces of Bessel potentials. *Ann. Probab.* **31** 1338–1376. MR1989436 https://doi.org/10.1214/aop/1055425783
[35] Rozanov, J.A. (1977). Markovian random fields and stochastic partial differential equations. *Mat. Sb. (N.S.)* **103** 590–613, 631. MR0467916 https://doi.org/10.1070/SM1977v032n04ABEH002404 MR467916
[36] Rozanov, Y.A. (1982). *Markov Random Fields*. *Applications of Mathematics*. New York-Berlin: Springer-Verlag. Translated from the Russian by Constance M. Elson. MR0676644 https://doi.org/10.1007/978-1-4613-8190-7 MR676644
[37] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. Boca Raton, FL: Chapman & Hall/CRC. MR2130347 https://doi.org/10.1201/9780203492024
[38] Schramm, O. (2000). Scaling limits of loop-erased random walks and uniform spanning trees. *Israel J. Math.* **118** 221–288. MR1776084 https://doi.org/10.1007/BF02803524
[39] Ver Hoef, J.M., Peterson, E. and Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environ. Ecol. Stat.* **13** 449–464. MR2297373 https://doi.org/10.1007/s10651-006-0022-8
[40] Werner, W. and Powell, E. (2021). Lecture Notes on the Gaussian Free Field. Cours Spécialisés [Specialized Courses] 28. Société Mathématique de France, Paris. MR4466634
[41] Yosida, K. (1995). *Functional Analysis*. *Classics in Mathematics*. Berlin: Springer-Verlag. MR1336382 https://doi.org/10.1007/978-3-642-61859-8

Bernoulli **32**(1), 2026, 179–204
https://doi.org/10.3150/25-BEJ1854

Quantifying and estimating dependence via sensitivity of conditional distributions

JONATHAN ANSARI[a], PATRICK B. LANGTHALER[b], SEBASTIAN FUCHS[c] and WOLFGANG TRUTSCHNIG[d]

Department of Artificial Intelligence and Human Interfaces, University of Salzburg, Salzburg, Austria,
[a]*jonathan.ansari@plus.ac.at,* [b]*patrick.langthaler@plus.ac.at,* [c]*sebastian.fuchs@plus.ac.at,*
[d]*wolfgang@trutschnig.net*

Recently established, directed dependence measures for pairs (X,Y) of random variables build upon the natural idea of comparing the conditional distributions of Y given $X = x$ with the unconditional distribution of Y. They assign pairs (X,Y) values in $[0,1]$, where the value is 0 if and only if X,Y are independent, and it is 1 exclusively for Y being a measurable function of X. Here we show that comparing randomly drawn conditional distributions with each other instead or, equivalently, analyzing how sensitive the conditional distribution of Y given $X = x$ is on x, opens the door to constructing novel families of dependence measures Λ_φ induced by general convex functions $\varphi : \mathbb{R} \to \mathbb{R}$, containing, e.g., Chatterjee's coefficient of correlation as special case. After establishing additional useful properties of Λ_φ we focus on continuous (X,Y), translate Λ_φ to the copula setting, consider the L^p-version and establish an estimator which is strongly consistent in full generality. A real data example and a simulation study illustrate the chosen approach and the performance of the estimator. Complementing the afore-mentioned results, we show how a slight modification of the sensitivity idea underlying Λ_φ can be used to define new measures of explainability generalizing the fraction of explained variance.

Keywords: Chatterjee's correlation coefficient; conditional distributions; copula; dependence measure; explainability; Schur order; sensitivity

References

Ansari, J. (2019). Ordering risk bounds in partially specified factor models. Freiburg im Breisgau: Univ. Freiburg, Fakultät für Mathematik und Physik (Diss.).

Ansari, J. and Fuchs, S. (2023). A simple extension of Azadkia & Chatterjee's rank correlation to a vector of endogenous variables. Preprint, arXiv:2212.01621v2.

Ansari, J. and Rockel, M. (2024). Dependence properties of bivariate copula families. *Depend. Model.* **12** 1–36. https://doi.org/10.1515/demo-2021-0113

Ansari, J. and Rüschendorf, L. (2021). Sklar's theorem, copula products, and ordering results in factor models. *Depend. Model.* **9** 267–306. https://doi.org/10.1515/demo-2021-0113

Ansari, J., Langthaler, P.B., Fuchs, S. and Trutschnig, W. (2026). Supplement to "Quantifying and estimating dependence via sensitivity of conditional distributions." https://doi.org/10.3150/25-BEJ1854SUPP

Azadkia, M. and Chatterjee, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. https://doi.org/10.1214/21-AOS2073

Bickel, P.J. (2022). Measures of independence and functional dependence. Preprint, arXiv:2206.13663v1.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 2009–2022. https://doi.org/10.1080/01621459.2020.1758115

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/047174882X

da Silva Rocha, A., von Diemen, L., Kretzer, D., Matos, S., Rombaldi Bernardi, J. and Magalhaes, J.A. (2020). Visceral adipose tissue measurements during pregnancy. *PhysioNet*. https://doi.org/10.13026/p729-7p53

1350-7265

Deb, N., Ghosal, P. and Sen, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. Preprint, arXiv:2010.01768.

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

Ding, A.A., Dy, J.G., Li, Y. and Chang, Y. (2017). A robust-equitable measure for feature ranking and selection. *J. Mach. Learn. Res.* **18** 1–46. Id/No 71.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography. Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4–7, 2006. Proceedings* 265–284. Berlin: Springer. https://doi.org/10.1007/11681878

Evfimievski, A., Gehrke, J. and Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222. https://doi.org/10.1145/773153.773174

Fuchs, S. (2024). Quantifying directed dependence via dimension reduction. *J. Multivariate Anal.* **201** 21. Id/No 105266. https://doi.org/10.1016/j.jmva.2023.105266

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *J. Bank. Financ.* **83** 70–84. https://doi.org/10.1016/j.jbankfin.2017.06.013

Gamboa, F., Gremaud, P., Klein, T. and Lagnoux, A. (2022). Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli* **28** 2345–2374. https://doi.org/10.3150/21-BEJ1421

Genest, C., Nešlehová, J.G. and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivariate Anal.* **159** 82–110. https://doi.org/10.1016/j.jmva.2017.04.002

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P.C., Mark, R., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101** 215–220. https://doi.org/10.1161/01.CIR.101.23.e215

Griessenberger, F., Junker, R.R. and Trutschnig, W. (2022). On a multivariate copula-based dependence measure and its estimation. *Electron. J. Stat.* **16** 2206–2251. https://doi.org/10.1214/22-EJS2005

Han, F. and Huang, Z. (2024). Azadkia-Chatterjee's correlation coefficient adapts to manifold data. *Ann. Appl. Probab.* **34** 5172–5210. https://doi.org/10.1214/24-AAP2088

Huang, Z., Deb, N. and Sen, B. (2022). Kernel partial correlation coefficient — a measure of conditional dependence. *J. Mach. Learn. Res.* **23** 1–58.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *J. Statist. Plann. Inference* **142** 1189–1197. https://doi.org/10.1016/j.jspi.2011.11.020

Junker, R.R., Griessenberger, F. and Trutschnig, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.* **153** 22. Id/No 107058. https://doi.org/10.1016/j.csda.2020.107058

Kasper, T.M., Fuchs, S. and Trutschnig, W. (2021). On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation. *Bernoulli* **27** 2217–2240. https://doi.org/10.3150/20-BEJ1306

Kinney, J.B. and Atwal, G.S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **111** 3354–3359. https://doi.org/10.1073/pnas.1309933111

Li, X., Mikusiński, P. and Taylor, M.D. (1998). Strong approximation of copulas. *J. Math. Anal. Appl.* **225** 608–623. https://doi.org/10.1006/jmaa.1998.6056

Mroz, T., Fernández-Sánchez, J., Fuchs, S. and Trutschnig, W. (2023). On distributions with fixed marginals maximizing the joint or the prior default probability, estimation, and related results. *J. Statist. Plann. Inference* **223** 33–52. https://doi.org/10.1016/j.jspi.2022.07.005

Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. New York, NY: Springer. https://doi.org/10.1007/0-387-28678-0

Nies, T.G., Staudt, T. and Munk, A. (2023). Transport dependency: Optimal transport based dependency measures. Preprint, arXiv:2105.02073.

Rohde, A. and Steinberger, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. https://doi.org/10.1214/19-AOS1901

Rüschendorf, L. (2013). *Mathematical Risk Analysis*. Berlin: Springer. https://doi.org/10.1007/978-3-642-33590-7

Shaked, M., Sordo, M.A. and Suárez-Llorens, A. (2013). A global dependence stochastic order based on the presence of noise. In *Stochastic Orders in Reliability and Risk. In Honor of Professor Moshe Shaked* 3–39. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-6892-9_1

Shih, J.-H. and Emura, T. (2021). On the copula correlation ratio and its generalization. *J. Multivariate Anal.* **182** 15. Id/No 104708. https://doi.org/10.1016/j.jmva.2020.104708

Strothmann, C., Dette, H. and Siburg, K.F. (2024). Rearranged dependence measures. *Bernoulli* **30** 1055–1078. https://doi.org/10.3150/23-BEJ1624

Sungur, E.A. (2005). A note on directional dependence in regression setting. *Comm. Statist. Theory Methods* **34** 1957–1965. https://doi.org/10.1080/03610920500201228

Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. https://doi.org/10.1016/j.jmaa.2011.06.013

Wiesel, J.C.W. (2022). Measuring association with Wasserstein distances. *Bernoulli* **28** 2816–2832. https://doi.org/10.3150/21-BEJ1438

Bernoulli **32**(1), 2026, 205–225
https://doi.org/10.3150/25-BEJ1855

Haldane's asymptotics for supercritical branching processes in an iid random environment

FLORIN BOENKOST[1,a] and GÖTZ KERSTING[2,b]

[1] *Institut für Mathematik, Universität Wien, Wien, Austria,* [a] *florin.boenkost@univie.ac.at*
[2] *Institut für Mathematik, Goethe Universität, Frankfurt am Main, Germany,* [b] *kersting@math.uni-frankfurt.de*

Branching processes in a random environment are natural generalisations of Galton-Watson processes. In this paper we analyse the asymptotic decay of the survival probability for a sequence of slightly supercritical branching processes in an iid random environment, where the offspring expectation converges from above to 1. We prove that Haldane's asymptotics, known from classical Galton-Watson processes, turns up again in the random environment case, provided that one stays away from the critical/subcritical regime. A central building block is a connection to and a limit theorem for perpetuities with asymptotically vanishing interest rates.

Keywords: Branching process; perpetuity; random environment; supercriticality; survival probability

References

[1] Alsmeyer, G. (2021). Linear fractional Galton-Watson processes in random environment and perpetuities. *Stoch. Qual. Control* **36** 111–127. MR4361568 https://doi.org/10.1515/eqc-2021-0037
[2] Alsmeyer, G., Iksanov, A. and Rösler, U. (2009). On distributional properties of perpetuities. *J. Theor. Probab.* **22** 666–682. MR2530108 https://doi.org/10.1007/s10959-008-0156-8
[3] Athreya, K.B. (1992). Rates of decay for the survival probability of a mutant gene. *J. Math. Biol.* **30** 577–581. MR1173109 https://doi.org/10.1007/BF00948892
[4] Bansaye, V. and Böinghoff, C. (2014). Small positive values for supercritical branching processes in random environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 770–805. MR3224289 https://doi.org/10.1214/13-AIHP538
[5] Blanchet, J. and Glynn, P. (2022). Approximations for the distribution of perpetuities with small discount rates. *Naval Res. Logist.* MR4617531 https://doi.org/10.1002/nav.22058
[6] Borovkov, A.A. (1972). Notes on inequalities for sums of independent variables. *Theory Probab. Appl.* **17** 556–557. MR0309187 https://doi.org/10.1137/1117068
[7] Borovkov, A.A., Fayolle, G. and Korshunov, D.A. (1992). Transient phenomena for Markov chains and applications. *Adv. Appl. Probab.* **24** 322–342. MR1167262 https://doi.org/10.2307/1427695
[8] Buraczewski, D., Damek, E. and Mikosch, T. (2016). *Stochastic Models with Power-Law Tails. The Equation $X = AX + B$. Springer Series in Operations Research and Financial Engineering.* Springer. MR3497380 https://doi.org/10.1007/978-3-319-29679-1
[9] Collamore, J.F. and Vidyashankar, A.N. (2013). Tail estimates for stochastic fixed point equations via nonlinear renewal theory. *Stochastic Process. Appl.* **123** 3378–3429. MR3071384 https://doi.org/10.1016/j.spa.2013.04.015
[10] Dufresne, D. (1990). The distribution of a perpetuity, with applications to risk theory and pension funding. *Scand. Actuar. J.* **1990** 39–79. MR1129194 https://doi.org/10.1080/03461238.1990.10413872
[11] Eshel, I. (1981). On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size—a branching process model. *J. Math. Biol.* **12** 355–362. MR0632151 https://doi.org/10.1007/BF00276922

[12] Fisher, R.A. (1923). On the dominance ratio. *Proc. R. Soc. Edinb.* **42**.
[13] Geiger, J. and Kersting, G. (2000). The survival probability of a critical branching process in random environment. *Theory Probab. Appl.* **45** 607–615. MR1967796 https://doi.org/10.1137/S0040585X97978440
[14] Goldie, C.M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** 126–166. MR1097468 https://doi.org/10.1214/aoap/1177005985
[15] Goldie, C.M. and Maller, R.A. (2000). Stability of perpetuities. *Ann. Probab.* **28** 1195–1218. MR1797309 https://doi.org/10.1214/aop/1019160331
[16] Grama, I., Liu, Q. and Miqueu, E. (2017). Berry-Esseen's bound and Cramér's large deviation expansion for a supercritical branching process in a random environment. *Stochastic Process. Appl.* **127** 1255–1281. MR3619270 https://doi.org/10.1016/j.spa.2016.07.014
[17] Haldane, J.B.S. (1927). A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Proc. Camb. Philos. Soc.* **23** 838–844.
[18] Hambly, B. (1992). On the limiting distribution of a supercritical branching process in a random environment. *J. Appl. Probab.* **29** 499–518. MR1174427 https://doi.org/10.2307/3214889
[19] Hitczenko, P. and Wesołowski, J. (2009). Perpetuities with thin tails revisited. *Ann. Appl. Probab.* **19** 2080–2101. MR2588240 https://doi.org/10.1214/09-AAP603
[20] Hitczenko, P. and Wesołowski, J. (2011). Renorming divergent perpetuities. *Bernoulli* **17** 880–894. MR2817609 https://doi.org/10.3150/10-bej297
[21] Hoppe, F.M. (1992). Asymptotic rates of growth of the extinction probability of a mutant gene. *J. Math. Biol.* **30** 547–566. MR1173107 https://doi.org/10.1007/BF00948890
[22] Huang, C. and Liu, Q. (2014). Convergence in L^p and its exponential rate for a branching process in a random environment. *Electron. J. Probab.* **19** 104. MR3275856 https://doi.org/10.1214/EJP.v19-3388
[23] Iksanov, A., Marynych, A. and Nikitin, A. (2023). Limit theorems for discounted convergent perpetuities. II. *Electron. J. Probab.* **28** 22. MR4536690 https://doi.org/10.1214/23-ejp907
[24] Iksanov, A., Nikitin, A. and Samoilenko, I. (2021). Limit theorems for discounted convergent perpetuities. *Electron. J. Probab.* **26** 25. MR4343568 https://doi.org/10.1214/21-ejp705
[25] Kersting, G. (2020). A unifying approach to branching processes in varying environments. *J. Appl. Probab.* **57** 196–200. MR4094390 https://doi.org/10.1017/jpr.2019.84
[26] Kersting, G. and Vatutin, V.A. (2017). *Discrete Time Branching Processes in Random Environment*. Wiley, New York. MR3587982 https://doi.org/10.1002/9781119452898
[27] Kolmogorov, A.N. (1938). Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk* **2** 1–12.
[28] Korshunov, D.A. (1994). Transient phenomena for real-valued Markov chains. *Theory Probab. Appl.* **38** 149–152. MR1317792 https://doi.org/10.1137/1138012
[29] Nagaev, S.V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129 https://doi.org/10.1214/aop/1176994938
[30] Patwa, Z. and Wahl, L. (2008). The fixation probability of beneficial mutations. *J. R. Soc. Interface* **5** 1279–1289. https://doi.org/10.1098/rsif.2008.0248
[31] Smith, W.L. and Wilkinson, W.E. (1969). On branching processes in random environments. *Ann. Math. Statist.* **40** 814–827. MR0246380 https://doi.org/10.1214/aoms/1177697589
[32] Tanny, D. (1977). Limit theorems for branching processes in a random environment. *Ann. Probab.* **5** 100–116. MR0426189 https://doi.org/10.1214/aop/1176995894
[33] Tanny, D. (1988). A necessary and sufficient condition for a branching process in a random environment to grow like the product of its means. *Stochastic Process. Appl.* **28** 123–139. MR0936379 https://doi.org/10.1016/0304-4149(88)90070-1
[34] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.* **11** 750–783. MR0544194 https://doi.org/10.2307/1426858
[35] Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16** 97. https://doi.org/10.1093/genetics/16.3.290

Bernoulli **32**(1), 2026, 226–248
https://doi.org/10.3150/25-BEJ1856

Bayesian model selection consistency for high-dimensional discrete graphical models

LYNDSAY ROACH[1,a], HÉLÈNE MASSAM[1,b], NANWEI WANG[2,d] and XIN GAO[1,c]

[1]*Department of Mathematics and Statistics, York University, Toronto, Canada,* [a]*lyndsay.roach@gmail.com,* [b]*massamh@mathstat.yorku.ca,* [c]*xingao@yorku.ca*
[2]*Department of Mathematics and Statistics, University of New Brunswick, Fredericton, Canada,* [d]*nanwei.wang@unb.ca*

The Bayes factor is a popular method of model selection that compares the posterior probabilities of two competing models. Consider data given in the form of a contingency table where N objects are classified according to q random variables and the conditional independence structure of these random variables are represented by a discrete graphical model. We assume the cell counts follow a multinomial distribution with a hyper Dirichlet prior distribution imposed on the cell probability parameters. We examine the behaviour of the Bayes factor when the dimension increases to infinity with the sample size. Our main result is proving strong model selection consistency for increasing dimension both when the true graph is decomposable and when the true graph is non-decomposable. When the true graph is non-decomposable, we prove that the Bayes factor selects a minimal triangulation of the true graph with the least fill-in edges.

Keywords: Bayesian; consistency; decomposable; graph selection; hyper Dirichlet; marginal likelihood

References

Batir, N. (2008). Inequalities for the gamma gunction. *Arch. Math.* **91** 554–563. MR2465874 https://doi.org/10.1007/s00013-008-2856-9

Bhattacharya, A. and Pati, D. (2020). Nonasymptotic laplace approximation under model misspecification. arXiv preprint. arXiv:2005.07844. https://doi.org/10.48550/arXiv.2005.07844

Cao, X., Khare, K. and Ghosh, M. (2019). Posterior graph selection and estimation consistency for high–dimensional Bayesian DAG models. *Ann. Statist.* **47** 319–348. MR3909935 https://doi.org/10.1214/18-AOS1689

Cao, X., Khare, K. and Ghosh, M. (2020). Consistent Bayesian sparsity selection for high–dimensional Gaussian DAG models with multiplicative and beta-mixture. *J. Multivariate Anal.* **179** 104628. MR4103343 https://doi.org/10.1016/j.jmva.2020.104628

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759–771. MR2443189 https://doi.org/10.1093/biomet/asn034

Darroch, J.N. and Speed, T.P. (1983). Additive and multiplicative models and interactions. *Ann. Statist.* **11** 724–738. MR0707924 https://doi.org/10.1214/aos/1176346240

Dawid, A.P. and Lauritzen, S.L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267 https://doi.org/10.1214/aos/1176349260

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238 https://doi.org/10.1214/aos/1176344611

Fitch, A.M., Jones, M.B. and Massam, H. (2014). The performance of covariance selection methods that consider decomposable models only. *Bayesian Anal.* **9** 659–684. MR3256059 https://doi.org/10.1214/14-BA874

Gao, X. and Song, P.X.K. (2010). Composite likelihood Bayesian information criteria for model selection in high–dimensional data. *J. Amer. Statist. Assoc.* **105** 1531–1540. MR2796569 https://doi.org/10.1198/jasa.2010.tm09414

Gao, X., Pu, D.Q., Wu, Y. and Xu, H. (2012). Tuning parameter selection for penalized likelihood estimation of Gaussian graphical model. *Statist. Sinica* **22** 1123–1146. MR2987486 https://doi.org/10.5705/ss.2009.210

1350-7265

Heggernes, P. (2006). Minimal triangulations of graphs: A survey. *Discrete Math.* **306** 297–317. MR2204109 https://doi.org/10.1016/j.disc.2005.12.003

Lauritzen, S.L. (1996). *Graph. Models*. *Oxford Science Publications*. MR1419991

Lee, K., Lee, J. and Lin, L. (2019). Minimax posterior convergence rates and model selection consistency in high–dimensional DAG models based on sparse Cholesky factors. *Ann. Statist.* **47** 3413–3437. MR4025747 https://doi.org/10.1214/18-AOS1783

Letac, G. and Massam, H. (2012). Bayes factors and the geometry of discrete hierarchical loglinear models. *Ann. Statist.* **40** 1–30. MR2985936 https://doi.org/10.1214/12-AOS974

Massam, H., Liu, J. and Dobra, A. (2009). A conjugate prior for discrete hierarchical log–linear models. *Ann. Statist.* **37** 3431–3467. MR2549565 https://doi.org/10.1214/08-AOS669

Meinshausen, N. and Bühlmann, P. (2006). High–dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

Niu, Y., Debdeep, P. and Mallick, B.K. (2021). Bayesian graph selection consistency under model misspecification. *Bernoulli* **27** 637–672. MR4177384 https://doi.org/10.3150/20-BEJ1253

Raskutti, G., Yu, B., Wainwright, M.J. and Ravikumar, P.K. (2009). Model selection in Gaussian graphical models: High–dimensional consistency of ℓ_1–regularized MLE. In *Advances in Neural Information Processing Systems* 1329–1336.

Ravikumar, P., Wainwright, M.J. and Lafferty, J.D. (2010). High–dimensional Ising model selection using ℓ_1–regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691

Roach, L., Massam, H., Wang, N. and Gao, X. (2026). Supplement to "Bayesian model selection consistency for high-dimensional discrete graphical models." https://doi.org/10.3150/25-BEJ1856SUPP

Rose, D.J., Tarjan, R.E. and Lueker, G.S. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM J. Comput.* **5** 266–283. MR0408312 https://doi.org/10.1137/0205021

Wainwright, M.J., Ravikumar, P. and Lafferty, J.D. (2007). High–dimensional graphical model selection using ℓ_1–regularized logistic regression. In *Adv. Neural Inf. Process. Syst.* (B. Schölkopf, J. Platt and T. Hoffman, eds.) **19** 1465–1472. https://doi.org/10.7551/mitpress/7503.003.0188

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

Bernoulli **32**(1), 2026, 249–273
https://doi.org/10.3150/25-BEJ1857

Large deviations for fully local monotone stochastic partial differential equations driven by gradient-dependent noise

TIANYI PAN[a], SHIJIE SHANG[b], JIANLIANG ZHAI[c] and TUSHENG ZHANG[d]

School of Mathematical Sciences, University of Science and Technology of China, Hefei, China,
[a]*pty0512@mail.ustc.edu.cn,* [b]*sjshang@ustc.edu.cn,* [c]*zhaijl@ustc.edu.cn,* [d]*tusheng.zhang@manchester.ac.uk*

Consider stochastic partial differential equations (SPDEs) with fully local monotone coefficients in a Gelfand triple $V \subseteq H \subseteq V^*$,

$$\begin{cases} du_t^{\varepsilon} = A(t, u_t^{\varepsilon})dt + \varepsilon B(t, u_t^{\varepsilon})dW_t,\ t \in (0,T], \\ u_0^{\varepsilon} = x \in H, \end{cases}$$

where

$$A : [0,T] \times V \to V^*, \quad B : [0,T] \times V \to L_2(U,H)$$

are measurable maps, $L_2(U,H)$ is the space of Hilbert-Schmidt operators from U to H and W is a U-cylindrical Wiener process. In this paper, we establish a small noise large deviation principle (LDP) for the solutions $\{u^{\varepsilon}\}_{\varepsilon>0}$ of the above SPDEs. The main contribution of this paper is the much broader generality of our framework than that of the existing results. In particular, the diffusion coefficient $B(t,\cdot)$ may depend on the gradient of the solutions, which is of great interest in the field of SPDEs, but there are few existing results on the topic of LDP. The broader scope of the fully local monotone setting leads us to use different strategies and techniques. A combination of pseudomonotone technique and compactness arguments plays a crucial role in the whole paper. Our framework is very general to include many interesting models that could not be covered by existing work, including stochastic quasilinear SPDEs, stochastic convection diffusion equation, stochastic 2D liquid crystal equation, stochastic p-Laplace equation with gradient-dependent noise, stochastic 2D Navier-Stokes equation with gradient-dependent noise, etc.

Keywords: Fully local monotone coefficient; gradient-dependent noise; large deviation principle; stochastic partial differential equations; transport noise; variational approach; weak convergence method

References

[1] Agresti, A. and Veraar, M. (2022). Nonlinear parabolic stochastic evolution equations in critical spaces part I: Stochastic maximal regularity and local existence. *Nonlinearity* **35** 4100–4210. MR4459102 https://doi.org/10.1088/1361-6544/abd613

[2] Agresti, A. and Veraar, M. (2022). Nonlinear parabolic stochastic evolution equations in critical spaces part II: Blow-up criteria and instaneous regularization. *J. Evol. Equ.* **22** 1–92. MR4437443 https://doi.org/10.1007/s00028-022-00786-7

[3] Agresti, A. and Veraar, M. (2024). The critical variational setting for stochastic evolution equations. *Probab. Theory Related Fields* **188** 957–1015. MR4716343 https://doi.org/10.1007/s00440-023-01249-x

[4] Brzeźniak, Z., Capiński, M. and Flandoli, F. (1992). Stochastic Navier-Stokes equations with multiplicative noise. *Stoch. Anal. Appl.* **10** 523–532. MR1185046 https://doi.org/10.1080/07362999208809288

[5] Brzeźniak, Z., Flandoli, F. and Maurelli, M. (2016). Existence and uniqueness for stochastic 2D Euler flows with bounded vorticity. *Arch. Ration. Mech. Anal.* **221** 107–142. MR3483892 https://doi.org/10.1007/s00205-015-0957-8

[6] Brzeźniak, Z., Liu, W. and Zhu, J. (2014). Strong solutions for SPDE with locally monotone coefficients driven by Lévy noise. *Nonlinear Anal. Real World Appl.* **17** 283–310. MR3158475 https://doi.org/10.1016/j.nonrwa.2013.12.005

[7] Brzeźniak, Z., Manna, U. and Panda, A. (2020). Large deviations for stochastic nematic liquid crystals driven by multiplicative Gaussian noise. *Potential Anal.* **53** 799–838. MR4140079 https://doi.org/10.1007/s11118-019-09788-6

[8] Budhiraja, A. and Dupuis, P. (2000). A variational representation for positive functionals of infinite dimensional Brownian motion. *Probab. Math. Statist.* **20** 39–61. MR1785237

[9] Chueshov, I. and Millet, A. (2010). Stochastic 2D hydrodynamical type systems: Well posedness and large deviations. *Appl. Math. Optim.* **61** 379–420. MR2609596 https://doi.org/10.1007/s00245-009-9091-z

[10] Damascelli, L. (1998). Comparison theorems for some quasilinear degenerate elliptic operators and applications to symmetry and monotonicity results. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **15** 493–516. MR1632933 https://doi.org/10.1016/S0294-1449(98)80032-2

[11] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. *Appl. Math. (N. Y.)* **38**. New York: Springer. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[12] Deugoué, G. and Medjo, T. (2020). Large deviation for a 2D Cahn-Hilliard-Navier-Stokes model under random influences. *J. Math. Anal. Appl.* **486** 1–34. MR4054511 https://doi.org/10.1016/j.jmaa.2020.123863

[13] Deugoué, G. and Medjo, T. (2021). Large deviation for a 2D Allen-Cahn-Navier-Stokes model under random influences. *Asymptot. Anal.* **123** 41–78. MR4255997 https://doi.org/10.3233/asy-201625

[14] Dong, Z., Zhang, R. and Zhang, T. (2020). Large deviations for quasilinear parabolic stochastic partial differential equations. *Potential Anal.* **53** 183–202. MR4117985 https://doi.org/10.1007/s11118-019-09763-1

[15] Filippov, A.F. (1988). *Differential Equations with Discontinuous Righthand Sides*. *Math. Appl. (Soviet Ser.)* **18**. Dordrecht: Kluwer Academic Publishers Group. MR1028776 https://doi.org/10.1007/978-94-015-7793-9

[16] Flandoli, F., Galeati, L. and Luo, D. (2021). Delayed blow-up by transport noise. *Comm. Partial Differential Equations* **46** 1757–1788. MR4304694 https://doi.org/10.1080/03605302.2021.1893748

[17] Flandoli, F., Gubinelli, M. and Priola, E. (2010). Well-posedness of the transport equation by stochastic perturbation. *Invent. Math.* **180** 1–53. MR2593276 https://doi.org/10.1007/s00222-009-0224-4

[18] Flandoli, F. and Luo, D. (2021). High mode transport noise improves vorticity blow-up control in 3D Navier-Stokes equations. *Probab. Theory Related Fields* **180** 309–363. MR4265023 https://doi.org/10.1007/s00440-021-01037-5

[19] Galeati, L. and Luo, D. (2024). LDP and CLT for SPDEs with transpot noise. *Stoch. Partial Differ. Equ. Anal. Comput.* **12** 736–793. MR4709553 https://doi.org/10.1007/s40072-023-00292-y

[20] Gyöngy, I. (1982). On stochastic equations with respect to semimartingales III. *Stochastics* **7** 231–254. MR0674448 https://doi.org/10.1080/17442508208833220

[21] Gyöngy, I. and Krylov, N. (2003). On the splitting-up method and stochastic partial differential equations. *Ann. Probab.* **31** 564–591. MR1964941 https://doi.org/10.1214/aop/1048516528

[22] Holm, D. (2015). Variational principles for stochastic fluid dynamics. *Proc. A* **471** 1–19. MR3325187 https://doi.org/10.1098/rspa.2014.0963

[23] Jakubowski, A. (1986). On the Skorokhod topology. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 263–285. MR0871083

[24] Krylov, N.V. and Rozovskiĭ, B.L. (1979). Stochastic evolution equations. *Current Problems in Mathematics* **14** 71–147. MR0570795

[25] Kumar, A. and Mohan, M.T. (2024). Well-posedness of a class of stochastic partial differential equations with fully monotone coefficients perturbed by Lévy noise. *Anal. Math. Phys.* **14** 1–59. MR4732962 https://doi.org/10.1007/s13324-024-00898-y

[26] Lang, O. and Crisan, D. (2023). Well-posedness for a stochastic 2D Euler equation with transport noise. *Stoch. PDE: Anal. Comp.* **11** 433–480. MR4588615 https://doi.org/10.1007/s40072-021-00233-7

[27] Lin, F. and Liu, C. (1995). Nonparabolic dissipative systems modeling the flow of liquid crystals. *Comm. Pure Appl. Math.* **48** 501–537. MR1329830 https://doi.org/10.1002/cpa.3160480503

[28] Liu, W. (2010). Large deviations for stochastic evolution equations with small multiplicative noise. *Appl. Math. Optim.* **61** 27–56. MR2575313 https://doi.org/10.1007/s00245-009-9072-2

[29] Liu, W. (2013). Well-posedness of stochastic partial differential equations with Lyapunov condition. *J. Differ. Equ.* **255** 572–592. MR3053478 https://doi.org/10.1016/j.jde.2013.04.021

[30] Liu, W. and Röckner, M. (2010). SPDE in Hilbert space with locally monotone coefficients. *J. Funct. Anal.* **259** 2902–2922. MR2719279 https://doi.org/10.1016/j.jfa.2010.05.012

[31] Liu, W. and Röckner, M. (2015). *Stochastic Partial Differential Equations: An Introduction. Univertext*. Cham: Springer. MR3410409 https://doi.org/10.1007/978-3-319-22354-4

[32] Liu, W., Tao, C. and Zhu, J. (2020). Large deviation principle for a class of SPDE with locally monotone coefficients. *Sci. China Math.* **63** 1181–1202. MR4119552 https://doi.org/10.1007/s11425-018-9440-3

[33] Manna, U., Sritharan, S.S. and Sundar, P. (2009). Large deviations for the stochastic shell model of turbulence. *Nonlinear Differential Equations Appl.* **16** 493–521. MR2525514 https://doi.org/10.1007/s00030-009-0023-z

[34] Matoussi, A., Sabbagh, W. and Zhang, T. (2021). Large deviation principles of obstacle problems for quasilinear stochastic PDEs. *Appl. Math. Optim.* **83** 849–879. MR4239802 https://doi.org/10.1007/s00245-019-09570-5

[35] Medjo, T. (2022). Large deviation principles for a 2D stochastic Allen-Cahn-Navier-Stokes driven by jump noise. *Stoch. Dyn.* **22** 1–36. MR4439603 https://doi.org/10.1142/S0219493722500058

[36] Mikulevicius, R. and Rozovskiĭ, B.L. (2004). Stochastic Navier-Stokes equations for turbulent flows. *SIAM J. Math. Anal.* **35** 1250–1310. MR2050201 https://doi.org/10.1137/S0036141002409167

[37] Mikulevicius, R. and Rozovskiĭ, B.L. (2005). Global L_2-solutions of stochastic Navier–Stokes equations. *Ann. Probab.* **33** 137–176. MR2118862 https://doi.org/10.1214/009117904000000630

[38] Neelima and Šiška, D. (2020). Coercivity condition for higher moment a priori estimates for nonlinear SPDEs and existence of a solution under local monotonicity. *Stochastics* **92** 684–715. MR4137060 https://doi.org/10.1080/17442508.2019.1650043

[39] Nguyen, P., Tawri, K. and Temam, R. (2021). Nonlinear stochastic parabolic partial differential equations with a monotone operator of the Ladyzenskaya-Smagorinsky type, driven by a Lévy noise. *J. Funct. Anal.* **281** 109–157. MR4280274 https://doi.org/10.1016/j.jfa.2021.109157

[40] Pardoux, E. (1975). Equations aux dérivées partielles stochastiques non linéairnes monotones: étude de solutions fortes de type Itô. Université de Paris-Sud. Thèse. MR0651582

[41] Pardoux, E. (1979). Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3** 127–167. MR0553909 https://doi.org/10.1080/17442507908833142

[42] Peng, X., Yang, J. and Zhai, J. (2022). Well-posedness of stochastic 2D hydrodynamics type systems with multiplicative Lévy noises. *Electron. J. Probab.* **27** 1–31. MR4416127 https://doi.org/10.1214/22-ejp779

[43] Qiu, Z. and Wang, H. (2020). Large deviation principle for the 2D stochastic Cahn-Hilliard-Navier-Stokes equations. *Z. Angew. Math. Phys.* **71** 1–29. MR4102834 https://doi.org/10.1007/s00033-020-01312-w

[44] Ren, J. and Zhang, X. (2008). Freidlin–Wentzell's large deviations for stochastic evolution equations. *J. Funct. Anal.* **254** 3148–3172. MR2418622 https://doi.org/10.1016/j.jfa.2008.02.010

[45] Röckner, M., Shang, S. and Zhang, T. (2024). Well-posedness of stochastic partial differential equations with fully local monotone coefficients. *Math. Ann.* **390** 3419–3469. MR4803456 https://doi.org/10.1007/s00208-024-02836-6

[46] Simon, J. (1987). Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl.* **146** 65–96. MR0916688 https://doi.org/10.1007/BF01762360

[47] Slavík, J. (2022). Large and moderate deviations principles and central limit theorem for the stochastic 3D primitive equations with gradient-dependent noise. *J. Theoret. Probab.* **35** 1736–1781. MR4488557 https://doi.org/10.1007/s10959-021-01125-1

[48] Sritharan, S. and Sundar, P. (2006). Large deviations for the two-dimensional Navier-Stokes equations with multiplicative noise. *Stochastic Process. Appl.* **116** 1636–1659. MR2269220 https://doi.org/10.1016/j.spa.2006.04.001

[49] Theewis, E. and Veraar, M. (2024). Large deviations for stochastic evolution equations in the critical variational setting. arXiv:2402.16622.

[50] Vallet, G. and Zimmermann, A. (2019). Well-posedness for a pseudomonotone evolution problem with multiplicative noise. *J. Evol. Equ.* **19** 153–202. MR3918520 https://doi.org/10.1007/s00028-018-0472-0

[51] Xiong, J. and Zhai, J. (2018). Large deviations for locally monotone stochastic partial differential equations driven by Lévy noise. *Bernoulli* **24** 2842–2874. MR3779704 https://doi.org/10.3150/17-BEJ947

[52] Zeidler, E. (1990). *Nonlinear Functional Analysis and Its Applications II/B*. New York: Springer. MR1033498 https://doi.org/10.1007/978-1-4612-0985-0

[53] Zhang, R. and Zhou, G. (2018). Large deviations for nematic liquid crystals driven by pure jump noise. *Math. Methods Appl. Sci.* **41** 5552–5581. MR3845928 https://doi.org/10.1002/mma.5097

Bernoulli **32**(1), 2026, 274–298
https://doi.org/10.3150/25-BEJ1858

A procedure for multiple testing of partial conjunction hypotheses based on a hazard rate inequality

THORSTEN DICKHAUS[1,a], RUTH HELLER[2,c], ANH-TUAN HOANG[1,b] and YOSEF RINOTT[3,d]

[1] *Institute for Statistics, University of Bremen, Bremen, Germany,* [a]*dickhaus@uni-bremen.de,* [b]*anhtuan.hoang@uni-bremen.de*
[2] *Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel,* [c]*ruheller@post.tau.ac.il*
[3] *Department of Statistics, Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem, Israel,* [d]*yosef.rinott@mail.huji.ac.il*

The partial conjunction null hypothesis is tested in order to discover a signal that is present in multiple studies. The standard approach of carrying out a multiple test procedure on the partial conjunction (PC) p-values can be extremely conservative. We suggest alleviating this conservativeness by eliminating many of the large PC p-values prior to the application of a multiple test procedure. This leads to the following two step procedure: first, select the set with PC p-values below a selection threshold; second, within the selected set only, apply a family-wise error rate or false discovery rate controlling procedure on the conditional PC p-values. The conditional PC p-values are valid if the null p-values are uniform and the combining method is Fisher. The proof of their validity is based on a novel inequality in hazard rate order of partial sums of order statistics which may be of independent interest. We also provide the conditions for which the false discovery rate controlling procedures considered will be below the nominal level. We demonstrate the potential usefulness of our new method, `CoFilter` (conditional testing after filtering), for analyzing multiple genome-wide association studies of Crohn's disease.

Keywords: Composite hypotheses; false discovery rate; hazard rate order; intersection hypotheses; meta analysis

References

Benjamini, Y. and Heller, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B, Methodol.* **57** 289–300. MR1325392 https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., Krieger, A.M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438 https://doi.org/10.1093/biomet/93.3.491

Bogomolov, M. and Heller, R. (2023). Replicability across multiple studies. *Statist. Sci.* **38** 602–620. MR4665028 https://doi.org/10.1214/23-sts892

Chi, Z. (2007). On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **35** 1409–1431. MR2351091 https://doi.org/10.1214/009053607000000037

Dickhaus, T. (2013). Randomized p-values for multiple testing of composite null hypotheses. *J. Statist. Plann. Inference* **143** 1968–1979. MR3095086 https://doi.org/10.1016/j.jspi.2013.06.011

Dickhaus, T. (2014). *Simultaneous Statistical Inference with Applications in the Life Sciences*. Berlin, Heidelberg: Springer. MR3184277 https://doi.org/10.1007/978-3-642-45182-9

Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist.* **35**. 1432–1455. MR2351092 https://doi.org/10.1214/009053607000000046

1350-7265

Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biom. J.* **43** 985–1005. MR1878272 https://doi.org/10.1002/1521-4036(200112)43:8<985::AID-BIMJ985>3.0.CO;2-4

Fisher, R.A. (1934). *Statistical Methods for Research Workers*, 5th ed. Edinburgh and London: Oliver and Boyd.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R. et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42** 1118–1125. https://doi.org/10.1038/ng.717

Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517. MR1924303 https://doi.org/10.1111/1467-9868.00347

Gontscharuk, V. (2010). Asymptotic and Exact Results on FWER and FDR in Multiple Hypotheses Testing PhD thesis, Heinrich-Heine-Universität Düsseldorf. https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=16990.

Heller, R. and Yekutieli, D. (2014). Replicability analysis for genome-wide association studies. *Ann. Appl. Stat.* **8** 481–498. MR3191999 https://doi.org/10.1214/13-AOAS697

Hoang, A.-T. and Dickhaus, T. (2022). Combining independent p-values in replicability analysis: A comparative study. *J. Stat. Comput. Simul.* **92** 2184–2204. MR4437521 https://doi.org/10.1080/00949655.2021.2022678

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

Karmakar, B. and Small, D.S. (2020). Assessment of the extent of corroboration of an elaborate theory of a causal hypothesis using partial conjunctions of evidence factors. *Ann. Statist.* **48** 3283–3311. MR4185809 https://doi.org/10.1214/19-AOS1929

Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43. MR0054208 https://doi.org/10.1214/aoms/1177729080

Li, S., Sesia, M., Romano, Y., Candes, E. and Sabatti, C. (2021). Searching for consistent associations with a multi-environment knockoff filter. Preprint. Available at https://doi.org/10.48550/arXiv.2106.04118

Liang, B., Zhang, L. and Janson, L. (2023). Powerful partial conjunction hypothesis testing via conditioning. Preprint. Available at https://doi.org/10.48550/arXiv.2212.11304

Nagaraja, H.N. (2006). Order statistics from independent exponential random variables and the sum of the top order statistics. In *Advances in Distribution Theory, Order Statistics, and Inference* 173–185. Boston, MA: Birkhäuser Boston. MR2226221 https://doi.org/10.1007/0-8176-4487-3_11

Parzen, E. (1960). *Modern Probability Theory and Its Applications. A Wiley Publication in Mathematical Statistics*. New York and London: John Wiley & Sons, Inc. MR0112166

Schweder, T. and Spjøtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69** 493–502. https://doi.org/10.1093/biomet/69.3.493

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders. Springer Series in Statistics*. New York, NY: Springer. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 https://doi.org/10.1093/biomet/73.3.751

Storey, J.D., Taylor, J.E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766 https://doi.org/10.1111/j.1467-9868.2004.00439.x

Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. and Williams, R.M. Jr (1949). *The American Soldier: Adjustment During Army Life. (Studies in Social Psychology in World War II)* **1**. Princeton, NJ: Princeton Univ. Press.

Sun, W. and Wei, Z. (2011). Multiple testing for pattern identification, with applications to microarray time-course experiments. *J. Amer. Statist. Assoc.* **106** 73–88. MR2816703 https://doi.org/10.1198/jasa.2011.ap09587

Wang, J., Gui, L., Su, W.J., Sabatti, C. and Owen, A.B. (2020). adaFilter. R package. Available at https://github.com/jingshuw/adaFilter.

Wang, J., Gui, L., Su, W.J., Sabatti, C. and Owen, A.B. (2022). Detecting multiple replicating signals using adaptive filtering procedures. *Ann. Statist.* **50** 1890–1909. MR4474476 https://doi.org/10.1214/21-aos2139

Yu, Y. (2009). Stochastic ordering of exponential family distributions and their mixtures. *J. Appl. Probab.* **46** 244–254. MR2508516 https://doi.org/10.1239/jap/1238592127

Zhao, Q., Small, D.S. and Su, W. (2019). Multiple testing when many p-values are uniformly conservative, with application to testing qualitative interaction in educational interventions. *J. Amer. Statist. Assoc.* **114** 1291–1304. MR4011780 https://doi.org/10.1080/01621459.2018.1497499

Bernoulli **32**(1), 2026, 299–324
https://doi.org/10.3150/25-BEJ1859

Minimax optimal goodness-of-fit testing with kernel Stein discrepancy

OMAR HAGRASS[1,a], BHARATH SRIPERUMBUDUR[2,b] and KRISHNAKUMAR BALASUBRAMANIAN[3,c]

[1]*Department of Operations Research & Financial Engineering, Princeton University, Princeton, NJ, USA,* [a]*oh2588@princeton.edu*

[2]*Department of Statistics, Pennsylvania State University, University Park, PA, USA,* [b]*bks18@psu.edu*

[3]*Department of Statistics, University of California, Davis, Davis, CA, USA,* [c]*kbala@ucdavis.edu*

We explore the minimax optimality of goodness-of-fit tests on general domains using the kernelized Stein discrepancy (KSD). The KSD framework offers a flexible approach for goodness-of-fit testing, avoiding strong distributional assumptions, accommodating diverse data structures beyond Euclidean spaces, and relying only on partial knowledge of the reference distribution, while maintaining computational efficiency. Although KSD is a powerful framework for goodness-of-fit testing, only the consistency of the corresponding tests has been established so far, and their statistical optimality remains largely unexplored. In this paper, we develop a general framework and an operator-theoretic representation of the KSD, encompassing many existing KSD tests in the literature, which vary depending on the domain. Building on this representation, we propose a modified discrepancy by applying the concept of spectral regularization to the KSD framework. We establish the minimax optimality of the proposed regularized test for a wide range of the smoothness parameter θ under a specific alternative space, defined over general domains, using the χ^2-divergence as the separation metric. In contrast, we demonstrate that the unregularized KSD test fails to achieve the minimax separation rate for the considered alternative space. Additionally, we introduce an adaptive test capable of achieving minimax optimality up to a logarithmic factor by adapting to unknown parameters. Through numerical experiments, we illustrate the superior performance of our proposed tests across various domains compared to their unregularized counterparts.

Keywords: Adaptivity; Bernstein's inequality; covariance operator; goodness-of-fit test; kernel Stein discrepancy; maximum mean discrepancy; minimax separation; reproducing kernel Hilbert space; spectral regularization; U-statistics

References

Arcones, M.A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586 https://doi.org/10.1214/aos/1176348650

Balakrishnan, S. and Wasserman, L. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. MR3953439 https://doi.org/10.1214/18-AOS1729

Balasubramanian, K., Li, T. and Yuan, M. (2021). On the optimality of kernel-embedding based goodness-of-fit tests. *J. Mach. Learn. Res.* **22** 1–45. MR4253694

Bauer, F., Pereverzev, S. and Rosasco, L. (2007). On regularization algorithms in learning theory. *J. Complexity* **23** 52–72. MR2297015 https://doi.org/10.1016/j.jco.2006.07.001

Berlinet, A. and Thomas-Agnan, C. (2001). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Acad. Publ. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

Bickel, P.J. and Breiman, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214. MR0682809 https://doi.org/10.1214/aop/1176993668

Boente, G., Rodriguez, D. and Manteiga, W.G. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.* **41** 259–275. MR3181143 https://doi.org/10.1111/sjos.12020

1350-7265

Burnashev, M.V. (1979). On the minimax detection of an inaccurately known signal in a white Gaussian noise background. *Theory Probab. Appl.* **24** 107–119. MR0522240 https://doi.org/10.1137/1124008

Caponnetto, A. and Vito, E.D. (2007). Optimal rates for regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8

Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. In *Proceedings of the 33rd International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **48** 2606–2615. New York, NY, USA: PMLR.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51** 392–413. MR1321305 https://doi.org/10.1006/jmva.1994.1069

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642–669. MR0083864 https://doi.org/10.1214/aoms/1177728174

Ebner, B., Henze, N. and Yukich, J.E. (2018). Multivariate goodness-of-fit on flat and curved spaces via nearest neighbor distances. *J. Multivariate Anal.* **165** 231–242. MR3768763 https://doi.org/10.1016/j.jmva.2017.12.009

Engl, H.W., Hanke, M. and Neubauer, A. (1996). *Regularization of Inverse Problems*. *Mathematics and Its Applications* **375**. Dordrecht, The Netherlands: Kluwer Acad. Publ. MR1408680

Fernandez, T., Rivera, N., Xu, W. and Gretton, A. (2020). Kernelized Stein discrepancy tests of goodness-of-fit for time-to-event data. In *Proceedings of the 37th International Conference on Machine Learning* (H.D. III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 3112–3122. PMLR. Available at arXiv:2008.08397.

Fischer, A. and Igel, C. (2012). An introduction to restricted Boltzmann machines. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (L. Alvarez, M. Mejail, L. Gomez and J. Jacobo, eds.) 14–36. Berlin: Springer. MR3075009 https://doi.org/10.1007/978-3-642-33275-3_2

Gorham, J. and Mackey, L. (2015). Measuring sample quality with Stein's method. In *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, eds.) **28**. Curran Associates Inc.

Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

Hagrass, O., Sriperumbudur, B. and Balasubramanian, K. (2026). Supplement to "Minimax optimal Goodness-of-Fit testing with Kernel Stein Discrepancy." https://doi.org/10.3150/25-BEJ1859SUPP

Hagrass, O., Sriperumbudur, B.K. and Li, B. (2024a). Spectral regularized kernel two-sample tests. *Ann. Statist.* **52** 1076–1101. MR4784070 https://doi.org/10.1214/24-aos2383

Hagrass, O., Sriperumbudur, B.K. and Li, B. (2024b). Spectral regularized kernel goodness-of-fit tests. *J. Mach. Learn. Res.* **25** 1–52. MR4829153 https://doi.org/10.48550/arXiv

Hallin, M., Mordant, G. and Segers, J. (2021). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.* **15** 1328–1371. MR4255302 https://doi.org/10.1214/21-ejs1816

Hušková, M. and Janssen, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics. *Ann. Statist.* **21** 1811–1823. MR1245770 https://doi.org/10.1214/aos/1176349399

Ingster, Y. and Suslina, I.A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models 169*. Springer Sci. Bus. Media, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

Ingster, Y.I. (1987). Minimax testing of nonparametric hypotheses on a distribution density in the L_p metrics. *Theory Probab. Appl.* **31** 333–337. MR0851000 https://doi.org/10.1137/1139029

Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.* **2** 85–114. MR1257978

Jupp, P.E. (2008). Data-driven Sobolev tests of uniformity on compact Riemannian manifolds. *Ann. Statist.* **36** 1246–1260. MR2418656 https://doi.org/10.1214/009053607000000541

Kanagawa, H., Jitkrittum, W., Mackey, L., Fukumizu, K. and Gretton, A. (2023). A kernel Stein test for comparing latent variable models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 986–1011. MR4726979 https://doi.org/10.1093/jrsssb/qkad050

Khmaladze, E.V. (1988). An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$. *Ann. Statist.* **16** 1503–1516. MR0964936 https://doi.org/10.1214/aos/1176351051

Khmaladze, E.V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829. MR1232520 https://doi.org/10.1214/aos/1176349152

Khmaladze, E.V. (2013). Note on distribution free testing for discrete distributions. *Ann. Statist.* **41** 2979–2993. MR3161454 https://doi.org/10.1214/13-AOS1176

Khmaladze, E.V. (2016). Unitary transformations, empirical processes and distribution free testing. *Bernoulli* **22** 563–588. MR3449793 https://doi.org/10.3150/14-BEJ668

Lehmann, E.L. and Romano, J.P. (2006). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer Science & Business Media. MR2135927 https://doi.org/10.1007/978-3-030-70578-7

Liu, Q., Lee, J. and Jordan, M. (2016). A kernelized Stein discrepancy for goodness-of-fit tests. In *Proceedings of the 33rd International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **48** 276–284. New York, NY, USA: PMLR.

Mardia, K.V., Holmes, D. and Kent, J.T. (1984). A goodness-of-fit test for the von Mises–Fisher distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **46** 72–78. MR0745218 https://doi.org/10.1111/j.2517-6161.1984.tb01278.x

Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069 https://doi.org/10.1214/aop/1176990746

Minh, H.Q., Niyogi, P. and Yao, Y. (2006). Mercer's theorem, feature maps, and smoothing. In *Learning Theory* (G. Lugosi and H.U. Simon, eds.). *Lecture Notes in Computer Science* **4005** 154–168. Berlin: Springer. MR2280604 https://doi.org/10.1007/11776420_14

Muandet, K., Fukumizu, K., Sriperumbudur, B. and Schölkopf, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **10** 1–141. https://doi.org/10.1561/2200000060

Qu, X., Fan, X. and Vemuri, B. (2024). Kernel Stein discrepancy on Lie groups: Theory and applications. *IEEE Trans. Inf. Theory* **70** 1–15. MR4832310 https://doi.org/10.1109/tit.2024.3468212

Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics: Functional Analysis I*. New York: Academic Press. MR0751959

Schilling, M.F. (1983a). Goodness of fit testing in $\mathbb{R}^m$ based on the weighted empirical distribution of certain nearest neighbor statistics. *Ann. Statist.* **11** 1–12. MR0684858 https://doi.org/10.1214/aos/1176346051

Schilling, M.F. (1983b). An infinite-dimensional approximation for nearest neighbor goodness of fit tests. *Ann. Statist.* **11** 13–24. MR0684859 https://doi.org/10.1214/aos/1176346052

Schrab, A., Guedj, B. and Gretton, A. (2022). KSD aggregated goodness-of-fit test. In *Adv. Neural Inf. Process. Syst* (K. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.) **35** 32624–32638. Curran Assoc., Inc. arXiv:2202.00824.

Schrab, A., Kim, I., Albert, M., Laurent, B., Guedj, B. and Gretton, A. (2023). MMD aggregated two-sample test. *J. Mach. Learn. Res.* **24** 1–81. MR4633583

Smola, A.J., Gretton, A., Song, L. and Schölkopf, B. (2007). A Hilbert space embedding for distributions. In *Algorithmic Learning Theory* (M. Hutter, R.A. Servedio and E. Takimoto, eds.) 13–31. Berlin, Germany: Springer-Verlag. MR2605631 https://doi.org/10.1007/978-3-540-75225-7_5

Sriperumbudur, B.K., Fukumizu, K. and Lanckriet, G.R.G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431

Steinwart, I. and Scovel, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. MR2914365 https://doi.org/10.1007/s00365-012-9153-3

Wahba, G. (1990). Spline models for observational data. In *CBMS-NSF Regional Conference Series in Applied Mathematics*. *Industrial and Applied Mathematics*. **59**. Society for. MR1045442 https://doi.org/10.1137/1.9781611970128

Wynne, G., Kasprzak, M.J. and Duncan, A.B. (2025). A Fourier representation of kernel Stein discrepancy with application to goodness-of-fit tests for measures on infinite dimensional Hilbert spaces. *Bernoulli*. **31** 868–893. MR4863060 https://doi.org/10.3150/23-bej1662

Xu, W. and Matsuda, T. (2020). A Stein goodness-of-fit test for directional distributions. In *Proceedings of the Twenty-Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 320–330. PMLR. arXiv:2002.06843.

Xu, W. and Matsuda, T. (2021). Interpretable Stein goodness-of-fit tests on Riemannian manifold. In *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.). *Proceedings of Machine Learning Research* **139** 11502–11513. New York, NY, USA: PMLR.

Xu, W. and Reinert, G. (2021). A Stein goodness-of-fit test for exponential random graph models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 415–423. PMLR. arXiv:2103.00580.

Xu, W. and Reinert, G. (2022). AgraSSt: Approximate graph Stein statistics for interpretable assessment of implicit graph generators. *Adv. Neural Inf. Process. Syst.* **35** 24268–24279. https://doi.org/10.48550/arXiv

Yang, J., Liu, Q., Rao, V. and Neville, J. (2018). Goodness-of-fit testing for discrete distributions via Stein discrepancy. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 5561–5570. New York, NY, USA: PMLR.

Yang, J., Rao, V. and Neville, J. (2019). A Stein–Papangelou goodness-of-fit test for point processes. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 226–235. New York, NY, USA: PMLR.

Bernoulli **32**(1), 2026, 325–349
https://doi.org/10.3150/25-BEJ1860

Smallest gaps between eigenvalues of real Gaussian matrices

PATRICK LOPATTO[a] and MATTHEW MEEKER[b]

Division of Applied Mathematics, Brown University, Providence, United States of America,
[a]*patrick_lopatto@brown.edu,* [b]*matthew_meeker@brown.edu*

We consider an $n \times n$ matrix of independent real Gaussian random variables and determine the asymptotic distribution of the smallest gaps between complex eigenvalues.

Keywords: Extreme value theory; Ginibre ensemble; random matrix

References

[1] Adhikari, K. (2018). Hole probabilities for β-ensembles and determinantal point processes in the complex plane. *Electron. J. Probab.* **23** 1–21. MR3814242 https://doi.org/10.1214/18-ejp176
[2] Adhikari, K. and Reddy, N.K. (2017). Hole probabilities for finite and infinite Ginibre ensembles. *Int. Math. Res. Not. IMRN* **2017** 6694–6730. MR3719476 https://doi.org/10.1093/imrn/rnw207
[3] Ameur, Y. (2018). Repulsion in low temperature β-ensembles. *Comm. Math. Phys.* **359** 1079–1089. MR3784540 https://doi.org/10.1007/s00220-017-3027-2
[4] Ameur, Y. and Cronvall, J. (2023). Szegő type asymptotics for the reproducing kernel in spaces of full-plane weighted polynomials. *Comm. Math. Phys.* **398** 1291–1348. MR4561803 https://doi.org/10.1007/s00220-022-04539-y
[5] Ameur, Y., Marceca, F. and Romero, J.-L. (2024). Gaussian beta ensembles: the perfect freezing transition and its characterization in terms of Beurling-Landau densities. Available at arXiv:2205.15054v2. https://doi.org/10.48550/arXiv.2205.15054
[6] Ameur, Y. and Romero, J.L. (2023). The planar low temperature Coulomb gas: Separation and equidistribution. *Rev. Mat. Iberoam.* **39** 611–648. MR4575375 https://doi.org/10.4171/rmi/1340
[7] Ameur, Y. and Troedsson, E. (2024). Remarks on the one-point density of Hele–Shaw β-ensembles. Available at arXiv:2402.13882. MR4912929 https://doi.org/10.48550/arXiv.2402.13882
[8] Ben Arous, G. and Bourgade, P. (2013). Extreme gaps between eigenvalues of random matrices. *Ann. Probab.* **41** 2648–2681. MR3112927 https://doi.org/10.1214/11-AOP710
[9] Billingsley, P. (2008). *Probability and Measure*, Third ed. Hoboken: John Wiley & Sons, Inc. MR2893652
[10] Bleher, P. and Mallison, R. Jr. (2006). Zeros of sections of exponential sums. *Int. Math. Res. Not. IMRN* **2006** 38937. MR2264712 https://doi.org/10.1155/IMRN/2006/38937
[11] Bohigas, O., Giannoni, M.-J. and Schmit, C. (1984). Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Phys. Rev. Lett.* **52** 1. MR0730191 https://doi.org/10.1103/PhysRevLett.52.1
[12] Bordenave, C. and Chafaï, D. (2012). Around the circular law. *Probab. Surv.* **9** 1–89. MR2908617 https://doi.org/10.1214/11-PS183
[13] Borodin, A. and Sinclair, C.D. (2009). The Ginibre ensemble of real random matrices and its scaling limits. *Comm. Math. Phys.* **291** 177–224. MR2530159 https://doi.org/10.1007/s00220-009-0874-5
[14] Bourgade, P. (2021). Extreme gaps between eigenvalues of Wigner matrices. *J. Eur. Math. Soc. (JEMS)* **24** 2823–2873. MR4416591 https://doi.org/10.4171/jems/1141
[15] Byun, S.-S. and Forrester, P.J. (2022). Progress on the study of the Ginibre ensembles I: GinUE. Available at arXiv:2211.16223. https://doi.org/10.48550/arXiv.2211.16223

[16] Byun, S.-S. and Forrester, P.J. (2023). Progress on the study of the Ginibre ensembles II: GinOE and GinSE. Available at arXiv:2301.05022. https://doi.org/10.48550/arXiv.2301.05022
[17] Charlier, C. (2023). Hole probabilities and balayage of measures for planar Coulomb gases. Available at arXiv:2311.15285. https://doi.org/10.48550/arXiv.2311.15285
[18] Charlier, C. (2024). Large gap asymptotics on annuli in the random normal matrix model. *Math. Ann.* **388** 3529–3587. MR4721776 https://doi.org/10.1007/s00208-023-02603-z
[19] Deift, P., Nanda, T. and Tomei, C. (1983). Ordinary differential equations and the symmetric eigenvalue problem. *SIAM J. Numer. Anal.* **20** 1–22. MR0687364 https://doi.org/10.1137/0720001
[20] Feng, R., Götze, F. and Yao, D. (2024). Smallest gaps between zeros of stationary Gaussian processes. *J. Funct. Anal.* **287** 110493. MR4746004 https://doi.org/10.1016/j.jfa.2024.110493
[21] Feng, R., Li, J. and Yao, D. (2024). Small gaps of GSE. Available at arXiv:2409.03324. https://doi.org/10.48550/arXiv.2409.03324
[22] Feng, R., Tian, G. and Wei, D. (2019). Small gaps of GOE. *Geom. Funct. Anal.* **29** 1794–1827. MR4034920 https://doi.org/10.1007/s00039-019-00520-5
[23] Feng, R. and Wei, D. (2018). Large gaps of CUE and GUE. Available at arXiv:1807.02149. https://doi.org/10.48550/arXiv.1807.02149
[24] Feng, R. and Wei, D. (2021). Small gaps of circular β-ensemble. *Ann. Probab.* **49** 997–1032. MR4255136 https://doi.org/10.1214/20-aop1468
[25] Figalli, A. and Guionnet, A. (2016). Universality in several-matrix models via approximate transport maps. *Acta Math.* **217** 115–159. MR3646880 https://doi.org/10.1007/s11511-016-0142-4
[26] Forrester, P.J. (2010). *Log-Gases and Random Matrices*. Princeton: Princeton University Press. MR2641363 https://doi.org/10.1515/9781400835416
[27] Forrester, P.J. and Mays, A. (2009). A method to calculate correlation functions for $\beta = 1$ random matrices of odd size. *J. Stat. Phys.* **134** 443–462. MR2485724 https://doi.org/10.1007/s10955-009-9684-6
[28] Forrester, P.J. and Nagao, T. (2007). Eigenvalue statistics of the real Ginibre ensemble. *Phys. Rev. Lett.* **99** 050603. https://doi.org/10.1103/PhysRevLett.99.050603
[29] Forrester, P.J. and Nagao, T. (2008). Skew orthogonal polynomials and the partly symmetric real Ginibre ensemble. *J. Phys. A* **41** 375003. MR2430570 https://doi.org/10.1088/1751-8113/41/37/375003
[30] Ge, S. (2017). The Eigenvalue Spacing of IID Random Matrices and Related Least Singular Value Results PhD thesis, University of California, Los Angeles. MR3706025
[31] Gebert, M. and Poplavskyi, M. (2019). On pure complex spectrum for truncations of random orthogonal matrices and Kac polynomials. Available at arXiv:1905.03154. https://doi.org/10.48550/arXiv.1905.03154
[32] Goel, A., Lopatto, P. and Xie, X. (2024). Central limit theorem for the complex eigenvalues of Gaussian random matrices. *Electron. Commun. Probab.* **29** 1–13. MR4735543 https://doi.org/10.1214/24-ECP585
[33] Horn, R. and Johnson, C. (2012). *Matrix Analysis*. Cambridge: Cambridge University Press. MR2978290 https://doi.org/10.1017/CBO9780511810817
[34] Kallenberg, O. (2017). *Random Measures, Theory and Applications*. Cham: Springer. MR3642325 https://doi.org/10.1007/978-3-319-41598-7
[35] Kopel, P. (2015). Linear statistics of non-Hermitian matrices matching the real or complex Ginibre ensemble to four moments. Available at arXiv:1510.02987. https://doi.org/10.48550/arXiv.1510.02987
[36] Kriecherbauer, T., Kuijlaars, A.B.J., McLaughlin, K.D.T.R. and Miller, P.D. (2008). Locating the zeros of partial sums of e^z with Riemann–Hilbert methods. In *Integrable Systems and Random Matrices: In Honor of Percy Deift: Conference on Integrable Systems, Random Matrices, and Applications in Honor of Percy Deift's 60th Birthday, May 22–26, 2006, Courant Institute of Mathematical Sciences, New York University, New York* **458** 183. Providence: American Mathematical Society. MR3971152
[37] Landon, B., Lopatto, P. and Marcinek, J. (2020). Comparison theorem for some extremal eigenvalue statistics. *Ann. Probab.* **48** 2894–2919. MR4164456 https://doi.org/10.1214/20-AOP1439
[38] Lev, N. and Ortega-Cerdà, J. (2016). Equidistribution estimates for Fekete points on complex manifolds. *J. Eur. Math. Soc.* **18** 425–464. MR3459956 https://doi.org/10.4171/JEMS/594
[39] Lopatto, P. and Luh, K. (2021). Tail bounds for gaps between eigenvalues of sparse random matrices. *Electron. J. Probab.* **26** 1–26. MR4343561 https://doi.org/10.1214/21-ejp669
[40] Lopatto, P. and Meeker, M. (2026). Supplement to "Smallest gaps between eigenvalues of real Gaussian matrices." https://doi.org/10.3150/25-BEJ1860SUPP

[41] Luh, K. and O'Rourke, S. (2021). Eigenvectors and controllability of non-Hermitian random matrices and directed graphs. *Electron. J. Probab.* **26** 1–43. MR4235467 https://doi.org/10.1214/21-EJP588
[42] Marceca, F. and Romero, J.L. (2024). Improved Discrepancy for the planar Coulomb gas at low temperatures. Available at arXiv:2212.14821v3. https://doi.org/10.48550/arXiv.2212.14821
[43] Mays, A. (2012). A geometrical triumvirate of real random matrices. Available at arXiv:1202.1218. https://doi.org/10.48550/arXiv.1202.1218
[44] Nguyen, H., Tao, T. and Vu, V. (2017). Random matrices: Tail bounds for gaps between eigenvalues. *Probab. Theory Related Fields* **167** 777–816. MR3627428 https://doi.org/10.1007/s00440-016-0693-5
[45] Odlyzko, A.M. (1987). On the distribution of spacings between zeros of the zeta function. *Math. Comp.* **48** 273–308. MR0866115 https://doi.org/10.2307/2007890
[46] Otto, M. (2024). Couplings and Poisson approximation for stabilizing functionals of determinantal point processes. Available at arXiv:2402.08588. https://doi.org/10.48550/arXiv.2402.08588
[47] Shi, D. and Jiang, Y. (2012). Smallest gaps between eigenvalues of random matrices with complex Ginibre, Wishart and universal unitary ensembles. Available at arXiv:1207.4240. https://doi.org/10.48550/arXiv.1207.4240
[48] Sinclair, C.D. (2009). Correlation functions for $\beta = 1$ ensembles of matrices of odd size. *J. Stat. Phys.* **136** 17–33. MR2525224 https://doi.org/10.1007/s10955-009-9771-8
[49] Sommers, H.-J. and Wieczorek, W. (2008). General eigenvalue correlations for the real Ginibre ensemble. *J. Phys. A* **41** 405003. MR2439268 https://doi.org/10.1088/1751-8113/41/40/405003
[50] Soshnikov, A. (1998). Level spacings distribution for large random matrices: Gaussian fluctuations. *Ann. Math.* **148** 573–617. MR1668559 https://doi.org/10.2307/121004
[51] Soshnikov, A. (2000). Determinantal random point fields. *Russian Math. Surveys* **55** 923–975. MR1799012 https://doi.org/10.1070/rm2000v055n05ABEH000321
[52] Soshnikov, A. (2005). Statistics of extreme spacing in determinantal random point processes. *Mosc. Math. J.* **5** 705–719. MR2241818 https://doi.org/10.17323/1609-4514-2005-5-3-705-719
[53] Thoma, E. (2023). Overcrowding and separation estimates for the Coulomb gas. *Comm. Pure Appl. Math.* **77** 3227–3276. MR4764743 https://doi.org/10.1002/cpa.22188
[54] Vinson, J.P. (2001). Closest Spacing of Eigenvalues PhD dissertation, Princeton University. Available at arXiv:1111.2743. https://doi.org/10.48550/arXiv.1111.2743

Bernoulli **32**(1), 2026, 350–369
https://doi.org/10.3150/25-BEJ1861

Non-linear wavelet density estimation on the real line

MATHIEU SART[a]

Université Jean Monnet, CNRS, Ecole Centrale de Lyon, INSA Lyon, Universite Claude Bernard Lyon 1, ICJ UMR5208, 42023 Saint-Etienne, France, [a]*mathieu.sart@univ-st-etienne.fr*

We investigate the problem of optimal density estimation on the real line $\mathbb{R}$ under $\mathbb{L}^1$ loss. We carry out a new way to select the important coefficients in some wavelet expansions. We study the resulting estimator when the density is smooth with dominated tails. This assumption is very mild and allows in particular to deal with singularities, spatially inhomogeneous smoothness, and fat-tailed distributions.

Keywords: Besov classes; minimax rates; non-parametric estimation; wavelet methods

References

[1] Akakpo, N. (2017). Multivariate intensity estimation via hyperbolic wavelet selection. *J. Multivariate Anal.* **161** 32–57. MR3698113 https://doi.org/10.1016/j.jmva.2017.07.005
[2] Autin, F. (2006). Maxiset for density estimation on R. *Math. Methods Statist.* **15** 123–145. MR2256472
[3] Baraud, Y. and Birgé, L. (2009). Estimating the intensity of a random measure by histogram type estimators. *Probab. Theory Related Fields* **143** 239–284. MR2449129 https://doi.org/10.1007/s00440-007-0126-6
[4] Birgé, L. (2006). Model selection via testing: An alternative to (penalized) maximum likelihood estimators. *Ann. Inst. Henri Poincaré B, Probab. Stat.* **42** 273–325. MR2219712 https://doi.org/10.1016/j.anihpb.2005.04.004
[5] Cao, K. and Liu, Y. (2020). Uncompactly supported density estimation with L^1 risk. *Comm. Pure Appl. Math.* **19** 4007–4022. MR4112009 https://doi.org/10.3934/cpaa.2020177
[6] Chicken, E. and Cai, T.T. (2005). Block thresholding for density estimation: Local and global adaptivity. *J. Multivariate Anal.* **95** 76–106. MR2164124 https://doi.org/10.1016/j.jmva.2004.07.003
[7] Cohen, A., Daubechies, I. and Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **45** 485–560. MR1162365 https://doi.org/10.1002/cpa.3160450502
[8] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3** 215–228. MR1400080 https://doi.org/10.1006/acha.1996.0017
[9] Delyon, B. and Juditsky, A. (1997). On the computation of wavelet coefficients. *J. Approx. Theory* **88** 47–79. MR1426459 https://doi.org/10.1006/jath.1996.3008
[10] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation. Grundlehren der mathematischen Wissenschaften* **303**. Springer Science & Business Media. ISBN 9780387506272. MR1261635
[11] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974 https://doi.org/10.1214/aos/1032894451
[12] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 https://doi.org/10.1214/11-AOS883
[13] Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory Related Fields* **159** 479–543. MR3230001 https://doi.org/10.1007/s00440-013-0512-1
[14] Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–942. MR1635418 https://doi.org/10.1214/aos/1024691082
[15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications 129*. Springer Science & Business Media. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

[16] Ibragimov, I.A. and Khas'minskii, R.Z. (1984). More on the estimation of distribution densities. *J. Sov. Math.* **25** 1155–1165. https://doi.org/10.1007/BF01084794
[17] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10** 187–220. MR2046772 https://doi.org/10.3150/bj/1082380217
[18] Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). Lp adaptive density estimation. *Bernoulli* **2** 229–247. MR1416864 https://doi.org/10.2307/3318521
[19] Lepski, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129 https://doi.org/10.1214/13-AOS1109
[20] Lepski, O. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43** 1178–1242. MR3346701 https://doi.org/10.1214/14-AOS1306
[21] Lepski, O. and Willer, T. (2019). Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Statist.* **47** 233–287. MR3909933 https://doi.org/10.1214/18-AOS1687
[22] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2011). Adaptive density estimation: A curse of support? *J. Statist. Plann. Inference* **141** 115–139. MR2719482 https://doi.org/10.1016/j.jspi.2010.05.017
[23] Sart, M. (2021). Minimax bounds for Besov classes in density estimation. *Electron. J. Stat.* **15** 3184–3216. MR4280167 https://doi.org/10.1214/21-ejs1856
[24] Sart, M. (2021). Estimating a density, a hazard rate, and a transition intensity via the ρ-estimation method. *Ann. Inst. Henri Poincaré B, Probab. Stat.* **57** 195–249. MR4255173 https://doi.org/10.1214/20-aihp1076
[25] Sart, M. (2026). Supplement to "Non-linear wavelet density estimation on the real line." https://doi.org/10.3150/25-BEJ1861SUPP

Bernoulli **32**(1), 2026, 370–392
https://doi.org/10.3150/25-BEJ1862

Fourier analysis of spatial point processes

JUNHO YANG[1,a] and YONGTAO GUAN[2,b]

[1] *Institute of Statistical Science, Academia Sinica, Taipei, Taiwan,* [a] *junhoyang@stat.sinica.edu.tw*
[2] *School of Data Science, The Chinese University of Hong Kong, Shenzhen, China,* [b] *guanyongtao@cuhk.edu.cn*

In this article, we develop comprehensive frequency domain methods for estimating and inferring the second-order structure of spatial point processes. The main element here is on utilizing the discrete Fourier transform (DFT) of the point pattern and its tapered counterpart. Under second-order stationarity, we show that both the DFTs and the tapered DFTs are asymptotically jointly independent Gaussian even when the DFTs share the same limiting frequencies. Based on these results, we establish an α-mixing central limit theorem for a statistic formulated as a quadratic form of the tapered DFT. As applications, we derive the asymptotic distribution of the kernel spectral density estimator and establish a frequency domain inferential method for parametric stationary point processes. For the latter, the resulting model parameter estimator is computationally tractable and yields meaningful interpretations even in the case of model misspecification. We investigate the finite sample performance of our estimator through simulations, considering scenarios of both correctly specified and misspecified models.

Keywords: Bartlett's spectrum; clustering and repulsive point processes; data tapering; discrete Fourier transform; stationary point processes; Whittle likelihood

References

[1] Baddeley, A.J., Møller, J. and Waagepetersen, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **54** 329–350. MR1804002 https://doi.org/10.1111/1467-9574.00144
[2] Bandyopadhyay, S. and Lahiri, S.N. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhya A* **71** 221–259. MR2639292
[3] Bartlett, M.S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika* **51** 299–311. MR0175254 https://doi.org/10.2307/2334136
[4] Biscio, C.A.N. and Lavancier, F. (2016). Brillinger mixing of determinantal point processes and statistical applications. *Electron. J. Stat.* **10** 582–607. MR3471989 https://doi.org/10.1214/16-EJS1116
[5] Biscio, C.A.N. and Lavancier, F. (2017). Contrast estimation for parametric stationary determinantal point processes. *Scand. J. Stat.* **44** 204–229. MR3619702 https://doi.org/10.1111/sjos.12249
[6] Biscio, C.A.N. and Waagepetersen, R. (2019). A general central limit theorem and a subsampling variance estimator for α-mixing point processes. *Scand. J. Stat.* **46** 1168–1190. MR4033808 https://doi.org/10.1111/sjos.12389
[7] Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, Expanded ed. San Francisco, CA: Holden-Day, INC. MR0595684
[8] Chandler, R.E. (1997). A spectral method for estimating parameters in rainfall models. *Bernoulli* **3** 301–322. MR1468307 https://doi.org/10.2307/3318594
[9] Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *Ann. Statist.* **50** 1722–1746. MR4441138 https://doi.org/10.1214/22-aos2173
[10] Choiruddin, A., Coeurjolly, J.F. and Waagepetersen, R. (2021). Information criteria for inhomogeneous spatial point processes. *Aust. N. Z. J. Stat.* **63** 119–143. MR4296145 https://doi.org/10.1111/anzs.12327
[11] Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882. MR0919857 https://doi.org/10.1093/biomet/74.4.877
[12] Dahlhaus, R. and Sahm, M. (2000). Likelihood methods for nonstationary time series and random fields. *Resenhas IME-USP (São Paulo J. Math. Sci.)* **4** 457–477. MR1844728

1350-7265

[13] Dahlhaus, R. and Wefelmeyer, W. (1996). Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.* **24** 952–974. MR1401832 https://doi.org/10.1214/aos/1032526951
[14] Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, 2nd ed. New York City, NY: Springer. MR1950431 https://doi.org/10.1007/b97277
[15] D'Angelo, N., Adelfio, G., Abbruzzo, A. and Mateu, J. (2022). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *Ann. Appl. Stat.* **16** 791–815. MR4438812 https://doi.org/10.1214/21-aoas1519
[16] Diggle, P.J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. New York, NY: CRC press. MR3113855 https://doi.org/10.1201/b15326
[17] Doukhan, P. (1994). *Mixing: Properties and Examples*. New York City, NY: Springer. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0
[18] Folland, G.B. (1999). *Real Analysis: Modern Techniques and Their Applications 40*, 2nd ed. New York, NY: John Wiley & Sons. MR1681462
[19] Gabriel, E. and Diggle, P.J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Stat. Neerl.* **63** 43–51. MR2656916 https://doi.org/10.1111/j.1467-9574.2008.00407.x
[20] Guillaumin, A.P., Sykulski, A.M., Olhede, S.C. and Simons, F.J. (2022). The debiased spatial Whittle likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1526–1557. MR4494169 https://doi.org/10.1111/rssb.12539
[21] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69** 95–105. MR0655674 https://doi.org/10.2307/2335857
[22] Hawkes, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410 https://doi.org/10.2307/2334319
[23] Hawkes, A.G. (1971). Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **33** 438–443. MR0358976 https://doi.org/10.1111/j.2517-6161.1971.tb01530.x
[24] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Hoboken, NJ: John Wiley & Sons. MR2384630 https://doi.org/10.1002/9780470725160
[25] Jovanović, S., Hertz, J. and Rotter, S. (2015). Cumulants of Hawkes point processes. *Phys. Rev. E* **91** 042802. https://doi.org/10.1103/PhysRevE.91.042802
[26] Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab.* **7** 83–122. MR0380979 https://doi.org/10.2307/1425855
[27] Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. MR2655530 https://doi.org/10.1111/j.1467-9868.2008.00685.x
[28] Møller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019 https://doi.org/10.1111/1467-9469.00115
[29] Møller, J. and Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL: Chapman and Hall/CRC. MR2004226 https://doi.org/10.1201/9780203496930
[30] Mugglestone, M.A. and Renshaw, E. (1996). A practical guide to the spectral analysis of spatial point processes. *Comput. Statist. Data Anal.* **21** 43–65. MR1380832 https://doi.org/10.1016/0167-9473(95)00007-0
[31] Neyman, J. and Scott, E.L. (1958). Statistical approach to problems of cosmology. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **20** 1–29. MR0105309 https://doi.org/10.1111/j.2517-6161.1958.tb00272.x
[32] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28** 329–348. MR0088833 https://doi.org/10.1214/aoms/1177706962
[33] Poinas, A., Delyon, B. and Lavancier, F. (2019). Mixing properties and central limit theorem for associated point processes. *Bernoulli* **25** 1724–1754. MR3961228 https://doi.org/10.3150/18-BEJ1033
[34] Poinas, A. and Lavancier, F. (2023). Asymptotic approximation of the likelihood of stationary determinantal point processes. *Scand. J. Stat.* **50** 842–874. MR4599935 https://doi.org/10.1111/sjos.12613
[35] Prokešová, M. and Jensen, E.V.B. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Ann. Inst. Statist. Math.* **65** 387–412. MR3011627 https://doi.org/10.1007/s10463-012-0376-7
[36] Rajala, T.A., Olhede, S.C., Grainger, J.P. and Murrell, D.J. (2023). What is the Fourier transform of a spatial point process? *IEEE Trans. Inf. Theory* **69** 5219–5252. MR4624729 https://doi.org/10.1109/tit.2023.3269514
[37] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

[38] Shlomovich, L., Cohen, E.A.K., Adams, N. and Patel, L. (2022). Parameter estimation of binned Hawkes processes. *J. Comput. Graph. Statist.* **31** 990–1000. MR4513364 https://doi.org/10.1080/10618600.2022.2050247

[39] Subba Rao, S. (2018). Statistical inference for spatial statistics defined in the Fourier domain. *Ann. Statist.* **46** 469–499. MR3782374 https://doi.org/10.1214/17-AOS1556

[40] Subba Rao, S. and Yang, J. (2021). Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Ann. Statist.* **49** 2774–2802. MR4338383 https://doi.org/10.1214/21-aos2055

[41] Tanaka, U., Ogata, Y. and Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes. *Biom. J.* **50** 43–57. MR2414637 https://doi.org/10.1002/bimj.200610339

[42] Tukey, J.W. (1967). An introduction to the calculations of numerical spectrum analysis. In *(Proc. Advanced Sem., Madison, Wis., 1966)* 25–46. New York: John Wiley. MR0212977

[43] Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 685–702. MR2749914 https://doi.org/10.1111/j.1467-9868.2008.00702.x

[44] Warton, D.I. and Shepherd, L.C. (2010). Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Ann. Appl. Stat.* **4** 1383–1402. MR2758333 https://doi.org/10.1214/10-aoas331

[45] Whittle, P. (1953). The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **15** 125–139. MR0056902 https://doi.org/10.1111/j.2517-6161.1953.tb00131.x

[46] Yang, J. and Guan, Y. (2026). Supplement to “Fourier analysis of spatial point processes.” https://doi.org/10.3150/25-BEJ1862SUPP

[47] Zhu, L., Yang, J., Jun, M. and Cook, S. (2025). On minimum contrast method for multivariate spatial point processes. *Electron. J. Stat.* **19** 1889–1941. MR4890854 https://doi.org/10.1214/25-ejs2378

[48] Zhuang, J., Ogata, Y. and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res., Solid Earth* **109**. https://doi.org/10.1029/2003JB002879

Bernoulli **32**(1), 2026, 393–416
https://doi.org/10.3150/25-BEJ1863

On adaptive confidence ellipsoids for sparse high-dimensional linear models

XIAOYANG XIE[a]

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK, [a]*xx765@cam.ac.uk*

In high-dimensional Gaussian linear models, the problem of constructing adaptive confidence sets for the full parameter is known to be generally impossible. We propose re-weighted 'ellipsoidal' loss functions parameterized by the power $\alpha \geq 0$ of the weights. For $0 \leq \alpha < 1/2$, the minimax convergence rates in these loss functions are slower than $1/\sqrt{n}$ and depend on the unknown sparsity level k. We show that confidence sets that are honest and adapt to k can be constructed if and only if $\alpha \geq 1/4$. To achieve this, we derive lower and upper bounds for certain composite minimax testing problems arising with sparse vectors.

Keywords: Adaptive inference; sparse regression; uncertainty quantification

References

Arjovsky, M., Chintala, S. and Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y.W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 214–223. PMLR.

Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

Bühlmann, P. and van de Geer, S. (2011). *Methods, Theory and Applications*. *Statistics for High-Dimensional Data*. *Springer Series in Statistics*. Heidelberg: Springer. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

Bull, A.D. and Nickl, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Related Fields* **156** 889–919. MR3078289 https://doi.org/10.1007/s00440-012-0446-z

Cai, T.T. and Guo, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

Carpentier, A. and Nickl, R. (2015). On signal detection and confidence sets for low rank inference problems. *Electron. J. Stat.* **9** 2675–2688. MR3432430 https://doi.org/10.1214/15-EJS1087

Carpentier, A. and Verzelen, N. (2021). Optimal sparsity testing in linear regression model. MR4255213 https://doi.org/10.3150/20-bej1224

Carpentier, A., Klopp, O., Löffler, M. and Nickl, R. (2018). Adaptive confidence sets for matrix completion. *Bernoulli* **24** 2429–2460. MR3779691 https://doi.org/10.3150/17-BEJ933

Deo, N. and Randrianarisoa, T. (2023). On adaptive confidence sets for the Wasserstein distances. MR4580910 https://doi.org/10.3150/22-bej1535

Dezeure, R., Bühlmann, P. and Zhang, C.-H. (2017). Rejoinder on: High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 751–758. MR3713592 https://doi.org/10.1007/s11749-017-0560-4

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

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Technical Report, North Carolina State University. Dept. of Statistics.

Hoffmann, M. and Nickl, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872 https://doi.org/10.1214/11-AOS903

Ingster, Y.I., Tsybakov, A.B. and Verzelen, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589

Löffler, M. and Nickl, R. (2017). Comments on: High-dimensional simultaneous inference with the bootstrap [MR3713586]. MR3713589 https://doi.org/10.1007/s11749-017-0558-y

Nickl, R. and van de Geer, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450 https://doi.org/10.1214/13-AOS1170

Robins, J. and van der Vaart, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241 https://doi.org/10.1214/009053605000000877

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

Zhang, C.-H. and Zhang, S.S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

Bernoulli **32**(1), 2026, 417–441
https://doi.org/10.3150/25-BEJ1864

Consistency of the oblique decision tree and its boosting and random forest

HAORAN ZHAN[1,a], YU LIU[2,b] and YINGCUN XIA[3,c]

[1]*Institute of Statistical Interdisciplinary Research, Southwestern University of Finance and Economics, Chengdu, China,* [a]*zhan.haoran.cn@gmail.com*
[2]*School of Mathematics Science, Sichuan Normal University, Chengdu, China,* [b]*liuy8stat@sicnu.edu.cn*
[3]*Department of Statistics and Data Science, National University of Singapore, Singapore, Singapore ,* [c]*staxyc@nus.edu.sg*

Classification and Regression Tree (CART), Random Forest (RF) and Gradient Boosting Tree (GBT) are probably the most popular set of statistical learning methods. However, their statistical consistency can only be proved under very restrictive assumptions on the underlying regression function. As an extension to standard CART, the oblique decision tree (ODT), which uses linear combinations of predictors as partitioning variables, has received much attention. ODT tends to perform numerically better than CART and requires fewer partitions. In this paper, we show that ODT is consistent for very general regression functions as long as they are L^2 integrable. Then, we prove the consistency of the ODT-based random forest (ODRF), whether fully grown or not. Finally, we propose an ensemble of GBT for regression by borrowing the technique of orthogonal matching pursuit and study its consistency under very mild conditions on the tree structure. After refining existing computer packages according to the established theory, extensive experiments on real data sets show that both our ensemble boosting trees and ODRF have noticeable overall improvements over RF and other forests.

Keywords: CART; consistency; feature bagging; gradient boosting tree; nonparametric regression; oblique decision tree; random forest

References

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

Bagirov, A.M., Clausen, C. and Kohler, M. (2009). Estimation of a regression function by maxima of minima of linear functions. *IEEE Trans. Inf. Theory* **55** 833–845. MR2597271 https://doi.org/10.1109/TIT.2008.2009835

Bartlett, P.L., Harvey, N., Liaw, C. and Mehrabian, A. (2019). Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.* **20** 2285–2301. MR3960917

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

Biau, G. and Devroye, L. (2010). On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. *J. Multivariate Anal.* **101** 2499–2518. MR2719877 https://doi.org/10.1016/j.jmva.2010.06.019

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

Blaser, R. and Fryzlewicz, P. (2016). Random rotation ensembles. *J. Mach. Learn. Res.* **17** 126–151. MR3482924

Blumer, A., Ehrenfeucht, A., Haussler, D. and Warmuth, M.K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *J. ACM* **36** 929–965. MR1072253 https://doi.org/10.1145/76359.76371

Breiman, L. (1984). *Classification and Regression Trees Regression Trees*, 1st ed. Chapman and Hall/CRC. MR0726392 https://doi.org/10.1201/9781315139470

Breiman, L. (2001a). Random forests. *J. Mach. Learn.* **45** 5–32. MR3874153

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statist. Sci.* **16** 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

Breiman, L. (2004). Consistency for a simple model of random forests at Berkeley. Technical Report 670, University of California.

Bühlmann, P.L. (2002). Consistency for L_2 boosting and matching pursuit with trees and tree-type basis functions. In *Research Report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH). Seminar für Statistik* **109**. Eidgenössische Technische Hochschule (ETH).

Cattaneo, M.D., Chandak, R. and Klusowski, J.M. (2022). Convergence Rates of Oblique Regression Trees for Flexible Function Libraries. ArXiv preprint. Available at arXiv:2210.14429. MR4744184 https://doi.org/10.1214/24-aos2354

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proc. 22nd ACM SIGKDD Int. Conf. KDD* 785–794. https://doi.org/10.1145/2939672.2939785

Chi, C.-M., Vossler, P., Fan, Y. and Lv, J. (2020). Asymptotic Properties of High-Dimensional Random Forests. ArXiv preprint. Available at arXiv:2004.13953. MR4524502 https://doi.org/10.1214/22-aos2234

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control. Signal.* **2** 303–314. MR1015670 https://doi.org/10.1007/BF02551274

da Silva, N., Cook, D. and Lee, E.K. (2021). A projection pursuit forest algorithm for supervised classification. *J. Comput. Graph. Statist.* **30** 1168–1180. MR4356612 https://doi.org/10.1080/10618600.2020.1870480

Denil, M., Matheson, D. and Freitas, N. (2013). Consistency of online random forests. In *International Conference on Machine Learning* 1256–1264. PMLR.

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

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

Genuer, R. (2012). Variance reduction in purely random forests. *J. Nonparametr. Stat.* **24** 543–562. MR2968888 https://doi.org/10.1080/10485252.2012.677843

Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression 1*. Springer. MR1920390 https://doi.org/10.1007/b97848

Heath, D., Kasif, S. and Salzberg, S. (1993). Induction of oblique decision trees. *J. Artificial Intelligence Res.* **2** 1–32.

Ishwaran, H. and Kogalur, U.B. (2010). Consistency of random survival forests. *Statist. Probab. Lett.* **80** 1056–1064. MR2651045 https://doi.org/10.1016/j.spl.2010.02.020

Kim, H. and Loh, W.-Y. (2001). Classification trees with unbiased multiway splits. *J. Amer. Statist. Assoc.* **96** 589–604. MR1946427 https://doi.org/10.1198/016214501753168271

Klusowski, J.M. (2021). Universal Consistency of Decision Trees in High Dimensions. ArXiv preprint. Available at arXiv:2104.13881.

Lee, D., Yang, M.-H. and Oh, S. (2015). Fast and accurate head pose estimation via random projection forests. In *Proc. IEEE Int. Conf. Comput. Vis* 1958–1966. https://doi.org/10.1109/ICCV.2015.227

Lin, Y. and Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176 https://doi.org/10.1198/016214505000001230

Majumder, T. (2020). *Ensembles of Oblique Decision Trees*. The University of Texas at Dallas.

Mentch, L. and Hooker, G. (2014). Ensemble trees and clts: Statistical inference for supervised learning. *Stat.* **1050** 25. MR3450160

Menze, B.H., Kelm, B.M., Splitthoff, D.N., Koethe, U. and Hamprecht, F.A. (2011). On oblique random forests. In *ECML PKDD 2011* 453–469. Springer. https://doi.org/10.1007/978-3-642-23783-6_29

Pati, Y.C., Rezaiifar, R. and Krishnaprasad, P.S. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. 27th Asil. Conf. SSC* 40–44. IEEE. https://doi.org/10.1109/ACSSC.1993.342465

Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. *Acta Numer.* **8** 143–195. MR1819645 https://doi.org/10.1017/S0962492900002919

Rainforth, T. and Wood, F. (2015). Canonical correlation forests. ArXiv preprint. Available at arXiv:1507.05444.

Scornet, E., Biau, G. and Vert, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876 https://doi.org/10.1214/15-AOS1321

Syrgkanis, V. and Zampetakis, M. (2020). Estimation and inference with trees and forests in high dimensions. In *Conference on Learning Theory* 3453–3454. PMLR.

Tomita, T.M., Browne, J., Shen, C., Chung, J., Patsolic, J.L., Falk, B., Priebe, C.E., Yim, J., Burns, R., Maggioni, M. et al. (2020). Sparse projection oblique randomer forests. *J. Mach. Learn. Res.* **21**. MR4119172

Wager, S. (2014). Asymptotic theory for random forests. ArXiv preprint. Available at arXiv:1405.0352.

Zhan, X., Liu, Y. and Xia, Y. (2026). Supplement to "Consistency of the oblique decision tree and its boosting and random forest." https://doi.org/10.3150/25-BEJ1864SUPP

Zhang, T. and Yu, B. (2005). Boosting with early stopping: Convergence and consistency. *Ann. Statist.* **33** 1538–1579. MR2166555 https://doi.org/10.1214/009053605000000255

Zhou, Y. and Hooker, G. (2022). Decision tree boosted varying coefficient models. *Data Min. Knowl. Discov.* **36** 2237–2271. MR4510524 https://doi.org/10.1007/s10618-022-00863-y

Zhu, R., Zeng, D. and Kosorok, M.R. (2015). Reinforcement learning trees. *J. Amer. Statist. Assoc.* **110** 1770–1784. MR3449072 https://doi.org/10.1080/01621459.2015.1036994

Bernoulli **32**(1), 2026, 442–466
https://doi.org/10.3150/25-BEJ1865

Balancing the edge effect and dimension of spectral spatial statistics under irregular sampling with applications to isotropy testing

THERESA ECKLE[1,a], ANNE VAN DELFT[2,b] and HOLGER DETTE[3,c]

[1]*Department of Mathematics, Ruhr University Bochum, Bochum, Germany,* [a]*Theresa.Eckle@rub.de*
[2]*Department of Statistics, Columbia University, New York, USA,* [b]*anne.vandelft@columbia.edu*
[3]*Department of Mathematics, Ruhr University Bochum, Bochum, Germany,* [c]*holger.dette@rub.de*

We investigate distributional properties of a class of spectral spatial statistics under irregular sampling of a random field that is defined on $\mathbb{R}^d$, and use this to obtain a novel test for isotropy via a minimum distance approach. More precisely, we derive an explicit expression for the minimum L^2-distance between the spectral density of the random field and its best approximation by a spectral density of an isotropic process in terms of certain integrals of the spectral density, which are then estimated from the data. Within this context, edge effects are well-known to create a bias in classical estimators commonly encountered in the analysis of spatial data. This bias increases with dimension d and, for $d > 1$, can become non-negligible in the limiting distribution of such statistics to the extent that a nondegenerate distribution does not exist. We provide a general theory for a class of (integrated) spectral statistics that are used to estimate the minimal L^2-distance, which enables to 1) significantly reduce this bias and 2) that ensures that asymptotically Gaussian limits can be derived for $d \leq 3$ for appropriately tapered versions of such statistics. We use this to address some crucial gaps in the literature, and demonstrate that tapering with a sufficiently smooth function is necessary to achieve such results. Our findings specifically shed a new light on a recent result in (*Ann. Statist.* **46** (2018) 469–499). In contrast to most of the literature, which validates this assumption on a finite number of spatial locations (or a finite number of Fourier frequencies), we develop a test for isotropy on the full spatial domain by means of its characterization in the frequency domain. We prove asymptotic normality of the corresponding in the mixed increasing domain framework and use this result to derive an asymptotic level α-test.

Keywords: Isotropy; mixed increasing domain asymptotics; spatial processes; tapering

References

Baczkowski, A.J. (1990). A test of spatial isotropy. In *Compstat* (K. Momirović and V. Mildner, eds.) 277–282. Heidelberg: Physica-Verlag HD. MR1118608 https://doi.org/10.1007/978-3-642-50096-1_42

Baczkowski, A.J. and Mardia, K.V. (1990). A test of spatial symmetry with general application. *Comm. Statist. Theory Methods* **19** 555–572. MR1067588 https://doi.org/10.1080/03610929008830218

Bagchi, P., Characiejus, V. and Dette, H. (2018). A simple test for white noise in functional time series. *J. Time Series Anal.* **39** 54–74. MR3743093 https://doi.org/10.1111/jtsa.12264

Bandyopadhyay, S. and Lahiri, S.N. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhya A* **71** 221–259. MR2639292

Bandyopadhyay, S. and Subba Rao, S. (2017). A test for stationarity for irregularly spaced spatial data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 95–123. MR3597966 https://doi.org/10.1111/rssb.12161

Bowman, A.W. and Crujeiras, R.M. (2013). Inference for variograms. *Comput. Statist. Data Anal.* **66** 19–31. MR3064021 https://doi.org/10.1016/j.csda.2013.02.027

Brillinger, D.R. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **1** 215–218. https://doi.org/10.1007/BF02532246

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*. Fransisco: San Holden-Day. MR0595684

1350-7265

Cabaña, E.M. (1987). Affine processes: A test of isotropy based on level sets. *SIAM J. Appl. Math.* **47** 886–891. MR0898839 https://doi.org/10.1137/0147059

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York: Wiley. MR1239641 https://doi.org/10.1002/9781119115151

Crujeiras, R.M. and Fernández-Casal, R. (2010). On the estimation of the spectral density for continuous spatial processes. *Statistics* **44** 587–600. MR2739414 https://doi.org/10.1080/02331880903270313

Dahlhaus, R. (1983). Spectral analysis with tapered data. *J. Time Series Anal.* **4** 163–175. MR0732895 https://doi.org/10.1111/j.1467-9892.1983.tb00366.x

Dahlhaus, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.* **16** 808–841. MR0947580 https://doi.org/10.1214/aos/1176350838

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. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882. MR0919857 https://doi.org/10.1093/biomet/74.4.877

Deo, R.S. and Chen, W.W. (2000). On the integral of the squared periodogram. *Stochastic Process. Appl.* **85** 159–176. MR1730613 https://doi.org/10.1016/S0304-4149(99)00071-X

Dette, H., Preuss, P. and Vetter, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768 https://doi.org/10.1198/jasa.2011.tm10811

Eckle, T., van Delft, A. and Dette, H. (2026). Supplement to "Balancing the edge effect and dimension of spectral spatial statistics under irregular sampling with applications to isotropy testing." https://doi.org/10.3150/25-BEJ1865SUPP

Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **10** 517–532. MR0840512 https://doi.org/10.1214/aos/1176349936

Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *J. Multivariate Anal.* **96** 30–54. MR2202399 https://doi.org/10.1016/j.jmva.2004.09.003

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545 https://doi.org/10.1198/016214506000000852

Guan, Y., Sherman, M. and Calvin, J.A. (2004). A nonparametric test for spatial isotropy using subsampling. *J. Amer. Statist. Assoc.* **99** 810–821. MR2090914 https://doi.org/10.1198/016214504000001150

Guan, Y., Sherman, M. and Calvin, J.A. (2007). On asymptotic properties of the mark variogram estimator of a marked point process. *J. Statist. Plann. Inference* **137** 148–161. MR2292847 https://doi.org/10.1016/j.jspi.2005.10.004

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69** 95–105. MR0655674 https://doi.org/10.1093/biomet/69.1.95

Harris, F.J. (1978). On the use of windows for harmonic analsis with the discrete Fourier transform. *Proc. IEEE* **66** 51–83. https://doi.org/10.1109/PROC.1978.10837

Jona-Lasinio, G. (2001). Modeling and exploring multivariate spatial variation: A test procedure for isotropy of multivariate spatial data. *J. Multivariate Anal.* **77** 295–317. MR1838690 https://doi.org/10.1006/jmva.2000.1938

Lu, H. and Zimmerman, D.L. (2005). Testing for directional symmetry in spatial dependence using the periodogram. *J. Statist. Plann. Inference* **129** 369–385. MR2126855 https://doi.org/10.1016/j.jspi.2004.06.058

Maity, A. and Sherman, M. (2012). Testing for spatial isotropy under general designs. *J. Statist. Plann. Inference* **142** 1081–1091. MR2879753 https://doi.org/10.1016/j.jspi.2011.11.013

Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. Roy. Statist. Soc. Ser. B, Methodol.* **71** 191–217. MR2655530 https://doi.org/10.1111/j.1467-9868.2008.00685.x

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press. MR2514435

Robinson, P.M. (2007). Nonparametric spectrum estimation for spatial data. *J. Statist. Plann. Inference* **137** 1024–1034. MR2301732 https://doi.org/10.1016/j.jspi.2006.06.021

Sahoo, I., Guinness, J. and Reich, B.J. (2019). A test for isotropy on a sphere using spherical harmonic functions. *Statist. Sinica* **29** 1253–1276. MR3932517 https://doi.org/10.5705/ss.202017.0475

Scaccia, L. and Martin, R.J. (2005). Testing axial symmetry and separability of lattice processes. *J. Statist. Plann. Inference* **131** 19–39. MR2136004 https://doi.org/10.1016/j.jspi.2004.01.003

Subba Rao, S. (2018a). Statistical inference for spatial statistics defined in the Fourier domain. *Ann. Statist.* **46** 469–499. MR3782374 https://doi.org/10.1214/17-AOS1556

Subba Rao, S. (2018b). Supplement to statistical inference for spatial statistics defined in the fourier domain.

Taniguchi, M. (1980). On estimation of the integrals of certain functions of spectral density. *J. Appl. Probab.* **17** 73–83. MR0557436 https://doi.org/10.2307/3212925

Tukey, J.W. (1967). An introduction to the calculations of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series* (B. Harris, ed.) 25–46. New York: Wiley. MR0212977

van Delft, A., Characiejus, V. and Dette, H. (2021). A nonparametric test for stationarity in functional time series. *Statist. Sinica* **31** 1375–1395. MR4297704 https://doi.org/10.5705/ss.202018.0320

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electron. J. Stat.* **12** 107–170. MR3746979 https://doi.org/10.1214/17-EJS1384

Van Hala, M., Bandyopadhyay, S., Lahiri, S.N. and Nordman, D.J. (2020). A general frequency domain method for assessing spatial covariance structures. *Bernoulli* **26** 2463–2487. MR4140518 https://doi.org/10.3150/19-BEJ1160

Watson, G.N. (1966). *A Treatise on the Theory of Bessel Functions*. Cambridge University Press. MR0010746

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Math.* **2** 423–434. MR0060797 https://doi.org/10.1007/BF02590998

Yang, J. and Guan, Y. (2024). *Fourier Analysis of Spatial Point Processes*. https://arxiv.org/abs/2401.06403.

Bernoulli **32**(1), 2026, 467–492
https://doi.org/10.3150/25-BEJ1866

Azadkia–Chatterjee's dependence coefficient for infinite dimensional data

SIEGFRIED HÖRMANN[a] and DANIEL STRENGER[b]

Institute of Statistics, Graz University of Technology, Graz, Austria, [a]*shoermann@tugraz.at,* [b]*daniel.strenger@tugraz.at*

We extend the scope of Azadkia–Chatterjee's dependence coefficient between a scalar response Y and a multivariate covariate X to the case where X takes values in a general metric space. Particular attention is paid to the case where X is a curve. Although extending this framework at the population level is relatively straightforward, analyzing the asymptotic behavior of the estimator proves to be complex. This complexity is largely related to the nearest neighbor structure of the infinite-dimensional covariate sample, leading us to explore a topic that has not been previously addressed in the literature. The primary contribution of this paper is to provide insights into this issue and propose strategies to address it. Our findings also have significant implications for other graph-based methods facing similar challenges.

Keywords: Dependence measure; functional data analysis; independence test; nearest neighbor graph

References

Auddy, A., Deb, N. and Nandy, S. (2024). Exact detection thresholds and minimax optimality of Chatterjee's correlation coefficient. *Bernoulli* **30** 1640–1668. MR4699567 https://doi.org/10.3150/23-bej1648

Azadkia, M. and Chatterjee, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. MR4352523 https://doi.org/10.1214/21-aos2073

Beyer, K., Goldstein, J., Ramakrishnan, R. and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *Database Theory — ICDT'99* (C. Beeri and P. Buneman, eds.) 217–235. Berlin: Springer. https://doi.org/10.1007/3-540-49257-7_15

Bickel, P. (2022). Measures of independence and functional dependence. https://doi.org/10.48550/arXiv.2206.13663

Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (BMSGPK) (2021). COVID-19 Schutzimpfungen – Impfungen in Gemeinden.

Cao, S. and Bickel, P. (2020). Correlations with tailored extremal properties. https://doi.org/10.48550/arXiv.2008.10177

Chatterjee, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859 https://doi.org/10.1214/07-AOP370

Chatterjee, S. (2020). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 1–21. MR4353729 https://doi.org/10.1080/01621459.2020.1758115

Chatterjee, S. (2024). A survey of some recent developments in measures of association. In *Probability and Stochastic Processes: A Volume in Honour of Rajeeva L. Karandikar* (S. Athreya, A.G. Bhatt and B.V. Rao, eds.) 109–128. Springer Nature. MR4824372 https://doi.org/10.1007/978-981-99-9994-1_4

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, 1st ed. *Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. New York: Academic Press. MR0666546

Deb, N., Ghosal, P. and Sen, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. https://doi.org/10.48550/arXiv.2010.01768

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

Durrant, R. and Kabán, A. (2009). When is 'nearest neighbour' meaningful: A converse theorem and implications. *J. Complexity* **25** 385–397. MR2542038 https://doi.org/10.1016/j.jco.2009.02.011

Durrett, R. (2019). *Probability: Theory and Examples*, 5th ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press. MR2722836 https://doi.org/10.1017/CBO9780511779398

García-Portugués, E., González-Manteiga, W. and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *J. Comput. Graph. Statist.* **23** 761–778. MR3224655 https://doi.org/10.1080/10618600.2013.812519

Griessenberger, F., Junker, R. and Trutschnig, W. (2022). On a multivariate copula-based dependence measure and its estimation. *Electron. J. Stat.* **16** 2206–2251. MR4401220 https://doi.org/10.1214/22-ejs2005

Han, F. and Huang, Z. (2024). Azadkia–Chatterjee's correlation coefficient adapts to manifold data. *Ann. Appl. Probab.* **34** 5172–5210. MR4840483 https://doi.org/10.1214/24-aap2088

Henze, N. (1987). On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction. *Adv. Appl. Probab.* **19** 873–895. MR0914597 https://doi.org/10.2307/1427106

Hörmann, S. and Strenger, D. (2026). Supplement to "Azadkia–Chatterjee's dependence coefficient for infinite dimensional data." https://doi.org/10.3150/25-BEJ1866SUPP

Huang, Z., Deb, N. and Sen, B. (2022). Kernel partial correlation coefficient — a measure of conditional dependence. *J. Mach. Learn. Res.* **23** 9699–9756. MR4577169 https://doi.org/10.48550/arXiv

Jeffreys, H. (1961). *The Theory of Probability*, 3rd ed. *International Series of Monographs on Physics*. Oxford: Oxford University Press. MR0187257

Kabatjanski, G. and Levenstein, V. (1978). Bounds for packings on the sphere and in space. *Problemy Peredachi Informatsii* **14** 3–25. MR0514023

Lin, Z. and Han, F. (2022). Limit theorems of Chatterjee's rank correlation. MR4589063 https://doi.org/10.48550/arXiv.2204.08031

Lin, Z. and Han, F. (2023). On boosting the power of Chatterjee's rank correlation. *Biometrika* **110** 283–299. MR4589063 https://doi.org/10.1093/biomet/asac048

Lin, Z. and Han, F. (2024). On the failure of the bootstrap for Chatterjee's rank correlation. *Biometrika* **111** 1063–1070. MR4794428 https://doi.org/10.1093/biomet/asae004

Ofner, M. (2021). Quantile regression for weakly dependent functional data Master's thesis, Graz University of Technology.

Pan, W., Wang, X., Zhang, H., Zhu, H. and Zhu, J. (2020). Ball covariance: A generic measure of dependence in Banach space. *J. Amer. Statist. Assoc.* **115** 307–317. MR4078465 https://doi.org/10.1080/01621459.2018.1543600

Pyke, R. (1965). Spacings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **27** 395–436. MR0216622 https://doi.org/10.1111/j.2517-6161.1965.tb00602.x

Rinott, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.* **55** 135–143. MR1327369 https://doi.org/10.1016/0377-0427(94)90016-7

Shi, H., Drton, M. and Han, F. (2022). On the power of Chatterjee's rank correlation. *Biometrika* **109** 317–333. MR4430960 https://doi.org/10.1093/biomet/asab028

Shi, H., Drton, M. and Han, F. (2024). On Azadkia–Chatterjee's conditional dependence coefficient. *Bernoulli* **30** 851–877. MR4699537 https://doi.org/10.3150/22-bej1529

Statistik Austria (2021). Bevölkerung zu Jahresbeginn 2021.

Székely, G., Rizzo, M. and Bakirov, N. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. MR2825218 https://doi.org/10.1016/j.jmaa.2011.06.013

van der Vaart, A. (1998). *Asymptotic Statistics*, 1st ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Zhang, Q. (2023). On the asymptotic null distribution of the symmetrized Chatterjee's correlation coefficient. *Statist. Probab. Lett.* **194**. MR4525660 https://doi.org/10.1016/j.spl.2022.109759

Bernoulli **32**(1), 2026, 493–518
https://doi.org/10.3150/25-BEJ1867

Data-driven fixed-point tuning for truncated realized variations

B. COOPER BONIECE[1,a], JOSÉ E. FIGUEROA-LÓPEZ[2,b] and YUCHEN HAN[2,c]

[1]*Department of Mathematics, Drexel University,* [a]*cooper.boniece@drexel.edu*
[2]*Department of Statistics and Data Science, Washington University in St. Louis,* [b]*figueroa-lopez@wustl.edu,* [c]*y.han@wustl.edu*

Many methods for estimating integrated volatility and related functionals of semimartingales in the presence of jumps require specification of tuning parameters for their use in practice. In much of the available theory, tuning parameters are assumed to be deterministic and their values are specified only up to asymptotic constraints. However, in empirical work and in simulation studies, they are typically chosen to be random and data-dependent, with explicit choices often relying entirely on heuristics. In this paper, we consider novel data-driven tuning procedures for the truncated realized variations of a semimartingale with jumps based on a type of random fixed-point iteration. Being effectively automated, our approach alleviates the need for delicate decision-making regarding tuning parameters in practice and can be implemented using information regarding sampling frequency alone. We demonstrate our methods can lead to asymptotically efficient estimation of integrated volatility and exhibit superior finite-sample performance compared to popular alternatives in the literature.

Keywords: High-frequency data; integrated volatility estimation; semimartingales

References

Andersen, T.G., Dobrev, D. and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *J. Econometrics* **169** 75–93. MR2935265 https://doi.org/10.3386/w15533

Barndorff-Nielsen, O.E. (2004). Power and bipower variation with stochastic volatility and jumps. *J. Financ. Econom.* **2** 1–37. https://doi.org/10.1093/jjfinec/nbh001

Barndorff-Nielsen, O.E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl.* **116** 796–806. MR2218336 https://doi.org/10.1016/j.spa.2006.01.007

Boniece, B.C., Figueroa-López, J.E. and Han, Y. (2024). Efficient integrated volatility estimation in the presence of infinite variation jumps via debiased truncated realized variations. *Stochastic Process. Appl.* **176** 104429. MR4774175 https://doi.org/10.1016/j.spa.2024.104429

Boniece, B.C., Figueroa-López, J.E. and Han, Y. (2026). Supplement to "Data-driven fixed-point tuning for truncated realized variations." https://doi.org/10.3150/25-BEJ1867SUPP

Cont, R. and Mancini, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli* **17** 781–813. MR2787615 https://doi.org/10.3150/10-bej293

Corsi, F., Pirino, D. and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *J. Econometrics* **159** 276–288. MR2733121 https://doi.org/10.1016/j.jeconom.2010.07.008

Ding, Y., Li, Y., Liu, G. and Zheng, X. (2023). Stock co-jump networks. *J. Econometrics* **105420**. MR4708617 https://doi.org/10.1016/j.jeconom.2023.01.026

Fan, J. and Wang, Y. (2008). Spot volatility estimation for high-frequency data. *Stat. Interface* **1** 279–288. MR2476744 https://doi.org/10.4310/sii.2008.v1.n2.a5

Figueroa-López, J.E., Gong, R. and Han, Y. (2022). Estimation of tempered stable Lévy models of infinite variation. *Methodol. Comput. Appl. Probab.* **24** 713–747. MR4446864 https://doi.org/10.1007/s11009-022-09940-7

Figueroa-López, J.E. and Li, C. (2020). Optimal kernel estimation of spot volatility of stochastic differential equations. *Stochastic Process. Appl.* **130** 4693–4720. MR4108468 https://doi.org/10.1016/j.spa.2020.01.013

Figueroa-López, J.E. and Mancini, C. (2019). Optimum thresholding using mean and conditional mean squared error. *J. Econometrics* **208** 179–210. MR3906967 https://doi.org/10.1016/j.jeconom.2018.09.011

Figueroa-López, J.E. and Nisen, J. (2013). Optimally thresholded realized power variations for Lévy jump diffusion models. *Stochastic Process. Appl.* **123** 2648–2677. MR3054540 https://doi.org/10.1016/j.spa.2013.04.006

Figueroa-López, J.E. and Wu, B. (2022). Kernel estimation of spot volatility with microstructure noise using pre-averaging. *Econom. Theory* **1** 50. MR4740935 https://doi.org/10.1017/s0266466622000470

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

Jacod, J. and Protter, P. (2011). *Discretization of Processes*. Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42** 1029–1069. MR3210995 https://doi.org/10.1214/14-aos1213

Jacod, J. and Todorov, V. (2018). Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation. *Ann. Appl. Probab.* **28**. MR3770883 https://doi.org/10.1214/17-aap1311

Kawai, R. (2010). On sequential calibration for an asset price model with piecewise Lévy processes. *IAENG Int. J. Appl. Math.* **40** 1–8. MR2778987

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econom. Theory* **26** 60–93. MR2587103 https://doi.org/10.1017/s0266466609090616

Li, J., Todorov, V. and Tauchen, G. (2017). Jump regressions. *Econometrica* **85** 173–195. MR3611769 https://doi.org/10.3982/ecta12962

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Gornale dell'Istituto Italiano degli Attuari* **64** 19–47.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.* **36** 270–296. MR2528985 https://doi.org/10.1111/j.1467-9469.2008.00622.x

Mancini, C. (2011). The speed of convergence of the threshold estimator of integrated variance. *Stochastic Process. Appl.* **121** 845–855. MR2770909 https://doi.org/10.1016/j.spa.2010.12.001

Mancini, C. and Renò, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. *J. Econometrics* **160** 77–92. MR2745869 https://doi.org/10.1016/j.jeconom.2010.03.019

Palmes, C. and Woerner, J.H.C. (2016). A mathematical analysis of the Gumbel test for jumps in stochastic volatility models. *Stoch. Anal. Appl.* **34** 852–881. MR3529199 https://doi.org/10.1080/07362994.2016.1182870

Sato, K-i. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press. MR1739520

Todorov, V. and Tauchen, G. (2012). The realized Laplace transform of volatility. *Econometrica* **80** 1105–1127. MR2963883 https://doi.org/10.3982/ecta9133

Bernoulli **32**(1), 2026, 519–543
https://doi.org/10.3150/25-BEJ1868

Concentration inequalities for classical and smoothed empirical processes of independent and dependent random variables

ERIC BEUTNER[1,a] and HENRYK ZÄHLE[2,b]

[1]*Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands,* [a]*e.a.beutner@vu.nl*
[2]*Department of Mathematics, Saarland University, D-66123 Saarbücken, Germany,* [b]*zaehle@math.uni-sb.de*

We present new concentration inequalities for classical and smoothed empirical processes of independent and dependent real-valued random variables. In the case of i.i.d. random variables, we show that Massart's celebrated refinement of the Dvoretzky–Kiefer–Wolfowitz inequality for the canonical empirical process extends to the empirical process indexed by classes of functions of uniformly bounded variation. The famous Talagrand inequality for empirical processes does not seem to give a similar result. Furthermore, we show that versions of both Massart's exponential bound and the generalisation we prove can also be obtained for the smoothed empirical process. In the case where the underlying random variables follow a linear process, we use a recently shown concentration inequality for the canonical empirical process to obtain, to the best of our knowledge, the first concentration inequalities for the smoothed empirical process of dependent random variables.

Keywords: Concentration inequalities; empirical process; function of bounded variation; linear time series; kernel smoothing; smoothed empirical process

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] Alexander, K. (1984). Probability inequalities for empirical processes and a law of iterated logarithm. *Ann. Probab.* **12** 1041–1067. MR0757769 https://doi.org/10.1214/aop/1176993141
[3] Bahnholzer, D., Fliege, J. and Werner, R. (2022). On rates of convergence for sample average approximations in the almost sure sense and in mean. *Math. Program.* **191** 307–345. MR4380094 https://doi.org/10.1007/s10107-019-01400-4
[4] Beutner, E. and Zähle, H. (2023). Donsker results for the empirical process indexed by functions of locally bounded variation and applications to the smoothed empirical process. *Bernoulli* **29** 205–228. MR4497245 https://doi.org/10.3150/21-bej1455
[5] Beutner, E. and Zähle, H. (2026). Supplement to "Concentration inequalities for classical and smoothed empirical processes of independent and dependent random variables." https://doi.org/10.3150/25-BEJ1868SUPP
[6] Bousquet, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications* (E. Giné, C. Houdré and D. Nualart, eds.) 180–212. Boston: Birkhäuser. MR2073435
[7] Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85** 799–808. MR1666695 https://doi.org/10.1093/biomet/85.4.799
[8] Chen, L. and Wu, W.B. (2018). Concentration inequalities for empirical processes of linear time series. *J. Mach. Learn. Res.* **18** 1–46. MR3845530
[9] Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* **10** 5931–606. MR2676786 https://doi.org/10.1080/14697681003685597

1350-7265

[10] Dudley, R.M. (2002). *Real Analysis and Probability*. Cambridge: Cambridge University Press. MR1932358 https://doi.org/10.1017/CBO9780511755347

[11] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669. MR0083864 https://doi.org/10.1214/aoms/1177728174

[12] Elstrodt, J. (2018). *Maß- und Integrationstheorie*. Berlin: Springer-Verlag. MR2257838

[13] Fernholz, L.T. (1997). Reducing the variance by smoothing. *J. Statist. Plann. Inference* **57** 29–38. MR1440227 https://doi.org/10.1016/S0378-3758(96)00033-X

[14] Giné, E. and Nickl, R. (2008). Adaptation on the space of finite signed measures. *Math. Methods Statist.* **17** 113–122. MR2429123 https://doi.org/10.3103/S1066530708020026

[15] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673 https://doi.org/10.1007/s00440-008-0137-y

[16] Giné, E. and Nickl, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge: Cambridge University Press. MR3588285 https://doi.org/10.1017/CBO9781107337862

[17] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley. MR0829458

[18] Hu, I. (1985). A uniform bound for the tail probability of Kolomogorov–Smirnov statistics. *Ann. Statist.* **13** 821–826. MR0790579 https://doi.org/10.1214/aos/1176349561

[19] Klein, T. (2002). Une inégalité de concentration à gauche pour les processus empiriques. *C. R. Math. Acad. Sci. Paris* **334** 501–504. MR1890641 https://doi.org/10.1016/S1631-073X(02)02303-8

[20] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312 https://doi.org/10.1214/009117905000000044

[21] Kozek, A.S. and Jersky, B. (2013). Variance reduction by smoothing revisited. *Probab. Math. Statist.* **33** 79–92. MR3119814

[22] Krätschmer, V. (2024). First order asymptotics of the sample average approximation method to solve risk averse stochastic programs. *Math. Program.* **208** 209–242. MR4809475 https://doi.org/10.1007/s10107-023-02036-1

[23] Massart, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 381–423. MR0871904

[24] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069 https://doi.org/10.1214/aop/1176990746

[25] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884. MR1782276 https://doi.org/10.1214/aop/1019160263

[26] Mitzenmacher, M. and Upfal, E. (2005). *Probability and Computing*. Cambridge: Cambridge University Press. MR2144605 https://doi.org/10.1017/CBO9780511813603

[27] Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018). *Foundations of Machine Learning*. Cambridge, MA: MIT Press. MR3931734

[28] Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.* **20** 1514–1533. MR1186263 https://doi.org/10.1214/aos/1176348782

[29] Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag. MR0762984 https://doi.org/10.1007/978-1-4612-5254-2

[30] Schilling, R. (2005). *Measures, Integrals and Martingales*. Cambridge: Cambridge University Press. MR3644418 https://doi.org/10.1017/CBO9780511810886

[31] Scott, D.W. (1992). *Multivariate Density Estimation*. New York: Wiley. MR3329609 https://doi.org/10.1002/9780470316849

[32] Shapiro, A., Dentcheva, D. and Ruszczýnski, A. (2014). *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia: SIAM. MR4362585

[33] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76. MR1258865 https://doi.org/10.1214/aop/1176988847

[34] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006 https://doi.org/10.1007/s002220050108

[35] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. New York: Springer-Verlag. MR2724359 https://doi.org/10.1007/b13794

[36] van de Geer, S. (1991). The entropy bound for monotone functions Technical Report, TW 91–10, University of Leiden.

[37] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[38] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag. MR4628026 https://doi.org/10.1007/978-1-4757-2545-2

Bernoulli **32**(1), 2026, 544–566
https://doi.org/10.3150/25-BEJ1869

On the power of private likelihood-ratio tests for goodness-of-fit in frequency tables

MARIO BERAHA[1,a], EMANUELE DOLERA[2,b] and STEFANO FAVARO[3,c]

[1]*Department of Economics, Management and Statistics, University of Milano - Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy* , [a]*mario.beraha@unimib.it*
[2]*Department of Mathematics, University of Pavia, Via Adolfo Ferrata 5, 27100 Pavia, Italy,* [b]*emanuele.dolera@unipv.it*
[3]*Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto, Corso Unione Sovietica 218/bis, 10134, Turin, Italy,* [c]*stefano.favaro@unito.it*

Privacy-protecting data analysis is a rising challenge in modern statistics, as the achievement of data confidentiality guarantees, which typically occur through a perturbation of the data, may determine a loss in the "statistical" utility of the data. In this paper, we consider the likelihood-ratio (LR) test for private goodness-of-fit in frequency tables, and study its large sample properties. Under the global (ε, δ)-differential privacy (DP) for frequency tables, with $\varepsilon, \delta \geq 0$ being parameters controlling the level of privacy against intruders, our main result is sharp large deviation principle (LDP) for the power of the LR test, providing a private version of the classical Bahadur–Rao LDP for the Multinomial LR test in the absence of perturbation. This is achieved through a novel sharp LDP for the sum of i.i.d. random vectors, which is of independent interest. The private Bahadur–Rao LDP brings out a critical quantity as a function of the sample size n, the dimension k of the table, and the DP parameters (ε, δ), which determines the loss of power in LR test due to the perturbation of the data, showing how n and k interact with the level of privacy, through ε and δ. This allows to control the (sample) cost of global (ε, δ)-DP, namely the additional sample size to recover the power of the Multinomial LR test. We validate empirically our results on both synthetic data and real data.

Keywords: Differential privacy; Edgeworth-type expansion; exponential mechanism; goodness-of-fit; likelihood-ratio test; power analysis; Bahadur–Rao large deviation principle; truncated Laplace distribution

References

Acharya, J., Sun, Z. and Zhang, H. (2018). Differentially private testing of identity and closeness of discrete distributions. *Adv. Neural Inf. Process. Syst.* **32**.

Aleshkyavichene, A.K. (1983). Multidimensional integral limit theorems for large deviation probabilities. *Theory Probab. Appl.* **28** 65–88. MR0691468

Aliakbarpour, M., Diakonikolas, I. and Rubinfeld, R. (2018). Differentially private identity and equivalence testing of discrete distributions. *Int. Conf. Mach. Learn.* **35**.

Avella-Medina, M. (2021). Privacy-preserving parametric inference: A case for robust statistics. *J. Amer. Statist. Assoc.* **116** 969–983. MR4270037 https://doi.org/10.1080/01621459.2019.1700130

Awan, J. and Slavković, A. (2018). Differentially private uniformly most powerful tests for Binomial data. *Adv. Neural Inf. Process. Syst.* **32**.

Awan, J. and Slavković, A. (2020). Differentially private inference for Binomial data. *J. Priv. Confid.* **10** 1–40.

Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295. MR0116413 https://doi.org/10.1214/aoms/1177705894

Bahadur, R.R. (1967). An optimal property of the likelihood-ratio statistic. *Proc. Berkeley Symp. Math. Stat. Probab.* **5** 13–26. MR0216637

Beraha, M., Dolera, E. and Favaro, S. (2026). Supplement to "On the power of private likelihood-ratio tests for goodness-of-fit in frequency tables." https://doi.org/10.3150/25-BEJ1869SUPP

1350-7265

Berrett, T.B. and Butucea, C. (2018). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Adv. Neural Inf. Process. Syst.* **34**.

Brenner, H. and Nissim, K. (2014). Impossibility of differentially private universally optimal mechanisms. *SIAM J. Comput.* **43** 1513–1540. MR3256803 https://doi.org/10.1137/110846671

Bun, M., Kamath, G., Steinke, T. and Wu, S.Z. (2021). Private hypothesis selection. *IEEE Trans. Inf. Theory* **67** 1981–2000. MR4282337 https://doi.org/10.1109/TIT.2021.3049802

Butucea, C., Rohde, A. and Steinberger, L. (2020). Interactive versus non-interactive locally, differentially private estimation: Two elbows for the quadratic functional. *Ann. Statist.* **51** 464–486. MR4600989 https://doi.org/10.1214/22-aos2254

Cai, B., Daskalakis, C. and Kamath, G. (2017). Priv'IT: Private and sample efficient identity testing. *Int. Conf. Mach. Learn.* **34**.

Cai, T.T., Wang, Y. and Zhang, L. (2020). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *Ann. Statist.* **49** 2825–2850. MR4338894 https://doi.org/10.1214/21-aos2058

Cai, T.T., Wang, Y. and Zhang, L. (2023). Score attack: a lower bound technique for optimal differentially private learning. Preprint. Available at arXiv:2303.07152.

Canonne, C.L., Kamath, G., McMillan, A., Smith, A. and Ullman, J. (2019). The structure of optimal private tests for simple hypotheses. In *Proc. Annu. ACM Symp. Theory Comput.* MR4003341

Canonne, C.L., Kamath, G., McMillan, A., Ullman, J. and Zakynthinou, L. (2020). Private identity testing for high-dimensional distributions. *Adv. Neural Inf. Process. Syst.* **34**.

Chen, W. and Özgür, A. (2024). L_q lower bounds on distributed estimation via Fisher information. In *Proc. ISIT.*

Couch, S., Kazan, Z., Shi, K., Bray, A. and Groce, A. (2019). Differentially private nonparametric hypothesis testing. In *Proc. ACM Conf. Comput. Commun. Secur.*

D'Orazio, V., Honaker, J. and King, G. (2018). Differential privacy for social science inference. Sloan Foundation Economics Research Paper.

Dubois, A., Berrett, T.B. and Butucea, C. (2022). Goodness-of-fit testing for Holder continuous densities under local differential privacy. In *Foundations of Modern Statistics - Festschrift in Honor of Vladimir Spokoiny*. MR4647367 https://doi.org/10.1007/978-3-031-30114-8\protect \relax \T1\textunderscore 2

Duchi, J.C., Jordan, M.I. and Wainwright, M.J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

Dwork, C. (2006). Differential privacy. In *Proc. ICALP*. MR2307219 https://doi.org/10.1007/11787006\protect \relax \T1\textunderscore 1

Dwork, C. and Lei, J. (2009). Differential privacy and robust statistics. In *Proc. Annu. ACM Symp. Theory Comput.* MR2780083

Dwork, C. and Roth, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–407. MR3254020 https://doi.org/10.1561/0400000042

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proc. TCC*. MR2241676 https://doi.org/10.1007/11681878\protect \relax \T1\textunderscore 14

Efron, B. (1967). The power of the likelihood-ratio test. *Ann. Math. Statist.* **38** 802–806. MR0212935 https://doi.org/10.1214/aoms/1177698874

Efron, B. and Truax, D. (1968). Large deviations theory in exponential families. *Ann. Math. Statist.* **39** 1402–1424. MR0234552 https://doi.org/10.1214/aoms/1177698121

Ferguson, T.S. (2002). *A Course in Large Sample Theory*. Chapman and Hall/CRC. MR1699953 https://doi.org/10.1007/978-1-4899-4549-5

Fienberg, S.E., Rinaldo, A. and Yang, G. (2010). Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In *Proc. PSD*.

Fienberg, S.E., Slavković, A. and Uhler, C. (2011). Privacy preserving GWAS data sharing. In *Proc. IEEE Int. Conf. Data Min.* MR2324074 https://doi.org/10.1214/088342306000000240

Gaboardi, M. and Rogers, R. (2018). Local private hypothesis testing: Chi-square tests. *Int. Conf. Mach. Learn.* **35**.

Gaboardi, M., Lim, H., Rogers, R. and Vdhan, S. (2016). Differentially private chi-squared hypothesis testing: Goodness-of-fit and independence testing. *Int. Conf. Mach. Learn.* **33**.

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 657–683. MR1185217

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408. MR0173322 https://doi.org/10.1214/aoms/1177700150

Hoeffding, W. (1967). On probabilities of large deviations. *Proc. Berkeley Symp. Math. Stat. Probab.* **5** 203–219. MR0212912

Johnson, A. and Shmatikov, V. (2013). Privacy-preserving data exploration in genome-wide association studies. In *Proc. IEEE Int. Conf. Data Min.* MR4433396 https://doi.org/10.1109/tsp.2022.3169432

Kairouz, P., Oh, S. and Viswanath, P. (2016). Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.* **17** 1–51. MR3491111

Kakizaki, K., Fukuchi, K. and Sakuma, J. (2017). Differentially private chi-squared test by unit circle mechanism. *Int. Conf. Mach. Learn.* **34**.

Kifer, D. and Rogers, R. (2017). A new class of private chi-square tests. *Int. Conf. Artif. Intell. Stat.* **20**.

Lam-Weil, J., Laurent, B. and Loubes, J.M. (2020). Minimax optimal goodness-of-fit testing for densities under a local differential privacy constraint. *Bernoulli* **28** 579–600. MR4337717 https://doi.org/10.3150/21-bej1358

Li, M., Berrett, T.B. and Yu, Y. (2022). On robustness and local differential privacy. *Ann. Statist.* **51** 717–737. MR4600999 https://doi.org/10.1214/23-aos2267

McSherry, F. and Talwar, K. (2007). Mechanism design via differential privacy. In *Proc. FOCS*.

Osipov, L.V. (1982). Large deviation probabilities of sums of independent random vectors for certain classes of sets. *Math. Notes* **31** 75–79. MR0646921

Pensia, A., Asadi, A.R., Jog, V. and Loh, P. (2025). Simple binary hypothesis testing under local differential privacy and communication constraints. *IEEE Trans. Inf. Theory* **71** 592–617. MR4843539 https://doi.org/10.1109/tit.2024.3461361

Petrov, V.V. (1975). *Sum of Independent Random Variables*. Springer. MR0388499

Rao, C.R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **24** 46–72. MR0293766

Rinott, Y., O'Keefe, C.M., Shlomo, N. and Skinner, C. (2017). Confidentiality and differential privacy in the dissemination of frequency tables. *Statist. Sci.* **33** 358–385. MR3843381 https://doi.org/10.1214/17-STS641

Saulis, L. (1983). Large deviations for random vectors for certain classes of sets. *Lith. Math. J.* **23** 308–317. MR0737667

Saulis, L. and Statulevicius, V.A. (1991). *Limit Theorems for Large Deviations*. Kluwer. MR1171883 https://doi.org/10.1007/978-94-011-3530-6

Sheffet, O. (2018). Locally private hypothesis testing. *Int. Conf. Artif. Intell. Stat.* **35**.

Steinberger, L. (2024). Efficiency in local differential privacy. *Ann. Statist.* **52** 2139–2166. MR4829483 https://doi.org/10.1214/24-aos2425

Suresh, A.T. (2021). Robust hypothesis testing and distribution estimation in Hellinger distance. *Int. Conf. Artif. Intell. Stat.* **24**.

Uhlerop, C., Slavković, A. and Fienberg, S.E. (2013). Privacy-preserving data sharing for genome-wide association studies. *J. Priv. Confid.* **5** 137–166.

von Bahr, B. (1967). Multi-dimensional integral limit theorems. *Arch. Math.* **7** 71–88. MR0228044 https://doi.org/10.1007/BF02591678

Vu, D. and Slavković, A. (2009). Differential privacy for clinical trial data: Preliminary evaluations. In *Proc. IEEE Int. Conf. Data Min.*

Wang, Y., Lee, J. and Kifer, D. (2015). Revisiting differentially private hypothesis tests for categorical data. Preprint. Available at arXiv:1511.03376.

Wasserman, L. and Zhou, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

Wilks, S.S. (1938). The large-sample distribution of the likelihood-ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.

Bernoulli **32**(1), 2026, 567–589
https://doi.org/10.3150/25-BEJ1870

Rates of convergence to the local time of oscillating and skew Brownian motion

SARA MAZZONETTO[a]

Université de Lorraine, CNRS, Inria, IECL, Nancy, France, [a]*sara.mazzonetto@univ-lorraine.fr*

In this paper, a class of statistics based on high frequency observations of oscillating and skew Brownian motion is considered. Their convergence rate towards the local time of the underlying process is obtained in form of a functional limit theorem. Oscillating and skew Brownian motions are solutions to stochastic differential equations with *singular* coefficients: piecewise constant diffusion coefficient or additive local time finite variation term. The result is applied to provide estimators of the skewness parameter and study their asymptotic behavior, and diffusion coefficient estimation is discussed as well. Moreover, in the case of the classical statistics given by the normalized number of crossings, the result is proved to hold for a larger class of Itô processes with singular coefficients. Up to our knowledge, this is the first result proving the convergence rates for estimators of the skewness parameter of skew Brownian motion.

Keywords: Central limit theorem; functional limit theorems; local time; oscillating Brownian motion; parameter estimation; skew Brownian motion; threshold diffusions

References

[1] Aldous, D.J. and Eagleson, G.K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331. MR0517416 https://doi.org/10.1214/aop/1176995577
[2] Altmeyer, R. (2021). Approximation of occupation time functionals. *Bernoulli* **27** 2714–2739. MR4303901 https://doi.org/10.3150/21-BEJ1328
[3] Athreya, S., Butkovsky, O., Lê, K. and Mytnik, L. (2024). Well-posedness of stochastic heat equation with distributional drift and skew stochastic heat equation. *Comm. Pure Appl. Math.* **77** 2708–2777. MR4720220 https://doi.org/10.1002/cpa.22157
[4] Azaïs, J.-M. (1989). Approximation des trajectoires et temps local des diffusions. (Approximation of trajectories and local times of diffusions). *Ann. Inst. Henri Poincaré Probab. Stat.* **25** 175–194. MR1001025
[5] Azaïs, J.-M. (1991). Convergence presque sûre du nombre de franchissements normalisé vers le temps local (Almost sure convergence of the normalized crossings number to local time). *C. R. Acad. Sci. Paris Sér. I Math.* **312** 689–694. MR1105627
[6] Bardou, O. and Martinez, M. (2010). Statistical estimation for reflected skew processes. *Stat. Inference Stoch. Process.* **13** 231–248. MR2729650 https://doi.org/10.1007/s11203-010-9047-6
[7] Barlow, M.T. (1988). Skew Brownian motion and a one dimensional stochastic differential equation. *Stochastics* **25** 1–2. MR1008231 https://doi.org/10.1080/17442508808833528
[8] Bass, R.F. and Chen, Z.-Q. (2005). One-dimensional stochastic differential equations with singular and degenerate coefficients. *Sankhyā* **67** 19–45. MR2203887
[9] Borodin, A.N. (1986). On the character of convergence to Brownian local time. I. *Probab. Theory Related Fields* **72** 231–250. MR0836277 https://doi.org/10.1007/BF00699105
[10] Borodin, A.N. (1989). Brownian local time. *Russian Math. Surveys* **44** 1–51. MR0998360 https://doi.org/10.1070/RM1989v044n02ABEH002050
[11] Borodin, A. and Salminen, P. (2019). On the local time process of a skew Brownian motion. *Trans. Amer. Math. Soc.* **372** 3597–3618. MR3988620 https://doi.org/10.1090/tran/7852
[12] Bounebache, S.K. and Zambotti, L. (2014). A skew stochastic heat equation. *J. Theor. Probab.* **27** 168–201. MR3174222 https://doi.org/10.1007/s10959-012-0421-8

1350-7265

[13] Decamps, M., Goovaerts, M. and Schoutens, W. (2006). Self exciting threshold interest rates models. *Int. J. Theor. Appl. Finance* **9** 1093–1122. MR2269906 https://doi.org/10.1142/S0219024906003937
[14] Ding, K., Cui, Z. and Wang, Y. (2021). A Markov chain approximation scheme for option pricing under skew diffusions. *Quant. Finance* **21** 461–480. MR4213546 https://doi.org/10.1080/14697688.2020.1781235
[15] Étoré, P. and Martinez, M. (2013). Exact simulation of one-dimensional stochastic differential equations involving the local time at zero of the unknown process. *Monte Carlo Methods Appl.* **19** 41–71. MR3039402 https://doi.org/10.1515/mcma-2013-0002
[16] Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804. MR1242012 https://doi.org/10.2307/3214513
[17] Forien, R. (2019). Gene flow across geographical barriers – scaling limits of random walks with obstacles. *Stochastic Process. Appl.* **129** 3748–3773. MR3997660 https://doi.org/10.1016/j.spa.2018.10.006
[18] Gairat, A. and Shcherbakov, V. (2017). Density of skew Brownian motion and its functionals with application in finance. *Math. Finance* **27** 1069–1088. MR3705163 https://doi.org/10.1111/mafi.12120
[19] Hajri, H. (2015). On flows associated to Tanaka's SDE and related works. *Electron. Commun. Probab.* **20** 16. MR3314651 https://doi.org/10.1214/ECP.v20-4058
[20] Harrison, J.M. and Shepp, L.A. (1981). On skew Brownian motion. *Ann. Probab.* **9** 309–313. MR0606993 https://doi.org/10.1214/aop/1176994472
[21] Jacod, J. (1998). Rates of convergence to the local time of a diffusion. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 505–544. MR1632849 https://doi.org/10.1016/S0246-0203(98)80026-5
[22] Jacod, J. and Protter, P. (2012). Discretization of processes. In *Stoch. Model. Appl. Probab.* **67**. Berlin: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7
[23] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren Math. Wiss.* **288**. Berlin: Springer. MR1943877 https://doi.org/10.1007/978-3-662-05265-5
[24] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Grad. Texts Math.* **113**. New York: Springer. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2
[25] Keilson, J. and Wellner, J.A. (1978). Oscillating Brownian motion. *J. Appl. Probab.* **15** 300–310. MR0474526 https://doi.org/10.2307/3213403
[26] Le Gall, J.F. (1984). One-dimensional stochastic differential equations involving the local times of the unknown process. In *Stochastic Analysis and Applications*. *Proc. Int. Conf., Swansea 1983, Lect. Notes Math.* **1095** 51–82. MR0777514 https://doi.org/10.1007/BFb0099122
[27] Le Gall, J.F. and Yor, M. (1986). Étude asymptotique de certains mouvements browniens complexes avec drift. *Probab. Theory Related Fields* **71** 183–229. MR0816704 https://doi.org/10.1007/BF00332310
[28] Lejay, A. (2006). On the constructions of the skew Brownian motion. *Probab. Surv.* **3** 413–466. MR2280299 https://doi.org/10.1214/154957807000000013
[29] Lejay, A. (2018). Estimation of the bias parameter of the skew random walk and application to the skew Brownian motion. *Stat. Inference Stoch. Process.* **21** 539–551. MR3846991 https://doi.org/10.1007/s11203-017-9161-9
[30] Lejay, A. and Mazzonetto, S. (2024). Maximum likelihood estimator for skew Brownian motion: The convergence rate. *Scand. J. Stat.* **51** 612–642. MR4764756 https://doi.org/10.1111/sjos.12694
[31] Lejay, A., Mordecki, E. and Torres, S. (2014). Is a Brownian motion skew? *Scand. J. Stat.* **41** 346–364. MR3207175 https://doi.org/10.1111/sjos.12033
[32] Lejay, A., Mordecki, E. and Torres, S. (2019). Two consistent estimators for the skew Brownian motion. *ESAIM, Probab. Stat.* **23** 567–583. MR3991913 https://doi.org/10.1051/ps/2018018
[33] Lejay, A. and Pichot, G. (2016). Simulating diffusion processes in discontinuous media: Benchmark tests. *J. Comput. Phys.* **314** 384–413. MR3484940 https://doi.org/10.1016/j.jcp.2016.03.003
[34] Lejay, A. and Pigato, P. (2018). Statistical estimation of the oscillating Brownian motion. *Bernoulli* **24** 3568–3602. MR3788182 https://doi.org/10.3150/17-BEJ969
[35] Lejay, A. and Pigato, P. (2019). A threshold model for local volatility: Evidence of leverage and mean reversion effects on historical data. *Int. J. Theor. Appl. Finance* **22** 1950017. MR3969553 https://doi.org/10.1142/S0219024919500171
[36] Lejay, A. and Pigato, P. (2020). Maximum likelihood drift estimation for a threshold diffusion. *Scand. J. Stat.* **47** 609–637. MR4157153 https://doi.org/10.1111/sjos.12417

[37] Lipton, A. and Sepp, A. (2011). Filling the gaps. Risk Magazine, October 66–71.
[38] Mazzonetto, S. (2026). Supplement to "Rates of convergence to the local time of Oscillating and Skew Brownian Motion." https://doi.org/10.3150/25-BEJ1870SUPP
[39] Mazzonetto, S. and Pigato, P. (2024). Drift estimation of the threshold Ornstein-Uhlenbeck process from continuous and discrete observations. *Statist. Sinica* **34** 313–336. MR4683574 https://doi.org/10.5705/ss.202021.0275
[40] Podolskij, M. and Rosenbaum, M. (2017). Comment on: Limit of random measures associated with the increments of a Brownian semimartingale: Asymptotic behavior of local times related statistics for fractional Brownian motion. *J. Financ. Econom.* **16** 588–598.
[41] Portenko, N.I. (1994). The development of I. I. Gikham's idea concerning the methods for investigating local behavior of diffusion processes and their weakly convergent sequences. *Theory Probab. Math. Statist.* **50** 6–21. MR1445514
[42] Ramirez, J.M. (2007). Skew Brownian motion and branching processes applied to diffusion-advection in heterogeneous media and fluid flow. In *ProQuest LLC* State University. Ann Arbor. MI Thesis (Ph.D.)–Oregon. MR2710544
[43] Rényi, A. (1963). On stable sequences of events. *Sankhyā Ser. A* **25** 293–302. MR0170385
[44] Robert, C.Y. (2023). How large is the jump discontinuity in the diffusion coefficient of a time-homogeneous diffusion? *Econom. Theory* **39** 848–880. MR4623502 https://doi.org/10.1017/S0266466622000214
[45] Roberts, G.O., Rosenthal, J.S. and Tawn, N.G. (2022). Skew Brownian motion and complexity of the ALPS algorithm. *J. Appl. Probab.* **59** 777–796. MR4480080 https://doi.org/10.1017/jpr.2021.78
[46] Su, F. and Chan, K.-S. (2015). Quasi-likelihood estimation of a threshold diffusion process. *J. Econometrics* **189** 473–484. MR3414915 https://doi.org/10.1016/j.jeconom.2015.03.038
[47] Taranto, A., Addie, R. and Khan, S. (2021). Bi-Directional Grid Constrained Stochastic Processes' Link to Multi-Skew Brownian Motion. Preprint, arXiv:2107.12554 [math.PR].
[48] Walsh, J.B. (1978). A diffusion with discontinuous local time. In *Temps Locaux 52-53* 37–45. Société Mathématique de France.

Bernoulli **32**(1), 2026, 590–614
https://doi.org/10.3150/25-BEJ1871

Series ridge regression for spatial data on $\mathbb{R}^d$

DAISUKE KURISU[1,a] and YASUMASA MATSUDA[2,b]

[1] *Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan,* [a]*daisukekurisu@csis.u-tokyo.ac.jp*
[2] *Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan,* [b]*yasumasa.matsuda.a4@tohoku.ac.jp*

This paper develops a general asymptotic theory of series estimators for spatial data collected at irregularly spaced locations within a sampling region $R_n \subset \mathbb{R}^d$. We employ a stochastic sampling design that can flexibly generate irregularly spaced sampling sites, encompassing both pure increasing and mixed increasing domain frameworks. Specifically, we focus on a spatial trend regression model and a nonparametric regression model with spatially dependent covariates. For these models, we investigate L^2-penalized series estimation of the trend and regression functions. We establish uniform and L^2 convergence rates and multivariate central limit theorems for general series estimators as main results. Additionally, we show that spline and wavelet series estimators achieve optimal uniform and L^2 convergence rates and propose methods for constructing confidence intervals for these estimators. Finally, we demonstrate that our dependence structure conditions on the underlying spatial processes cover a broad class of random fields, including Lévy-driven continuous autoregressive and moving average random fields.

Keywords: Irregularly spaced spatial data; Lévy-driven moving average random field; series ridge regression; spatial regression model

References

Bandyopadhyay, S., Lahiri, S.N. and Nordman, D.J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.* **43** 519–545. MR3316189 https://doi.org/10.1214/14-aos1291

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd Edition ed. Chapman and Hall/CRC. MR3362184

Barnett, L.A.K., Ward, E.J. and Anderson, S.C. (2021). Improving estimates of species distribution change by incorporating local trends. *Ecography* **44** 427–439.

Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press. MR1406564

Bradley, R.C. (1989). A caution on mixing conditions for random fields. *Statist. Probab. Lett.* **8** 489–491. MR1040812 https://doi.org/10.1016/0167-7152(89)90032-1

Bradley, R.C. (1993). Some examples of mixing random fields. *Rocky Mountain J. Math.* **23** 495–519. MR1226185 https://doi.org/10.1216/rmjm/1181072573

Brockwell, P.J. and Matsuda, Y. (2017). Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 833–857. MR3641410 https://doi.org/10.1111/rssb.12197

Brunsdon, C., Fotheringham, A.S. and Charlton, M.E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **28** 281–298.

Chan, N.H., Zhang, R. and Yau, C.Y. (2022). Inference for structural breaks in spatial models. *Statist. Sinica* **32** 1961–1981. MR4478185

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

Chen, X. and Christensen, T.M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *J. Econometrics* **188** 447–465. MR3383220 https://doi.org/10.1016/j.jeconom.2015.03.010

Chen, X., Liao, Y. and Weichen, W. (2024). Inference on time series nonparametric conditional moment restrictions using general sieves. *J. Econometrics*. forthcoming. MR4906007 https://doi.org/10.1016/j.jeconom.2024.105920

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. MR1256527 https://doi.org/10.1006/acha.1993.1005

Cressie, N. (2015). *Statistics for Spatial Data*, 2nd Edition ed. John Wiley & Sons. MR3559472

De Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer. MR1900298

De Jong, R.M. (2002). A note on "Convergence rates and asymptotic normality for series estimators": Uniform convergence rates. *J. Econometrics* **111** 1–9. MR1926535 https://doi.org/10.1016/S0304-4076(02)00113-6

DeVore, R.A. and Lorentz, G.G. (1993). Constructive Approximation 303. Springer. MR1261635

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

El Machkouri, M., Es-Sebaiy, K. and Ouassou, I. (2017). On local linear regression for strongly mixing random fields. *J. Multivariate Anal.* **156** 103–115. MR3624688 https://doi.org/10.1016/j.jmva.2017.02.002

El Machkouri, M. and Stoica, R. (2010). Asymptotic normality of kernel estimates in a regression model for random fields. *J. Nonparametr. Stat.* **22** 955–971. MR2738877 https://doi.org/10.1080/10485250903505893

Gaetan, C. and Guyan, X. (2010). *Spatial Statistics and Modeling*. Springer. MR2569034 https://doi.org/10.1007/978-0-387-92257-7

Gollini, I., Lu, B., Charlton, M., Brunsdon, C. and Harris, P. (2015). GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.* **63** 1–50.

Grenander, U. (1981). *Abstract Inference*. New York: Wiley. MR0599175

Gromenko, O., Kokoszka, P. and Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 29–50. MR3597963 https://doi.org/10.1111/rssb.12156

Hallin, M., Lu, Z. and Tran, L.T. (2004). Local linear spatial regression. *Ann. Statist.* **32** 2469–2500. MR2153992 https://doi.org/10.1214/009053604000000850

Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli* **15** 659–686. MR2555194 https://doi.org/10.3150/08-bej168

Huang, J.Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.* **26** 242–272. MR1611780 https://doi.org/10.1214/aos/1030563984

Huang, J.Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635. MR2012827 https://doi.org/10.1214/aos/1065705120

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *J. Econometrics* **167** 224–239. MR2885448 https://doi.org/10.1016/j.jeconom.2011.11.008

Johnstone, I.M. (2017). Gaussian estimation: Sequence and wavelet models. Manuscript.

Kato, T. (1995). *Perturbation Theory for Linear Operators 132*. Springer. MR1335452

Kato, K. and Kurisu, D. (2020). Bootstrap confidence bands for spectral estimation of Lévy densities under high-frequency observations. *Stochastic Process. Appl.* **130** 1159–1205. MR4058270 https://doi.org/10.1016/j.spa.2019.04.012

Kurisu, D. (2019). On nonparametric inference for spatial regression models under domain expanding and infill asymptotics. *Statist. Probab. Lett.* **154** 1108543. MR3980502 https://doi.org/10.1016/j.spl.2019.06.019

Kurisu, D. (2022). Nonparametric regression for locally stationary random fields under stochastic sampling design. *Bernoulli* **28** 1250–1275. MR4388937 https://doi.org/10.3150/21-bej1385

Kurisu, D., Kato, K. and Shao, X. (2024). Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data. *J. Amer. Statist. Assoc.* **119** 1820–1832. MR4797904 https://doi.org/10.1080/01621459.2023.2218578

Kurisu, D. and Matsuda, Y. (2024). Local polynomial trend regression for spatial data on $\mathbb{R}^d$. *Bernoulli* **30** 2770–2794. MR4779848 https://doi.org/10.3150/23-bej1694

Kurisu, D. and Matsuda, Y. (2026). Supplement to "Series ridge regression for spatial data on $\mathbb{R}^d$." https://doi.org/10.3150/25-BEJ1871SUPP

Lahiri, S.N. (1996). On inconsistency of estimators under infill asymptotics for spatial data. *Sankhya A* **58** 403–417. MR1659130

Lahiri, S.N. (2003a). Central limit theorems for weighted sum of a spatial process under a class of stochastic and fixed design. *Sankhyā* **65** 356–388. MR2028905

Lahiri, S.N. (2003b). *Resampling Methods for Dependent Data*. Springer. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

Lahiri, S.N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34** 1774–1813. MR2283717 https://doi.org/10.1214/009053606000000551

Lee, J. and Robinson, P.M. (2016). Series estimation under cross-sectional dependence. *J. Econometrics* **190** 1–17. MR3425269 https://doi.org/10.1016/j.jeconom.2015.08.001

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

Lu, Z. and Chen, X. (2004). Spatial kernel regression estimation: Weak consistency. *Statist. Probab. Lett.* **68** 125–136. MR2066167 https://doi.org/10.1016/j.spl.2003.08.014

Lu, Z. and Tjøstheim, D. (2014). Nonparametric estimation of probability density functions for irregularly observed spatial data. *J. Amer. Statist. Assoc.* **109** 1546–1564. MR3293609 https://doi.org/10.1080/01621459.2014.947376

Lu, B., Charlton, M., Harris, P. and Fotheringham, A.S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *Int. J. Geogr. Inf. Sci.* **28** 660–681.

Martínez-Hernández, I. and Genton, M.G. (2020). Recent developments in complex and spatially correlated functional data. *Braz. J. Probab. Stat.* **34** 204–229. MR4093256 https://doi.org/10.1214/20-BJPS466

Matsuda, Y. and Yajima, Y. (2018). Locally stationary spatio-temporal processes. *Jpn. J. Stat. Data Sci.* **1** 41–57. MR4010751 https://doi.org/10.1007/s42081-018-0003-9

Meyer, Y. (1992). *Wavelets and Operators 37*. Cambridge University Press. MR1228209

Newey, W.K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. MR1457700 https://doi.org/10.1016/S0304-4076(97)00011-0

Rajput, B.S. and Rosinski, J. (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82** 451–487. MR1001524 https://doi.org/10.1007/bf00339998

Robinson, P.M. (2011). Asymptotic theory for nonparametric regression with spatial data. *J. Econometrics* **165** 5–19. MR2842798 https://doi.org/10.1016/j.jeconom.2011.05.002

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press. MR1739520

Schumaker, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press. MR2348176 https://doi.org/10.1017/CBO9780511618994

Shen, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591. MR1604416 https://doi.org/10.1214/aos/1030741085

Sherwood, S.C. (2007). Simultaneous detection of climate change and observing biases in a network with incomplete sampling. *J. Climate* **20** 4047–4062.

Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

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

Tropp, J.A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94–116. MR1258867 https://doi.org/10.1214/aop/1176988849

Bernoulli **32**(1), 2026, 615–637
https://doi.org/10.3150/25-BEJ1872

A CLT for the difference of eigenvalue statistics of sample covariance matrices

NINA DÖRNEMANN[1,a] and HOLGER DETTE[2,b]

[1] *Department of Mathematics, Aarhus University, Aarhus, Denmark*, [a] *ndoernemann@math.au.dk*
[2] *Department of Mathematics, Ruhr University Bochum, Bochum, Germany*, [b] *holger.dette@ruhr-uni-bochum.de*

In the case where the dimension of the data grows at the same rate as the sample size, we prove a central limit theorem for the difference of a linear spectral statistic of the sample covariance and a linear spectral statistic of the matrix that is obtained from the sample covariance matrix by deleting a column and the corresponding row. Unlike previous works, we do neither require that the population covariance matrix is diagonal nor that moments of all order exist. Our proof methodology incorporates subtle enhancements to existing strategies, which meet the challenges introduced by determining the mean and covariance structure for the difference of two such eigenvalue statistics. Moreover, we also establish the asymptotic independence of the difference-type spectral statistic and the usual linear spectral statistic of sample covariance matrices.

Keywords: Central limit theorem; linear spectral statistic; sample covariance matrix

References

Ahlfors, L.V. (1953). *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable 177*. New York, London: McGraw-Hill. MR0054016

Bai, Z. and Silverstein, J.W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices 20*. Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

Bai, Z.D. and Yin, Y.Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083 https://doi.org/10.1214/aop/1176991792

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

Baik, J., Ben Arous, G. and Pèche, 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

Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

Bao, Z. (2012). Strong convergence of ESD for the generalized sample covariance matrices when $p/n \to 0$. *Statist. Probab. Lett.* **82** 894–901. MR2910035 https://doi.org/10.1016/j.spl.2012.01.012

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

Chen, B. and Pan, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. MR3338658 https://doi.org/10.3150/14-BEJ599

Cipolloni, G. and Erdős, L. (2020). Fluctuations for differences of linear eigenvalue statistics for sample covariance matrices. *Random Matrices Theory Appl.* **9** 2050006. MR4119592 https://doi.org/10.1142/S2010326320500069

Dörnemann, N. (2022). Asymptotics for linear spectral statistics of sample covariance matrices. Dissertation, Ruhr-Universität Bochum. https://doi.org/10.13154/294-8999

Dörnemann, N. and Dette, H. (2023). Fluctuations of the diagonal entries of a large sample precision matrix. *Statist. Probab. Lett.* **198** 109838. MR4568824 https://doi.org/10.1016/j.spl.2023.109838

Dörnemann, N. and Dette, H. (2024). Linear spectral statistics of sequential sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 946–970. MR4757513 https://doi.org/10.1214/22-aihp1339

Dörnemann, N. and Dette, H. (2026). Supplement to "A CLT for the difference of eigenvalue statistics of sample covariance matrices." https://doi.org/10.3150/25-BEJ1872SUPP

Erdős, L. and Schröder, D. (2018). Fluctuations of rectangular Young diagrams of interlacing Wigner eigenvalues. *Int. Math. Res. Not.* **2018** 3255–3298. MR3805203 https://doi.org/10.1093/imrn/rnw330

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

Li, Z., Han, F. and Yao, J. (2020). Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model. *Ann. Statist.* **48** 3138–3160. MR4185803 https://doi.org/10.1214/aos/1009210544

Liu, H., Aue, A. and Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140 https://doi.org/10.1214/14-AOS1294

Marčenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457. https://doi.org/10.1070/SM1967v001n04ABEH001994

Mei, T., Wang, C. and Yao, J. (2023). On singular values of data matrices with general independent columns. *Ann. Statist.* **51** 624–645. MR4600995 https://doi.org/10.1214/23-aos2263

Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.* **26** 1837–1887. MR3513608 https://doi.org/10.1214/15-AAP1135

Pan, G. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865 https://doi.org/10.5705/ss.2006.161

Qiu, J., Li, Z. and Yao, J. (2023). Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when $p/n \to \infty$ and applications. *Ann. Statist.* **51** 1427–1451. MR4630955 https://doi.org/10.1214/23-aos2300

Silverstein, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408 https://doi.org/10.1006/jmva.1995.1083

Silverstein, J.W. and Bai, Z. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534 https://doi.org/10.1006/jmva.1995.1051

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444 https://doi.org/10.1023/A:1019739414239

Van Der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

Wang, L., Aue, A. and Paul, D. (2017). Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions. *Bernoulli* **23** 2181–2209. MR3648029 https://doi.org/10.3150/16-BEJ818

Wang, L. and Paul, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when p/n→ 0. *J. Multivariate Anal.* **126** 25–52. MR3173080 https://doi.org/10.1016/j.jmva.2013.12.015

Zhang, Z., Zheng, S., Pan, G. and Zhong, P.-S. (2022). Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices. *Ann. Statist.* **50** 2205–2230. MR4474488 https://doi.org/10.1214/22-aos2183

Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

Bernoulli **32**(1), 2026, 638–663
https://doi.org/10.3150/25-BEJ1873

Faithlessness in Gaussian graphical models

MATHIAS DRTON[1,a], LEONARD HENCKEL[2,d], BENJAMIN HOLLERING[1,b] and PRATIK MISRA[1,c]

[1] *Department of Mathematics, Technical University of Munich, Garching b. München, Germany,* [a] *mathias.drton@tum.de,* [b] *benjamin.hollering@tum.de,* [c] *pratik.misra@tum.de*
[2] *School of Mathematics and Statistics, University College Dublin, Dublin, Ireland,* [d] *leonard.henckel@ucd.ie*

The implication problem for conditional independence (CI) asks whether the fact that a probability distribution obeys a given finite set of CI relations implies that a further CI statement also holds in this distribution. This problem has a long and fascinating history, cumulating in positive results about implications now known as the semigraphoid axioms as well as impossibility results about a general finite characterization of CI implications. Motivated by violation of faithfulness assumptions in causal discovery, we study the implication problem in the special setting where the CI relations are obtained from a directed acyclic graphical (DAG) model along with one additional CI statement. Focusing on the Gaussian case, we show that this implication problem reduces to a principal ideal membership problem. Using this we derive necessary and sufficient combinatorial criteria for when the set of faithless distributions decomposes into graphical components. Moreover, prompted by the relevance of strong faithfulness in statistical guarantees for causal discovery algorithms, we give a graphical solution for an approximate CI implication problem, in which we ask whether small values of one additional partial correlation entail small values for yet a further partial correlation.

Keywords: Conditional independence; faithfulness; Graphical models

References

[1] Arellano-Valle, R.B., Conteras-Reyes, J.E. and Genton, M.G. (2013). Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scand. J. Stat.* **40** 42–62. MR3024031 https://doi.org/10.1111/j.1467-9469.2011.00774.x
[2] Basu, S., Pollack, R. and Roy, M.-F.c. (2006). *Algorithms in Real Algebraic Geometry*, second ed. *Algorithms and Computation in Mathematics* **10**. Berlin: Springer-Verlag. MR2248869
[3] Boege, T. (2022). The Gaussian conditional independence inference problem.
[4] Boege, T. CInet::Tools — A software for computations with conditional independence structures. Available at: https://github.com/CInet.
[5] Boege, T., D'Alì, A., Kahle, T. and Sturmfels, B. (2019). The geometry of gaussoids. *Found. Comput. Math.* **19** 775–812. MR3989713 https://doi.org/10.1007/s10208-018-9396-x
[6] Boege, T., Kubjas, K., Misra, P. and Solus, L. (2024). Colored Gaussian DAG models.
[7] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. Wiley. MR2239987
[8] Cox, D.A., Little, J. and O'Shea, D. (2015). Ideals, varieties, and algorithms. In *An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed. *Undergraduate Texts in Mathematics*. Cham: Springer. MR3330490 https://doi.org/10.1007/978-3-319-16721-3
[9] Draisma, J., Sullivant, S. and Talaska, K. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. MR2662356 https://doi.org/10.1214/09-AOS760
[10] Draisma, J., Sullivant, S. and Talaska, K. (2013). Positivity for Gaussian graphical models. *Adv. in Appl. Math.* **50** 661–674. MR3044565 https://doi.org/10.1016/j.aam.2013.03.001
[11] Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Semin.* **39**. Basel: Birkhäuser. MR2723140 https://doi.org/10.1007/978-3-7643-8905-5
[12] Geiger, D. and Pearl, J. (1990). On the logic of causal models. In *Machine Intelligence and Pattern Recognition* **9** 3–14. Elsevier. MR1166825 https://doi.org/10.1016/B978-0-444-88650-7.50006-8

[13] Geiger, D. and Pearl, J. (1993). Logical and algorithmic properties of conditional independence and graphical models. *Ann. Statist.* **21** 2001–2021. MR1245778 https://doi.org/10.1214/aos/1176349407
[14] Geiger, D., Verma, T. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534. MR1064736 https://doi.org/10.1002/net.3230200504
[15] Henckel, L., Buttenschoen, M. and Maathuis, M.H. (2024). Graphical tools for selecting conditional instrumental sets. *Biometrika* **111** 771–788. MR4794149 https://doi.org/10.1093/biomet/asad066
[16] Henckel, L., Perković, E. and Maathuis, M.H. (2022). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 579–599. MR4412998
[17] Kenig, B. (2021). Approximate implication with d-separation. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence* 301–311.
[18] Kenig, B. and Suciu, D. (2022). Integrity constraints revisited: From exact to approximate implication. *Log. Methods Comput. Sci.* **18**. MR4363118 https://doi.org/10.46298/lmcs-18(1:5)2022
[19] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT press. MR2816736
[20] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. London: The Clarendon Press. MR1419991
[21] Lauritzen, S. and Sadeghi, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46** 2251–2278. MR3845017 https://doi.org/10.1214/17-AOS1618
[22] Maathuis, M.H., Drton, M., Lauritzen, S. and Wainwright, M. (2018). *Handbook of Graphical Models*, 1st ed. USA: CRC Press. MR3889064
[23] Pearl, J. (2009). *Causality*, Second ed. Cambridge University Press. MR2548166 https://doi.org/10.1017/CBO9780511803161
[24] Rotnitzky, A. and Smucler, E. (2020). Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J. Mach. Learn. Res.* **21** 188–1. MR4209474
[25] Runge, J. (2021). Necessary and sufficient graphical conditions for optimal adjustment sets in causal graphical models with hidden variables. In *Proceeding of the Thirty-Fifth Conference on Neural Information Processing Systems* 15762–15773.
[26] Sadeghi, K. (2017). Faithfulness of probability distributions and graphs. *J. Mach. Learn. Res.* **18** 1–29. MR3763782
[27] Sadeghi, K. and Soo, T. (2022). Conditions and assumptions for constraint-based causal structure learning. *J. Mach. Learn. Res.* **23**. MR4576694
[28] Spirtes, P. (2001). An anytime algorithm for causal inference. In *International Workshop on Artificial Intelligence and Statistics* 278–285. PMLR.
[29] Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*, Second ed. Cambridge, MA: MIT Press. MR1815675
[30] Studenỳ, M. (1992). Conditional independence relations have no finite complete characterization. In *Information Theory, Statistical Decision Functions and Random Processes*. *Transactions of the 11th Prague Conference Vol. B* 377–396.
[31] Studenỳ, M. (1997). Semigraphoids and structures of probabilistic conditional independence. *Ann. Math. Artif. Intell.* **21** 71–98. MR1479009 https://doi.org/10.1023/A:1018905100242
[32] Sullivant, S. (2009). Gaussian conditional independence relations have no finite complete characterization. *J. Pure Appl. Algebra* **213** 1502–1506. https://doi.org/10.1016/j.jpaa.2008.11.026
[33] Sullivant, S. (2018). *Algebraic Statistics*. *Graduate Studies in Mathematics* **194**. Providence, RI: American Mathematical Society. MR3838364 https://doi.org/10.1090/gsm/194
[34] Uhler, C., Raskutti, G., Bühlmann, P. and Yu, B. (2013). Geometry of the faithfulness assumption in causal inference. *Ann. Statist.* 436–463. MR3099109 https://doi.org/10.1214/12-AOS1080
[35] Van Der Zander, B., Liśkiewicz, M. and Textor, J. (2019). Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence* **270** 1–40. MR3901676 https://doi.org/10.1016/j.artint.2018.12.006
[36] Witte, J., Henckel, L., Maathuis, M.H. and Didelez, V. (2020). On efficient adjustment in causal graphs. *J. Mach. Learn. Res.* **21** 1–45. MR4213427
[37] Zhang, J. and Spirtes, P. (2002). Strong faithfulness and uniform consistency in causal inference. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence* 632–639.

Bernoulli **32**(1), 2026, 664–693
https://doi.org/10.3150/25-BEJ1874

Nonparametric curve estimation in measurement error problems with conditionally heteroscedastic variances

AURORE DELAIGLE[1,a], ALEXANDER MEISTER[2,b] and JIYANG ZHANG[3,c]

[1] *School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia,* [a]*aurored@unimelb.edu.au*
[2] *Institut für Mathematik, Universität Rostock, Rostock, Germany,* [b]*alexander.meister@uni-rostock.de*
[3] *School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia,* [c]*jiyanzhang@student.unimelb.edu.au*

We consider the problem of estimating the density f_X of a latent variable X using replicated observations contaminated by additive errors. Unlike the classical setting, these errors depend on X, which makes standard Fourier deconvolution techniques inappropriate, and we suggest instead a basis expansion approach. Using a basis of Legendre polynomials, we show that it is possible to express the unknown coefficients of the expansion in terms of invertible equations of moments of the observed data. We deduce a nonparametric estimator of f_X which attains optimal minimax convergence rates in the case where the errors are conditionally Gaussian. To implement our density estimator in practice, we introduce data-driven approaches for choosing a truncating parameter and computing an upper bound to the support of f_X. We illustrate the numerical performance of our estimator on simulated examples and on replicated data from the National Health and Nutrition Examination Survey. We show how to use the same ideas in related errors-in-variables regression problems and to estimate the variance function.

Keywords: Contaminated data; deconvolution; inverse problems; minimax convergence rates; non-standard errors-in-variables

References

[1] Belomestny, D. and Goldenshluger, A. (2020). Nonparametric density estimation from observations with multiplicative measurement errors. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 36–67. MR4058981 https://doi.org/10.1214/18-AIHP954

[2] Belomestny, D. and Panov, V. (2018). Semiparametric estimation in the normal variance-mean mixture model. *Statistics* **52** 571–589. MR3806565 https://doi.org/10.1080/02331888.2018.1425865

[3] Brenner Miguel, S., Comte, F. and Johannes, J. (2021). Spectral cut-off regularisation for density estimation under multiplicative measurement errors. *Electron. J. Stat.* **15** 3551–3573. MR4298979 https://doi.org/10.1214/21-ejs1870

[4] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11** 309–340. MR2132729 https://doi.org/10.3150/bj/1116340297

[5] Butt, N.S. and Khalil, M.G. (2020). A new bimodal distribution for modeling asymmetric bimodal heavy-tail real lifetime data. *Symmetry* **12** 2058.

[6] Carroll, R.J., Delaigle, A. and Hall, P. (2007). Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 859–878. MR2368574 https://doi.org/10.1111/j.1467-9868.2007.00614.x

[7] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR0997599

[8] Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica* **18** 1025–1045. MR2440402

[9] Delaigle, A. (2014). Nonparametric kernel methods with errors-in-variables: Constructing estimators, computing them, and avoiding common mistakes. *Aust. N. Z. J. Stat.* **56** 105–124. MR3226432 https://doi.org/10.1111/anzs.12066

[10] Delaigle, A., Fan, J. and Carroll, R.J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Statist. Assoc.* **104** 348–359. MR2504382 https://doi.org/10.1198/jasa.2009.0114

[11] Delaigle, A. and Gijbels, I. (2006). Data-driven boundary estimation in deconvolution problems. *Comput. Statist. Data Anal.* **50** 1965–1994. MR2225553 https://doi.org/10.1016/j.csda.2005.02.012

[12] Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statist. Sinica* **16** 773–788. MR2281301

[13] Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Amer. Statist. Assoc.* **103** 280–287. MR2394636 https://doi.org/10.1198/016214507000001355

[14] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685. MR2396811 https://doi.org/10.1214/009053607000000884

[15] Delaigle, A. and Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *J. Amer. Statist. Assoc.* **102** 1416–1426. MR2372541 https://doi.org/10.1198/016214507000000987

[16] Delaigle, A. and Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli* **14** 562–579. MR2544102 https://doi.org/10.3150/08-BEJ121

[17] Delaigle, A., Meister, A. and Zhang, J. (2026). Supplement to "Nonparametric curve estimation in measurement error problems with conditionally heteroscedastic variances." https://doi.org/10.3150/25-BEJ1874SUPP

[18] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

[19] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925. MR1245773 https://doi.org/10.1214/aos/1176349402

[20] Gassiat, E., Le Corff, S. and Lehéricy, L. (2022). Deconvolution with unknown noise distribution is possible for multivariate signals. *Ann. Statist.* **50** 303–323. MR4382018 https://doi.org/10.1214/21-aos2106

[21] Goldenshluger, A. (2002). Density deconvolution in the circular structural model. *J. Multivariate Anal.* **81** 360–375. MR1906385 https://doi.org/10.1006/jmva.2001.2015

[22] Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92** 135–148. MR2158615 https://doi.org/10.1093/biomet/92.1.135

[23] Hu, Y. and Schennach, S. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* **76** 195–216. MR2374986 https://doi.org/10.1111/j.0012-9682.2008.00823.x

[24] Jeon, J.M., Park, B.U. and Van Keilegom, I. (2022). Nonparametric regression on Lie groups with measurement errors. *Ann. Statist.* **50** 2973–3008. MR4500632 https://doi.org/10.1214/22-aos2218

[25] Kim, P.T. and Richards, D.P. (2011). Deconvolution density estimation on the space of positive definite symmetric matrices. In *Nonparametric Statistics and Mixture Models – a Festschrift in Honor of Thomas P Hettmansperger* 147–168. MR2838725 https://doi.org/10.1142/9789814340564\protect \relax \T1\textunderscore 0010

[26] Li, M., Ma, Y. and Li, R. (2019). Semiparametric regression for measurement error model with heteroscedastic error. *J. Multivariate Anal.* **171** 320–338. MR3900731 https://doi.org/10.1016/j.jmva.2018.12.012

[27] McIntyre, J. and Stefanski, L.A. (2011). Density estimation with replicate heteroscedastic measurements. *Ann. Inst. Statist. Math.* **63** 81–99. MR2748935 https://doi.org/10.1007/s10463-009-0220-x

[28] Meister, A. (2006). Support estimation via moment estimation in presence of noise. *Statistics* **40** 259–275. MR2236207 https://doi.org/10.1080/02331880600723101

[29] Meister, A. (2007). Deconvolving compactly supported densities. *Math. Methods Statist.* **16** 63–76. MR2319471 https://doi.org/10.3103/S106653070701005X

[30] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Berlin/Heidelberg/New York: Springer. MR2768576 https://doi.org/10.1007/978-3-540-87557-4

[31] [dataset] NHANES (2022). Dietary Interview - Total Nutrient Intakes (First Day, Second Day): Energy and Dietary Fiber. Hyattsville, MD: HHS, CDC. Available at http://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Dietary&Cycle=2017-2020.
[32] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627 https://doi.org/10.1214/aos/1017939249
[33] Sarkar, A., Mallick, B.K. and Carroll, R.J. (2014). Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics* **70** 823–834. MR3295743 https://doi.org/10.1111/biom.12197
[34] Sarkar, A., Mallick, B.K., Staudenmayer, J., Pati, D. and Carroll, R.J. (2014). Bayesian semiparametric density deconvolution in the presence of conditionally heteroscedastic measurement errors. *J. Comput. Graph. Statist.* **23** 1101–1125. MR3270713 https://doi.org/10.1080/10618600.2014.899237
[35] Sarkar, A., Pati, D., Chakraborty, A., Mallick, B.K. and Carroll, R.J. (2018). Bayesian semiparametric multivariate density deconvolution. *J. Amer. Statist. Assoc.* **113** 401–416. MR3803474 https://doi.org/10.1080/01621459.2016.1260467
[36] Staudenmayer, J., Ruppert, D. and Buonaccorsi, J.P. (2008). Density estimation in the presence of heteroscedastic measurement error. *J. Amer. Statist. Assoc.* **103** 726–736. MR2524005 https://doi.org/10.1198/016214508000000328
[37] Stefanski, L.A. and Carroll, R.J. (1990). Deconvolving kernel density estimators. *Statistics* **21** 169–184. MR1054861 https://doi.org/10.1080/02331889008802238
[38] Wand, M.P. and Jones, M.C. (1994). *Kernel Smoothing*. CRC press. MR1319818 https://doi.org/10.1007/978-1-4899-4493-1
[39] Wang, H. (2018). A new and sharper bound for Legendre expansion of differentiable functions. *Appl. Math. Lett.* **85** 95–102. MR3820286 https://doi.org/10.1016/j.aml.2018.05.022
[40] Yi, G., Delaigle, A. and Gustafson, P. (2021). *Handbook on Measurement Error Models*. CRC press.

Bernoulli **32**(1), 2026, 694–697
https://doi.org/10.3150/25-BEJ1875
Main article: https://doi.org/10.3150/10-BEJ329

Erratum: Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes

ROBERT STELZER[a,b]

Institute of Mathematical Finance, Ulm University, Helmholtzstrasse 18, D-89081 Ulm, Germany,
[a]*robert.stelzer@uni-ulm.de,* [b]*https://www.uni-ulm.de/mawi/finmath*

A serious flaw in the proof of the equivalence of continuous time state space models and MCARMA processes spotted in Fasen-Hartmann and Schenk (*J. Time Series Anal.* **46** (2025) 692–726) is corrected. We point out that likewise an issue in the proof of Theorem 3.2 in Brockwell and Schlemm (*J. Multivariate Anal.* **115** (2013) 217–251) can be resolved and, hence, any MCARMA process and linear state space model has both a controller and an observer canonical representation. Equivalently, the transfer function has both a left and right matrix fraction representation.

Keywords: Multivariate CARMA process; state space representation

References

Brockett, R.W. (2015). *Finite Dimensional Linear Systems*. *Classics in Applied Mathematics* **74**. Philadelphia: Society for Industrial and Applied Mathematics (SIAM). Reprint of the 1970 original. MR3486166 https://doi.org/10.1137/1.9781611973884

Brockwell, P.J. and Schlemm, E. (2013). Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *J. Multivariate Anal.* **115** 217–251. MR3004556 https://doi.org/10.1016/j.jmva.2012.09.004

Caines, P.E. (2018). *Linear Stochastic Systems*. *Classics in Applied Mathematics* **77**. Philadelphia: Society for Industrial and Applied Mathematics (SIAM). Reprint of the 1988 original. MR3908672

Fasen, V. and Kimmig, S. (2017). Information criteria for multivariate CARMA processes. *Bernoulli* **23** 2860–2886. MR3648048 https://doi.org/10.3150/16-BEJ830

Fasen-Hartmann, V. and Kimmig, S. (2020). Robust estimation of stationary continuous-time ARMA models via indirect inference. *J. Time Series Anal.* **41** 620–651. MR4176167 https://doi.org/10.1111/jtsa.12526

Fasen-Hartmann, V. and Mayer, C. (2022). Whittle estimation for continuous-time stationary state space models with finite second moments. *Ann. Inst. Statist. Math.* **74** 233–270. MR4396674 https://doi.org/10.1007/s10463-021-00802-6

Fasen-Hartmann, V. and Schenk, L. (2025). Mixed orthogonality graphs for continuous-time state space models and orthogonal projections. *J. Time Series Anal.* **46** 692–726. MR4920151 https://doi.org/10.1111/jtsa.12787

Fasen-Hartmann, V. and Scholz, M. (2019). Quasi-maximum likelihood estimation for cointegrated continuous-time linear state space models observed at low frequencies. *Electron. J. Stat.* **13** 5151–5212. MR4042406 https://doi.org/10.1214/19-EJS1636

Fasen-Hartmann, V. and Scholz, M. (2020). Cointegrated continuous-time linear state-space and MCARMA models. *Stochastics* **92** 1064–1099. MR4156002 https://doi.org/10.1080/17442508.2019.1691206

Kailath, T. (1980). *Linear Systems*. *Prentice-Hall Information and System Sciences Series*. Englewood Cliffs, NJ: Prentice-Hall, Inc. MR0569473

Marquardt, T. and Stelzer, R. (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* **117** 96–120. MR2287105 https://doi.org/10.1016/j.spa.2006.05.014

Schlemm, E. and Stelzer, R. (2012a). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.* **6** 2185–2234. MR3020261 https://doi.org/10.1214/12-EJS743

Schlemm, E. and Stelzer, R. (2012b). Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes. *Bernoulli* **18** 46–63. MR2888698 https://doi.org/10.3150/10-BEJ329

Tómasson, H. (2018). Implementation of multivariate continuous-time ARMA models. In *Continous-Time Modeling in the Behavioral and Related Sciences* 359–387. Cham: Springer.

Bernoulli **32**(1), 2026, 698–722
https://doi.org/10.3150/25-BEJ1876

Asymptotic breakdown point analysis for a general class of minimum divergence estimators

SUBHRAJYOTY ROY[a], ABIR SARKAR[b], ABHIK GHOSH[c] and AYANENDRANATH BASU[d]

Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India, [a]*roy.s@wustl.edu,* [b]*as4458@cornell.edu,* [c]*abhik.ghosh@isical.ac.in,* [d]*ayanbasu@isical.ac.in*

Robust inference based on the minimization of statistical divergences has proved to be a useful alternative to classical techniques based on maximum likelihood and related methods. Basu et al. (*Biometrika* **85** (1998) 549–559) introduced the density power divergence (DPD) family as a measure of discrepancy between two probability density functions and used this family for robust estimation of the parameter for independent and identically distributed data. Subsequently, Ghosh et al. (*Bernoulli* **23** (2017) 2746–2783) proposed a more general class of divergence measures, namely the S-divergence family, and discussed its usefulness in robust parametric estimation through several asymptotic properties and some numerical illustrations. In this paper, we develop the results concerning the asymptotic breakdown point for the minimum S-divergence estimators (in particular the minimum DPD estimator) under general model setups. The primary results of this paper provide lower bounds to the asymptotic breakdown point of these estimators which are free of the dimension of the data, in turn corroborating their usefulness in robust inference for high dimensional problems.

Keywords: Breakdown point; density power divergence; minimum S-divergence estimator; power divergence

References

Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach. CRC Monographs on Statistics & Applied Probability.* Chapman & Hall. MR2830561 https://doi.org/10.1201/b10956

Basu, A., Harris, I.R., Hjort, N.L. and Jones, M.C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559. MR1665873 https://doi.org/10.1093/biomet/85.3.549

Bhandari, S.K., Basu, A. and Sarkar, S. (2006). Robust inference in parametric models using the family of generalized negative exponential dispatches. *Aust. N. Z. J. Stat.* **48** 95–114. MR2234782 https://doi.org/10.1111/j.1467-842X.2006.00428.x

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

Cressie, N. and Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B, Methodol.* **46** 440–464. MR0790631

Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **8** 85–108. MR0164374

Dasiou, D. and Moyssiadis, C. (2001). The 50% breakdown point in simultaneous M-estimation of location and scale. *Statist. Papers* **42** 243–252. MR1854158 https://doi.org/10.1007/s003620100053

Davies, P.L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292. MR0902258 https://doi.org/10.1214/aos/1176350505

Donoho, D. and Huber, P.J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann. Wadsworth Statist./Probab. Ser.* 157–184. Belmont, CA: Wadsworth. MR0689745

Fishbone, J. and Mili, L. (2024). New highly efficient high-breakdown estimator of multivariate scatter and location for elliptical distributions. *Canad. J. Statist.* **52** 437–460. MR4745903 https://doi.org/10.1002/cjs.11770

Ghosh, A. (2015). Asymptotic properties of minimum S-divergence estimator for discrete models. *Sankhya A* **77** 380–407. MR3400120 https://doi.org/10.1007/s13171-014-0063-2

1350-7265

Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **7** 2420–2456. MR3117102 https://doi.org/10.1214/13-EJS847

Ghosh, A., Harris, I.R., Maji, A., Basu, A. and Pardo, L. (2017). A generalized divergence for statistical inference. *Bernoulli* **23** 2746–2783. MR3648044 https://doi.org/10.3150/16-BEJ826

Hampel, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896. MR0301858 https://doi.org/10.1214/aoms/1177693054

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (2011). *Robust Statistics: The Approach Based on Influence Functions*. *Wiley Series in Probability and Statistics*. Wiley. MR2488795 https://doi.org/10.1002/9781118186435

Hodges, J.L. Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics* 163–186. Berkeley, CA: Univ. California Press. MR0214251

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415 https://doi.org/10.1214/aoms/1177703732

Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc. MR2488795 https://doi.org/10.1002/9780470434697

Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67. MR0388656

Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M. (2019). *Robust Statistics*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc. Theory and methods (with R), Second edition of [MR2238141]. MR3839299

Morimoto, T. (1963). Markov processes and the H-theorem. *J. Phys. Soc. Jpn.* **18** 328–331. MR0167200 https://doi.org/10.1143/JPSJ.18.328

Park, C. and Basu, A. (2004). Minimum disparity estimation: Asymptotic normality and breakdown point results. *Bull. Inform. Cybernet.* **36** 19–33. MR2139489

Park, C., Kim, H. and Wang, M. (2022). Investigation of finite-sample properties of robust location and scale estimators. *Comm. Statist. Simulation Comput.* **51** 2619–2645. MR4422883 https://doi.org/10.1080/03610918.2019.1699114

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B (Bad Tatzmannsdorf, 1983)* 283–297. Dordrecht: Reidel. MR0851060

Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis (Heidelberg, 1983)*. *Lect. Notes Stat.* **26** 256–272. New York: Springer. MR0786313 https://doi.org/10.1007/978-1-4615-7821-5_15

Roy, S., Sarkar, A., Ghosh, A. and Basu, A. (2026). Supplement to "Asymptotic breakdown point analysis for a general class of minimum divergence estimators." https://doi.org/10.3150/25-BEJ1876SUPP

Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill Book Co. MR0924157

Simpson, D.G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *J. Amer. Statist. Assoc.* **82** 802–807. MR0909985

Smucler, E. (2019). Asymptotics for redescending M-estimators in linear models with increasing dimension. *Statist. Sinica* **29** 1065–1081. MR3931401

Tamura, R.N. and Boos, D.D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *J. Amer. Statist. Assoc.* **81** 223–229. MR0830585

Toma, A. (2007). Minimum Hellinger distance estimators for some multivariate models: Influence functions and breakdown point results. *C. R. Math. Acad. Sci. Paris* **345** 353–358. MR2359096 https://doi.org/10.1016/j.crma.2007.07.024

Toma, A. (2008). Minimum Hellinger distance estimators for multivariate distributions from the Johnson system. *J. Statist. Plann. Inference* **138** 803–816. MR2382888 https://doi.org/10.1016/j.jspi.2007.05.033

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

Bernoulli **32**(1), 2026, 723–751
https://doi.org/10.3150/25-BEJ1877

Benign overfitting in time-series linear models with over-parameterization

SHOGO NAKAKITA[1,a] and MASAAKI IMAIZUMI[2,b]

[1]*Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan,* [a]*nakakita@g.ecc.u-tokyo.ac.jp*
[2]*Graduate School of Arts and Sciences, The University of Tokyo / RIKEN Center for Advanced Intelligence Project, Tokyo, Japan,* [b]*imaizumi@g.ecc.u-tokyo.ac.jp*

The success of large-scale models in recent years has increased the importance of statistical models with numerous parameters. Several studies have analyzed over-parameterized linear models with high-dimensional data, which may not be sparse; however, existing results rely on the assumption of sample independence. In this study, we analyze a linear regression model with dependent time-series data in an over-parameterized setting. We consider an estimator using interpolation and develop a theory for the excess risk of the estimator. Then, we derive non-asymptotic risk bounds for the estimator for cases with dependent data. This analysis reveals that the coherence of the temporal covariance plays a key role; the risk bound is influenced by the product of temporal covariance matrices at different time steps. Moreover, we show the convergence rate of the risk bound and demonstrate that it is also influenced by the coherence of the temporal covariance. Finally, we provide several examples of specific dependent processes applicable to our setting.

Keywords: Benign overfitting; dependent data; linear model; over-parameterization; time series

References

Agarwal, A. and Duchi, J.C. (2012). The generalization ability of online algorithms for dependent data. *IEEE Trans. Inf. Theory* **59** 573–587. MR3008166 https://doi.org/10.1109/TIT.2012.2212414

Alquier, P. and Doukhan, P. (2011). Sparsity considerations for dependent variables. *Electron. J. Stat.* **5** 750–774. MR2824815 https://doi.org/10.1214/11-EJS626

Ba, J., Erdogdu, M., Suzuki, T., Wu, D. and Zhang, T. (2019). Generalization of two-layer neural networks: An asymptotic viewpoint. In *International Conference on Learning Representations*.

Bartlett, P.L., Long, P.M., Lugosi, G. and Tsigler, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

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

Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P. (2009). Rate of convergence of predictive distributions for dependent data. *Bernoulli* **15** 1351–1367. MR2597596 https://doi.org/10.3150/09-BEJ191

Bi, D., Han, X., Nie, A. and Yang, Y. (2022). Spiked eigenvalues of high-dimensional sample autocovariance matrices: CLT and applications. arXiv preprint. Available at arXiv:2201.03181.

Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

1350-7265

Bunea, F., Strimas-Mackey, S. and Wegkamp, M. (2022). Interpolating predictors in high-dimensional factor regression. *J. Mach. Learn. Res.* **23** 1–60. MR4420735

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523

Chen, W., Genton, M.G. and Sun, Y. (2021). Space-time covariance structures and models. *Annu. Rev. Stat. Appl.* **8** 191–215. MR4243546 https://doi.org/10.1146/annurev-statistics-042720-115603

Couillet, R., Greco, M.S., Ovarlez, J.-P. and Pascal, F. (2015). RMT for whitening space correlation and applications to radar detection. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)* 149–152. IEEE. https://doi.org/10.1109/CAMSAP.2015.7383758

Cressie, N.A.C. (1993). Statistics for Spatial Data. Wiley. MR1239641 https://doi.org/10.1002/9781119115151

Dagan, Y., Daskalakis, C., Dikkala, N. and Jayanti, S. (2019). Learning from weakly dependent data under Dobrushin's condition. In *Conference on Learning Theory* 914–928. PMLR.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *J. Financ. Econ.* **106** 157–181. https://doi.org/10.1016/j.jfineco.2012.04.003

Daskalakis, C., Dikkala, N. and Panageas, I. (2019). Regression from dependent observations. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 881–889. MR4003392 https://doi.org/10.1145/3313276.3316362

Davis, R.A., Zang, P. and Zheng, T. (2016). Sparse vector autoregressive modeling. *J. Comput. Graph. Statist.* **25** 1077–1096. MR3572029 https://doi.org/10.1080/10618600.2015.1092978

Dobriban, E. and Wager, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279. MR3766952 https://doi.org/10.1214/17-AOS1549

Guo, S., Wang, Y. and Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103** 889–903. MR3620446 https://doi.org/10.1093/biomet/asw046

Han, F., Lu, H. and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* **16** 3115–3150. MR3450535

Hastie, T., Tibshirani, R. and Wainwright, M. (2019). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC. MR3616141 https://doi.org/10.1201/b18401

Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R.J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

Kandiros, V., Dagan, Y., Dikkala, N., Goel, S. and Daskalakis, C. (2021). Statistical estimation from dependent data. In *International Conference on Machine Learning* 5269–5278. PMLR.

Karmakar, S., Richter, S. and Wu, W.B. (2022). Simultaneous inference for time-varying models. *J. Econometrics* **227** 408–428. MR4384679 https://doi.org/10.1016/j.jeconom.2021.03.002

Kock, A.B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *J. Econometrics* **186** 325–344. MR3343790 https://doi.org/10.1016/j.jeconom.2015.02.013

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

Kuznetsov, V. and Mohri, M. (2015). Learning theory and algorithms for forecasting non-stationary time series. *Adv. Neural Inf. Process. Syst.* **2015** 541–549. MR3563995

Kyriakidis, P.C. and Journel, A.G. (1999). Geostatistical space–time models: A. *Rev. Math. Geol.* **31** 651–684. MR1694654 https://doi.org/10.1023/A:1007528426688

Liang, T. and Rakhlin, A. (2020). Just interpolate: Kernel "ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849

Lototsky, S.V. (2009). Statistical inference for stochastic parabolic equations: A spectral approach. *Publ. Mat.* **53** 3–45. MR2474113 https://doi.org/10.5565/PUBLMAT\protect \relax \T1\textunderscore 53109\protect \relax \T1\textunderscore 01

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

Masini, R.P., Medeiros, M.C. and Mendes, E.F. (2022). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *J. Time Series Anal.* **43** 532–557. MR4446423 https://doi.org/10.1111/jtsa.12627

Medeiros, M.C. and Mendes, E.F. (2016). $\ell 1$-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *J. Econometrics* **191** 255–271. MR3434446 https://doi.org/10.1016/j.jeconom.2015.10.011

Mohri, M. and Rostamizadeh, A. (2008). Rademacher complexity bounds for non-I.I.D. processes. In *Proceedings of the 21st International Conference on Neural Information Processing Systems* 1097–1104.

Mohri, M. and Rostamizadeh, A. (2010). Stability bounds for stationary φ-mixing and β-mixing processes. *J. Mach. Learn. Res.* **11** 789–814. MR2600630

Nakakita, S. and Imaizumi, M. (2026). Supplement to "Benign overfitting in time-series linear models with over-parameterization." https://doi.org/10.3150/25-BEJ1877SUPP

Rudelson, M. and Vershynin, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** 1–9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

Song, S. and Bickel, P.J. (2011). Large vector auto regressions. arXiv preprint. Available at arXiv:1106.3915.

Tian, P. (2022). Joint CLT for top eigenvalues of sample covariance matrices of separable high dimensional long memory processes. *Random Matrices Theory Appl.* **11** 2250032. MR4440258 https://doi.org/10.1142/S2010326322500320

Tian, P. and Yao, J. (2022). Ratio-consistent estimation for long range dependent Toeplitz covariance with application to matrix data whitening. *Electron. J. Stat.* **16** 5035–5079. MR4491713 https://doi.org/10.1214/22-ejs2060

Tsigler, A. and Bartlett, P.L. (2023). Benign overfitting in ridge regression. *J. Mach. Learn. Res.* **24** 1–76. MR4583284

Van de Geer, S.A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809 https://doi.org/10.1214/009053607000000929

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge: Cambridge University Press. MR3837109 https://doi.org/10.1017/9781108231596

Wang, H., Li, G. and Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 63–78. MR2301500 https://doi.org/10.1111/j.1467-9868.2007.00577.x

Wong, K.C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690 https://doi.org/10.1214/19-AOS1840

Wu, W.-B. and Wu, Y.N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186 https://doi.org/10.1214/16-EJS1108

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94–116. MR1258867 https://doi.org/10.1214/aop/1176988849

Zhang, T. and Wu, W.B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029 https://doi.org/10.1214/12-AOS1010

Bernoulli **32**(1), 2026, 752–776
https://doi.org/10.3150/25-BEJ1878

Total variation bound for Hadwiger's functional using Stein's method

VALENTIN GARINO[1,a] and IVAN NOURDIN[2,b]

[1] *Uppsala University, Ångströmlaboratoriet, Lägerhyddsvägen 1, Box 480 751 06, Uppsala, Sweden,* [a] *valentin.garino@math.uu.se*
[2] *University of Luxembourg, Maison du Nombre, 6 avenue de la Fonte, L-4364 Esch-sur-Alzette, Grand Duchy of Luxembourg,* [b] *ivan.nourdin@uni.lu*

Let K be a convex body in $\mathbb{R}^d$, and let X_K be a d-dimensional random vector distributed according to the Hadwiger-Wills density μ_K, defined by $\mu_K(x) = ce^{-\pi \mathrm{dist}^2(x,K)}$, $x \in \mathbb{R}^d$. Finally, we define the information content H_K as $H_K = \mathrm{dist}^2(X_K, K)$. The goal of this paper is to study the fluctuations of H_K around its expectation as the dimension d tends to infinity. Relying on Stein's method and the Brascamp-Lieb inequality, we compute an explicit bound for the total variation distance between H_K and its Gaussian counterpart.

Keywords: Central limit theorem; convex body; intrinsic volumes; Stein's method; Steiner formula; Wills functional

References

[1] Bally, V. and Caramellino, L. (2016). Asymptotic development for the CLT in total variation distance. *Bernoulli* **22** 2442–2485. MR3498034 https://doi.org/10.3150/15-BEJ734
[2] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480 https://doi.org/10.1016/0022-1236(76)90004-5
[3] Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6
[4] Chevet, S. (1976). Processus gaussiens et volumes mixtes. *Z. Wahrsch. Verw. Gebiete* **36** 47–65. MR0426120 https://doi.org/10.1007/BF00533208
[5] Goldstein, L., Nourdin, I. and Peccati, G. (2017). Gaussian phase transitions and conic intrinsic volumes: Steining the Steiner formula. *Ann. Appl. Probab.* **27** 1–47. MR3619780 https://doi.org/10.1214/16-AAP1195
[6] Gruber, P.M. and Kenderov, P. (1982). Approximation of convex bodies by polytopes. *Rend. Circ. Mat. Palermo (2)* **31** 195–225. MR0670396 https://doi.org/10.1007/BF02844354
[7] Hadwiger, H. (1975). Das Wills'sche Funktional. *Monatsh. Math.* **79** 213–221. MR0379384 https://doi.org/10.1007/BF01304074
[8] Johnson, O. (2017). A discrete log-Sobolev inequality under a Bakry–Émery type condition. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1952–1970. MR3729642 https://doi.org/10.1214/16-AIHP778
[9] Ley, C., Reinert, G. and Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. MR3595350 https://doi.org/10.1214/16-PS278
[10] Lotz, M. and Tropp, J.A. (2022). Sharp phase transitions in Euclidean integral geometry. arXiv:2208.13919.
[11] Lotz, M., McCoy, M.B., Nourdin, I., Peccati, G. and Tropp, J.A. (2020). Concentration of the intrinsic volumes of a convex body. In *Geometric Aspects of Functional Analysis* (B. Klartag and E. Milman, eds.). *Lecture Notes in Mathematics* **2266**. Cham: Springer. MR4175761 https://doi.org/10.1007/978-3-030-46762-3_6
[12] Nourdin, I. and Peccati, G. (2012). *Normal Approximation with Malliavin Calculus: From Stein's Method to Universality* **192**. Cambridge University Press. MR2962301 https://doi.org/10.1017/CBO9781139084659

1350-7265

[13] Schneider, R. (2007). Integral geometric tools for stochastic geometry. In *Stochastic Geometry* (W. Weil, ed.). *Lecture Notes in Mathematics* **1892**. Berlin, Heidelberg: Springer. MR2327292 https://doi.org/10.1007/978-3-540-38175-4_3

[14] Wills, J.M. (1973). Zur Gitterpunktanzahl konvexer Mengen. *Elem. Math.* **28** 57–63. MR0400063

Bernoulli **32**(1), 2026, 777–797
https://doi.org/10.3150/25-BEJ1879

Sharp phase transitions in high-dimensional changepoint detection

DANIEL XIANG[a] and CHAO GAO[b]

Department of Statistics, University of Chicago, Chicago, U.S.A., [a]*dxiang@uchicago.edu,* [b]*chaogao@uchicago.edu*

We study a hypothesis testing problem in the context of high-dimensional changepoint detection. Given a matrix $X \in \mathbb{R}^{p\times n}$ with independent Gaussian entries, the goal is to determine whether or not a sparse, non-null fraction of rows in X exhibits a shift in mean at a common index between 1 and n. We focus on three aspects of this problem: the sparsity of non-null rows, the presence of a single, common changepoint in the non-null rows, and the signal strength associated with the changepoint. Within an asymptotic regime relating the data dimensions n and p to the signal sparsity and strength, the information-theoretic limits of this testing problem are characterized by a formula that determines whether or not there exists a testing procedure whose sum of Type I and II errors tends to zero as $n, p \to \infty$. The formula, called the *detection boundary*, partitions the parameter space into a two regions: one where it is possible to detect the presence of a single aligned changepoint (detectable region), and another where no test is able to consistently distinguish the mean matrix from one with constant rows (undetectable region).

Keywords: High-dimensional changepoint detection; minimax testing; phase transition; sparse mixture

References

Arias-Castro, E., Candes, E.J. and Durand, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847 https://doi.org/10.1214/10-AOS839

Aston, J.A.D. and Kirch, C. (2012a). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.* **109** 204–220. MR2922864 https://doi.org/10.1016/j.jmva.2012.03.006

Aston, J.A.D. and Kirch, C. (2012b). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688 https://doi.org/10.1214/12-AOAS565

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

Berk, R.H. and Jones, D.H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531 https://doi.org/10.1007/BF00533250

Bleakley, K. and Vert, J.-P. (2011). The group fused lasso for multiple change-point detection. arXiv preprint. arXiv:1106.4199.

Butucea, C. and Ingster, Y.I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

Cai, T.T., Jeng, J.X. and Jin, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452 https://doi.org/10.1111/j.1467-9868.2011.00778.x

Cai, T.T. and Wu, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inf. Theory* **60** 2217–2232. MR3181520 https://doi.org/10.1109/TIT.2014.2304295

Chan, H.P. and Walther, G. (2015). Optimal detection of multi-sample aligned sparse signals. *Ann. Statist.* **43** 1865–1895. MR3375870 https://doi.org/10.1214/15-AOS1328

Chen, Y., Wang, T. and Samworth, R.J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

Chen, Y., Wang, T. and Samworth, R.J. (2024). Inference in high-dimensional online changepoint detection. *J. Amer. Statist. Assoc.* **119** 1461–1472. MR4766002 https://doi.org/10.1080/01621459.2023.2199962

1350-7265

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

Cole, D.M., Smith, S.M. and Beckmann, C.F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state fMRI data. *Front. Syst. Neurosci.* **4** 1459.

Collier, O., Comminges, L. and Tsybakov, A.B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

Efron, B. and Zhang, N.R. (2011). False discovery rates and copy number variation. *Biometrika* **98** 251–271. MR2806427 https://doi.org/10.1093/biomet/asr018

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

Follain, B., Wang, T. and Samworth, R.J. (2022). High-dimensional changepoint estimation with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1023–1055. MR4460584

Gao, C., Han, F. and Zhang, C.-H. (2020). On estimation of isotonic piecewise constant signals. *Ann. Statist.* **48** 629–654. MR4102670 https://doi.org/10.1214/18-AOS1792

Horváth, L., Kokoszka, P. and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.* **68** 96–119. MR1668911 https://doi.org/10.1006/jmva.1998.1780

Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *TEST* **23** 219–255. MR3210268 https://doi.org/10.1007/s11749-014-0368-4

Hu, S., Huang, J., Chen, H. and Chan, H.P. (2023). Likelihood scores for sparse signal and change-point detection. *IEEE Trans. Inf. Theory* **69** 4065–4080. MR4604484 https://doi.org/10.1109/tit.2023.3242297

Ingster, Y.I. (1999). Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** 401–428. MR1680087

Jeng, X.J., Cai, T.T. and Li, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. MR3034330 https://doi.org/10.1093/biomet/ass059

Killick, R., Fearnhead, P. and Eckley, I.A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

Kovács, S., Li, H., Haubner, L., Munk, A. and Bühlmann, P. (2020). Optimistic search strategy: Change point detection for large-scale data via adaptive logarithmic queries. arXiv preprint. arXiv:2010.10194. MR4829141

Kovács, S., Bühlmann, P., Li, H. and Munk, A. (2023). Seeded binary segmentation: A general methodology for fast and optimal changepoint detection. *Biometrika* **110** 249–256. MR4565454 https://doi.org/10.1093/biomet/asac052

Li, M., Chen, Y., Wang, T. and Yu, Y. (2023). Robust mean change point testing in high-dimensional data with heavy tails. arXiv preprint. arXiv:2305.18987. MR4932558 https://doi.org/10.1016/j.econlet.2025.112477

Liu, H., Gao, C. and Samworth, R.J. (2021). Minimax rates in sparse, high-dimensional change point detection. *Ann. Statist.* **49** 1081–1112. MR4255120 https://doi.org/10.1214/20-aos1994

Page, E.S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

Pilliat, E., Carpentier, A. and Verzelen, N. (2023). Optimal multiple change-point detection for high-dimensional data. *Electron. J. Stat.* **17** 1240–1315. MR4576243 https://doi.org/10.1214/23-ejs2126

Shah, S.P., Lam, W.L., Ng, R.T. and Murphy, K.P. (2007). Modeling recurrent DNA copy number alterations in array CGH data. *Bioinformatics* **23** i450–i458.

Tukey, J. (1989). Higher criticism for individual significances in several tables or parts of tables. Princeton: Princeton University (Internal working paper).

Verzelen, N., Fromont, M., Lerasle, M. and Reynaud-Bouret, P. (2023). Optimal change-point detection and localization. *Ann. Statist.* **51** 1586–1610. MR4658569 https://doi.org/10.1214/23-aos2297

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

Xiang, D. and Gao, C. (2026). Supplement to "Sharp phase transitions in high-dimensional changepoint detection." https://doi.org/10.3150/25-BEJ1879SUPP

Xie, Y. and Siegmund, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 1–20. MR3099117 https://doi.org/10.1214/13-AOS1094

Yu, Y. (2020). A review on minimax rates in change point detection and localisation. arXiv preprint. arXiv:2011.01857.

Bernoulli **32**(1), 2026, 798–822
https://doi.org/10.3150/25-BEJ1880

Quasi-limiting behaviour of the sub-critical multitype bisexual Galton-Watson process

CORALIE FRITSCH[1,a], DENIS VILLEMONAIS[2,3,b] and NICOLÁS ZALDUENDO[4,c]

[1] *Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France,* [a]*coralie.fritsch@inria.fr*
[2] *Université de Strasbourg, IRMA, F-67000 Strasbourg, France,* [b]*denis.villemonais@unistra.fr*
[3] *Institut universitaire de France (IUF)*
[4] *MISTEA, Université de Montpellier, INRAE, Institut Agro, Montpellier, France,*
[c]*nicolas.zalduendo-vidal@inrae.fr*

We investigate the quasi-limiting behavior of bisexual subcritical Galton-Watson processes. While classical subcritical Galton-Watson processes have been extensively analyzed, bisexual Galton-Watson processes present unique difficulties because of the lack of the branching property. To prove the existence of and convergence to one or several quasi-stationary distributions, we leverage on recent developments linking bisexual Galton-Watson processes extinction to the eigenvalue of a concave operator.

Keywords: Branching processes with interactions; exponential convergence; Lyapunov function; multisexual processes; quasi-stationary distributions; two-sex processes

References

Alsmeyer, G. (2002). Bisexual Galton-Watson Processes: A Survey. https://www.uni-muenster.de/Stochastik/alsmeyer/bisex(survey).pdf.

Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*. *Die Grundlehren der mathematischen Wissenschaften* **196**. New York: Springer-Verlag. MR0373040

Ben-Ari, I. and Jiang, N. (2024). Representation and Characterization of Quasistationary Distributions for Markov Chains.

Champagnat, N. and Villemonais, D. (2022). Quasi-stationary distributions in reducible state spaces. arXiv preprint arXiv:2201.10151.

Champagnat, N. and Villemonais, D. (2023). General criteria for the study of quasi-stationarity. *Electron. J. Probab.* **28** 1–84. MR4546021 https://doi.org/10.1214/22-ejp880

Collet, P., Martínez, S. and San Martín, J. (2013). *Quasi-Stationary Distributions*. *Probability and its Applications (New York)*. Heidelberg: Springer. Markov chains, diffusions and dynamical systems. MR2986807 https://doi.org/10.1007/978-3-642-33131-2

Daley, D.J. (1968). Extinction conditions for certain bisexual Galton-Watson branching processes. *Z. Wahrsch. Verw. Gebiete* **9** 315–322. MR0230383 https://doi.org/10.1007/BF00531755

Daley, D.J., Hull, D.M. and Taylor, J.M. (1986). Bisexual Galton-Watson branching processes with superadditive mating functions. *J. Appl. Probab.* 585–600. MR0855367 https://doi.org/10.1017/s0021900200111763

Ferrari, P.A., Kesten, H., Martinez, S. and Picco, P. (1995). Existence of Quasi-Stationary Distributions. A Renewal Dynamical Approach. *Ann. Probab.* **23** 501–521. MR1334159

Fritsch, C., Villemonais, D. and Zalduendo, N. (2024). The multi-type bisexual Galton–Watson branching process. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 2975–3008. MR4828863 https://doi.org/10.1214/23-aihp1417

González, M. and Molina, M. (1996). On the limit behaviour of a superadditive bisexual Galton–Watson branching process. *J. Appl. Probab.* **33** 960–967. MR1416218 https://doi.org/10.1017/s0021900200100397

1350-7265

González, M., Hull, D.M., Martínez, R. and Mota, M. (2006). Bisexual branching processes in a genetic context: The extinction problem for Y-linked genes. *Math. Biosci.* **202** 227–247. MR2255142 https://doi.org/10.1016/j.mbs.2006.03.010

Hull, D.M. (2003). A survey of the literature associated with the bisexual Galton-Watson branching process. *Extracta Math.* **18** 321–343. MR2056600

Krause, U. (1994). Relative stability for ascending and positively homogeneous operators on Banach spaces. *J. Math. Anal. Appl.* **188** 182–202. MR1301725 https://doi.org/10.1006/jmaa.1994.1420

Maillard, P. (2018). The λ-invariant measures of subcritical Bienaymé-Galton-Watson processes. *Bernoulli* **24** 297–315. MR3706758 https://doi.org/10.3150/16-BEJ877

Méléard, S. and Villemonais, D. (2012). Quasi-stationary distributions and population processes. *Probab. Surv.* **9** 340–410. MR2994898 https://doi.org/10.1214/11-PS191

Molina, M. (2010). Two-sex branching process literature. In *Workshop on Branching Processes and Their Applications* 279–293. Springer. MR2730954 https://doi.org/10.1007/978-3-642-11156-3_20

Seneta, E. and Vere-Jones, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probab.* **3** 403–434. MR0207047 https://doi.org/10.2307/3212128

van Doorn, E.A. and Pollett, P.K. (2013). Quasi-stationary distributions for discrete-state models. *European J. Oper. Res.* **230** 1–14. MR3063313 https://doi.org/10.1016/j.ejor.2013.01.032

Yaglom, A.M. (1947). Certain limit theorems of the theory of branching random processes. *Dokl. Akad. Nauk* **56** 795–798. MR0022045

Zalduendo, N. (2023). Processus de branchement bi-sexués multi-types PhD thesis Thèse de doctorat dirigée par Villemonais, Denis et Fritsch, Coralie Mathématiques Université de Lorraine. 2023.

Bernoulli **32**(1), 2026, 823–847
https://doi.org/10.3150/25-BEJ1881

Equilibrium moderate deviations for occupation times of SSEP on regular trees

XIAOFENG XUE[a]

School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, China, [a]*xfxue@bjtu.edu.cn*

In this paper, we are concerned with the symmetric simple exclusion process on the regular tree $\mathcal{T}_d$ for $d \geq 2$. Our main result gives moderate deviation principles (MDP) of occupation times of the process starting from an invariant product measure. Two replacement lemmas play key roles in the proof of our main result. To obtain these replacement lemmas, we utilize duality relationships between the symmetric exclusion process and two types of random walks on $\mathcal{T}_d$ and $\mathcal{T}_d \times \mathcal{T}_d$ respectively.

Keywords: Dirichlet form; exclusion process; moderate deviation; occupation time; regular tree

References

Birkner, M. and Zähle, I. (2007). A functional CLT for the occupation time of a state-dependent branching random walk. *Ann. Probab.* **35** 2063–2090. MR2353383 https://doi.org/10.1214/009117907000000150

Bramson, M., Cox, J.T. and Griffeath, D. (1988). Occupation time large deviations of the voter model. *Probab. Theory Related Fields* **77** 401–413. MR0931506 https://doi.org/10.1007/BF00319297

Cox, J.T. and Griffeath, D. (1983). Occupation time limit theorems for the voter model. *Ann. Probab.* **11** 876–893. MR0714952

Deuschel, J.D. and Rosen, J. (1998). Occupation time large deviations for critical branching Brownian motion and related processes. *Ann. Probab.* **26** 602–643. MR1626503 https://doi.org/10.1214/aop/1022855645

Gao, F. and Quastel, J. (2003). Moderate deviations from the hydrodynamic limit of the symmetric exclusion process. *Sci. China Math.* **5** 577–592. MR2025926 https://doi.org/10.1360/02ys0114

Gao, F. and Quastel, J. (2024). Deviation inequalities and moderate deviations for the symmetric exclusion process. *Ann. Inst. Henri Poincaré Probab. Stat.* to appear.

Kipnis, C. (1987). Fluctuations des temps d'occupation d'un site dans l'exclusion simple symétrique. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 21–35. MR0877383

Kipnis, C. and Landim, C. (1999). *Scaling Limits of Interacting Particle Systems*. Berlin: Springer. MR1707314 https://doi.org/10.1007/978-3-662-03752-2

Kipnis, C., Olla, S. and Varadhan, S.R.S. (1989). Hydrodynamics and large deviation for simple exclusion processes. *Comm. Pure Appl. Math.* **42** 115–137. MR0978701 https://doi.org/10.1002/cpa.3160420202

Komorowski, T., Landim, C. and Olla, S. (2012). *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*. New York: Springer. MR2952852 https://doi.org/10.1007/978-3-642-29880-6

Landim, C. (1992). Occupation time large deviations for the symmetric simple exclusion process. *Ann. Probab.* **20** 206–231. MR1143419

Lee, T.Y., Landim, C. and Chang, C. (2004). Occupation time large deviations of two-dimensional symmetric simple exclusion process. *Ann. Probab.* **32** 661–691. MR2039939 https://doi.org/10.1214/aop/1079021460

Li, Q. and Ren, Y. (2011). A large deviation for occupation time of critical branching α-stable process. *Sci. China Math.* **54** 1445–1456. MR2817578 https://doi.org/10.1007/s11425-011-4209-6

Liggett, T.M. (1985). *Interacting Particle Systems*. New York: Springer. MR0776231 https://doi.org/10.1007/978-1-4613-8542-4

Liggett, T.M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. New York: Springer. MR1717346 https://doi.org/10.1007/978-3-662-03990-8

Maillard, G. and Mountford, T. (2009). Large deviations for voter model occupation times in two dimensions. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 577–588. MR2521414 https://doi.org/10.1214/08-AIHP178

Schonmann, H. (1986). Central limit theorem for the contact process. *Ann. Probab.* **14** 1291–1295. MR0866350

Sion, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

Xue, X. (2025). Limit theorems of occupation times of normalized binary contact path processes on lattices. *J. Appl. Probab.* **62** 44–66. MR4864600 https://doi.org/10.1017/jpr.2024.41

Xue, X. (2026). Supplement to "Equilibrium moderate deviations for occupation times of SSEP on regular trees." https://doi.org/10.3150/25-BEJ1881SUPP