STATISTICAL SCIENCE

Volume 39, Number 3 August 2024

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 39, Number 3, August 2024. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Statistical Science **Volume 39, Number 3 (375–517) August 2024**

Volume 39 Number 3 August 2024

Statistical Science
2024, Vol. 39, No. 3, 375–390
https://doi.org/10.1214/23-STS911

An Introduction to Proximal Causal Inference

Eric J. Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao

Abstract. A standard assumption for causal inference from observational data is that one has measured a sufficiently rich set of covariates to ensure that within covariate strata, subjects are exchangeable across observed treatment values. Skepticism about the exchangeability assumption in observational studies is often warranted because it hinges on investigators' ability to accurately measure covariates capturing all potential sources of confounding. Realistically, confounding mechanisms can rarely if ever, be learned with certainty from measured covariates. One can therefore only ever hope that covariate measurements are at best proxies of true underlying confounding mechanisms operating in an observational study, thus invalidating causal claims made on basis of standard exchangeability conditions. Causal inference from proxies is a challenging inverse problem which has to date remained unresolved. In this paper, we introduce a formal potential outcome framework for *proximal causal inference*, which while explicitly acknowledging covariate measurements as imperfect proxies of confounding mechanisms, offers an opportunity to learn about causal effects in settings where exchangeability on the basis of measured covariates fails. The proposed framework is closely related to the emerging literature on the use of proxies or negative control variables for nonparametric identification of causal effects in presence of hidden confounding bias (*Biometrika* **105** (2018) 987–993). However, while prior literature largely focused on point treatment settings, here we consider the more challenging setting of a complex longitudinal study with time-varying treatments and both measured and unmeasured time-varying confounding. Upon reviewing existing results for proximal identification in the point treatment setting, we provide new identification results for the time-varying setting, leading to the *proximal g-formula* and corresponding *proximal g-computation algorithm* for estimation. These may be viewed as generalizations of Robins' foundational g-formula and g-computation algorithm, which account explicitly for bias due to unmeasured confounding. Applications of proximal g-computation of causal effects are given for illustration in both point treatment and time-varying treatment settings.

Key words and phrases: Causality, counterfactual outcomes, proxies, confounding, negative control.

REFERENCES

[1] Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71** 1795–1843. MR2015420 https://doi.org/10.1111/1468-0262.00470

[2] An, Y. and Hu, Y. (2012). Well-posedness of measurement error models for self-reported data. *J. Econometrics* **168** 259–269. MR2923767 https://doi.org/10.1016/j.jeconom.2012.01.036

[3] Andrews, D. W. K. (2017). Examples of L^2-complete and boundedly-complete distributions. *J. Econometrics* **199** 213–220. MR3681027 https://doi.org/10.1016/j.jeconom.2017.05.011

[4] Blundell, R., Chen, X. and Kristensen, D. (2007). Semi-

Eric J. Tchetgen Tchetgen is Professor in the Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: ett@wharton.upenn.edu). Andrew Ying is postdoctoral fellow in the Department of Statistics and Data Science, the Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. Yifan Cui is Assistant Professor in the Center for Data Science and the School of Management, Zhejiang University, Hangzhou, China. Xu Shi is Assistant Professor in the Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. Wang Miao is Assistant Professor in the Department of Probability and Statistics, Peking University, Bejing, China.

nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* **75** 1613–1669. MR2351452 https://doi.org/10.1111/j.1468-0262.2007.00808.x

[5] CANAY, I. A., SANTOS, A. and SHAIKH, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica* **81** 2535–2559. MR3138554 https://doi.org/10.3982/ECTA10851

[6] CARROLL, R. J., CHEN, X. and HU, Y. (2010). Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. *J. Nonparametr. Stat.* **22** 379–399. MR2662599 https://doi.org/10.1080/10485250902874688

[7] CASELLA, G. and BERGER, R. L. (2001). *Statistical Inference*. *The Wadsworth & Brooks/Cole Statistics/Probability Series*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR1051420

[8] CHEN, X., CHERNOZHUKOV, V., LEE, S. and NEWEY, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica* **82** 785–809. MR3191719 https://doi.org/10.3982/ECTA9988

[9] CHEN, X. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ.* **9** 39–84. MR3789729 https://doi.org/10.3982/QE722

[10] CHEN, X. and HU, Y. (2006). Identification and inference of nonlinear models using two samples with arbitrary measurement errors. Technical report, Cowles, Foundation for Research in Economics, Yale Univ.

[11] CHERNOZHUKOV, V. and HANSEN, C. (2005). An IV model of quantile treatment effects. *Econometrica* **73** 245–261. MR2115636 https://doi.org/10.1111/j.1468-0262.2005.00570.x

[12] CHOI, H. K., HERNÁN, M. A., SEEGER, J. D., ROBINS, J. M. and WOLFE, F. (2002). Methotrexate and mortality in patients with rheumatoid arthritis: A prospective study. *Lancet* **359** 1173–1177.

[13] CONNORS, A. F., SPEROFF, T., DAWSON, N. V., THOMAS, C., HARRELL, F. E., WAGNER, D., DESBIENS, N., GOLDMAN, L., WU, A. W. et al. (1996). The effectiveness of right heart catheterization in the initial care of critically III patients. *JAMA* **276** 889–897.

[14] CUI, Y., PU, H., SHI, X. and MIAO, W. (2023). Semiparametric proximal causal inference. *J. Amer. Statist. Assoc.* 1–12.

[15] DAROLLES, S., FAN, Y., FLORENS, J. P. and RENAULT, E. (2011). Nonparametric instrumental regression. *Econometrica* **79** 1541–1565. MR2883763 https://doi.org/10.3982/ECTA6539

[16] EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

[17] FOLSTEIN, M. F., FOLSTEIN, S. E. and MCHUGH, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12** 189–198.

[18] FOX, J., KLEIBER, C. and ZEILEIS, A. (2021). ivreg: Instrumental-variables regression by '2SLS', '2SM', or '2SMM', with diagnostics. R package version 0.6-1.

[19] FREYBERGER, J. (2018). Non-parametric panel data models with interactive fixed effects. *Rev. Econ. Stud.* **85** 1824–1851. MR3818061 https://doi.org/10.1093/restud/rdx052

[20] GHASSAMI, A., YING, A., SHPITSER, I. and TCHETGEN TCHETGEN, E. J. (2021). Minimax kernel machine learning for a class of doubly robust functionals. arXiv preprint. Available at arXiv:2104.02929.

[21] GREENLAND, S. and ROBINS, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *Int. J. Epidemiol.* **15** 413–419.

[22] HALL, P., HOROWITZ, J. L. et al. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Ann. Statist.* **33** 2904–2929. MR2253107 https://doi.org/10.1214/009053605000000714

[23] HEATON, R. K., FRANKLIN, D. R., ELLIS, R. J., MCCUTCHAN, J. A., LETENDRE, S. L., LEBLANC, S., CORKRAN, S. H., DUARTE, N. A., CLIFFORD, D. B. et al. (2011). HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: Differences in rates, nature, and predictors. *J. Neurovirology* **17** 3–16.

[24] HERNÁN, M. A., BRUMBACK, B. A. and ROBINS, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat. Med.* **21** 1689–1709. https://doi.org/10.1002/sim.1144

[25] HERNÁN, M. A. and ROBINS, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology* **17** 360–372. https://doi.org/10.1097/01.ede.0000222409.00878.37

[26] HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference*: *What If*. CRC Press/CRC, Boca Raton, FL.

[27] HOROWITZ, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica* **79** 347–394. MR2809374 https://doi.org/10.3982/ECTA8662

[28] HU, Y. and SCHENNACH, S. M. (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

[29] HU, Y. and SHUM, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *J. Econometrics* **171** 32–44. MR2970334 https://doi.org/10.1016/j.jeconom.2012.05.023

[30] KALLUS, N., MAO, X. and UEHARA, M. (2021). Causal inference under unmeasured confounding with negative controls: A minimax learning approach. arXiv preprint. Available at arXiv:2103.14029.

[31] KRESS, R., MAZ'YA, V. and KOZLOV, V. (1989). *Linear Integral Equations*. *Applied Mathematical Sciences* **82**. Springer, Berlin. MR1007594 https://doi.org/10.1007/978-3-642-97146-4

[32] KUROKI, M. and PEARL, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101** 423–437. MR3215357 https://doi.org/10.1093/biomet/ast066

[33] LIPSITCH, M., TCHETGEN TCHETGEN, E. J. and COHEN, T. (2010). Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* **21** 383.

[34] MIAO, W., GENG, Z. and TCHETGEN TCHETGEN, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105** 987–993. MR3877879 https://doi.org/10.1093/biomet/asy038

[35] MIAO, W., HU, W., OGBURN, E. L. and ZHOU, X.-H. (2023). Identifying effects of multiple treatments in the presence of unmeasured confounding. *J. Amer. Statist. Assoc.* **118** 1953–1967. MR4646619 https://doi.org/10.1080/01621459.2021.2023551

[36] MIAO, W. and TCHETGEN TCHETGEN, E. J. (2017). Invited commentary: Bias attenuation and identification of causal effects with multiple negative controls. *Amer. J. Epidemiol.* **185** 950–953.

[37] MIAO, W., SHI, X. and TCHETGEN TCHETGEN, E. J. (2018). A confounding bridge approach for double negative control inference on causal effects. arXiv preprint. Available at arXiv:1808.04945.

[38] NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578. MR2000257 https://doi.org/10.1111/1468-0262.00459

[39] OGBURN, E. L. and VANDERWEELE, T. J. (2013). Bias attenuation results for nondifferentially mismeasured ordinal and

coarsened confounders. *Biometrika* **100** 241–248. MR3034338 https://doi.org/10.1093/biomet/ass054

[40] PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

[41] ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lect*. *Notes Stat*. **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

[42] ROBINS, J. M. (1999). Marginal structural models. In *Proceedings of the Section on Bayesian Statistical Science* 1–10. Amer. Statist. Assoc., Alexandria. MR1766776 https://doi.org/10.1023/A:1005285815569

[43] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[44] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J*. *Amer*. *Statist*. *Assoc*. **94** 1096–1120. MR1731478 https://doi.org/10.2307/2669923

[45] SHI, X., MIAO, W., NELSON, J. C. and TCHETGEN TCHETGEN, E. J. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **82** 521–540. MR4084174 https://doi.org/10.1111/rssb.12361

[46] SHI, X., MIAO, W. and TCHETGEN TCHETGEN, E. J. (2020). A selective review of negative control methods in epidemiology. *Curr*. *Epidemiol*. *Rep*. 1–13.

[47] SHIU, J.-L. and HU, Y. (2013). Identification and estimation of nonlinear dynamic panel data models with unobserved covariates. *J*. *Econometrics* **175** 116–131. MR3061934 https://doi.org/10.1016/j.jeconom.2013.03.001

[48] SOFER, T., RICHARDSON, D. B., COLICINO, E., SCHWARTZ, J. and TCHETGEN TCHETGEN, E. J. (2016). On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statist*. *Sci*. **31** 348–361. MR3552739 https://doi.org/10.1214/16-STS558

[49] STOKES, A. and MEHTA, N. K. (2013). Mortality and excess risk in US adults with pre-diabetes and diabetes: A comparison of two nationally representative cohorts, 1988–2006. *Popul*. *Health Metr*. **11** 1–7.

[50] TCHETGEN TCHETGEN, E. J., YING, A., CUI, Y., SHI, X. and MIAO, W. (2024). Supplement to "An introduction to proximal causal inference." https://doi.org/10.1214/23-STS911SUPP

[51] TOGLIA, J., FITZGERALD, K. A., O'DELL, M. W., MASTROGIOVANNI, A. R. and LIN, C. D. (2011). The mini-mental state examination and Montreal cognitive assessment in persons with mild subacute stroke: Relationship to functional outcome. *Arch*. *Phys*. *Med*. *Rehabil*. **92** 792–798.

[52] TOMBAUGH, T. N. and MCINTYRE, N. J. (1992). The mini-mental state examination: A comprehensive review. *J*. *Amer*. *Geriatr*. *Soc*. **40** 922–935.

[53] VANSTEELANDT, S. and DIDELEZ, V. (2018). Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. *Scand*. *J*. *Stat*. **45** 941–961. MR3884895 https://doi.org/10.1111/sjos.12329

[54] WANG, L. and TCHETGEN TCHETGEN, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **80** 531–550. MR3798877 https://doi.org/10.1111/rssb.12262

[55] WOODS, S. P., MOORE, D. J., WEBER, E. and GRANT, I. (2009). Cognitive neuropsychology of HIV-associated neurocognitive disorders. *Neuropsychol*. *Rev*. **19** 152–168.

[56] WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

[57] YING, A., MIAO, W., SHI, X. and TCHETGEN TCHETGEN, E. J. (2023). Proximal causal inference for complex longitudinal studies. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **85** 684–704.

Statistical Science
2024, Vol. 39, No. 3, 391–408
https://doi.org/10.1214/23-STS916

A General Construction of Multivariate Dependence Structures with Nonmonotone Mappings and Its Applications

Jean-François Quessy

Abstract. A famous theorem by Sklar (1959) provides an elegant and useful way to look at multivariate dependence structures. This paper explores the construction of copulas from nonmonotone transformations applied to the components of random vectors whose marginals are uniform on $(0, 1)$. This approach allows the creation of new families of multivariate copulas that generalize the chi-square, Fisher, squared and V-copulas, to name a few. The properties of the resulting dependence structures are studied, including tail dependence and tail asymmetry. The usefulness of the models created is illustrated for standard multivariate dependence modeling, nonmonotone copula regression and spatial dependence.

Key words and phrases: Semiparametric inference, tail asymmetry, tail dependence, V- and squared copulas, uniform-to-uniform transformations.

REFERENCES

[1] AAS, K., CZADO, C., FRIGESSI, A. and BAKKEN, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44** 182–198. MR2517884 https://doi.org/10.1016/j.insmatheco.2007.02.001

[2] ALIDOOST, F., STEIN, A. and SU, Z. (2018). Copula-based interpolation methods for air temperature data using collocated covariates. *Spat. Stat.* **28** 128–140. MR3887160 https://doi.org/10.1016/j.spasta.2018.08.003

[3] AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391

[4] BALAKRISHNAN, N. and LAI, C.-D. (2009). *Continuous Bivariate Distributions*, 2nd ed. Springer, Dordrecht. MR2840643 https://doi.org/10.1007/b101765

[5] BÁRDOSSY, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resour. Res.* **42** 1–12.

[6] BÁRDOSSY, A. and PEGRAM, G. (2009). Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.* **13** 2299–2314.

[7] BRANCO, M. D. and DEY, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79** 99–113. MR1867257 https://doi.org/10.1006/jmva.2000.1960

[8] BRUNNER, M. I., FURRER, R. and FAVRE, A.-C. (2019). Modeling the spatial dependence of floods using the Fisher copula. *Hydrol. Earth Syst. Sci.* **23** 107–124.

[9] CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8

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

[11] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication. MR1239641 https://doi.org/10.1002/9781119115151

[12] CZADO, C. (2019). *Analyzing Dependent Data with Vine Copulas*: *A Practical Guide with R. Lecture Notes in Statistics* **222**. Springer, Cham. MR3931334 https://doi.org/10.1007/978-3-030-13785-4

[13] DETTE, H., VAN HECKE, R. and VOLGUSHEV, S. (2014). Some comments on copula-based regression. *J. Amer. Statist. Assoc.* **109** 1319–1324. MR3265699 https://doi.org/10.1080/01621459.2014.916577

[14] ELTOFT, T., KIM, T. and LEE, T. W. (2006). On the multivariate Laplace distribution. *IEEE Signal Process. Lett.* **13** 300–303.

[15] FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions*. *Monographs on Statistics and Applied Probability* **36**. CRC Press, London. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

[16] FAVRE, A.-C., QUESSY, J.-F. and TOUPIN, M.-H. (2018). The new family of Fischer copulas to model upper tail dependence and radial asymmetry: Properties and application to high-dimensional rainfall data. *Environmetrics* **29** e2494, 17. MR3799916 https://doi.org/10.1002/env.2494

[17] GENEST, C., FAVRE, A., BÉLIVEAU, J. and JACQUES, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* **43** W09401.

Jean-François Quessy is Professor, Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Canada (e-mail: jean-francois.quessy@uqtr.ca; URL: www.uqtr.ca/jean-francois.quessy).

[18] GENEST, C., GHOUDI, K. and RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. MR1366280 https://doi.org/10.1093/biomet/82.3.543

[19] GENEST, C. and MACKAY, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *Amer. Statist.* **40** 280–283. MR0866908 https://doi.org/10.2307/2684602

[20] GENEST, C., NEŠLEHOVÁ, J. G. and RIVEST, L.-P. (2018). The class of multivariate max-id copulas with ℓ_1-norm symmetric exponent measure. *Bernoulli* **24** 3751–3790. MR3788188 https://doi.org/10.3150/17-BEJ977

[21] GUDENDORF, G. and SEGERS, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications*. *Lect. Notes Stat. Proc.* **198** 127–145. Springer, Heidelberg. MR3051266 https://doi.org/10.1007/978-3-642-12465-5_6

[22] HALD, A. (1998). *A History of Mathematical Statistics from* 1750 *to* 1930. *Wiley Series in Probability and Statistics*: *Texts and References Section*. Wiley, New York. MR1619032

[23] JOE, H. (1997). *Multivariate Models and Dependence Concepts*. *Monographs on Statistics and Applied Probability* **73**. CRC Press, London. MR1462613 https://doi.org/10.1201/b13150

[24] JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics*: *Continuous Multivariate Distributions*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0418337

[25] KAZIANKA, H. and PILZ, J. (2011). Bayesian spatial modeling and interpolation using copulas. *Comput. Geosci.* **37** 310–319.

[26] KOTZ, S., BALAKRISHNAN, N. and JOHNSON, N. L. (2000). *Continuous Multivariate Distributions*: *Models and Applications*, *Vol.* 1, 2nd ed. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley-Interscience, New York. MR1788152 https://doi.org/10.1002/0471722065

[27] KOTZ, S. and NADARAJAH, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge Univ. Press, Cambridge. MR2038227 https://doi.org/10.1017/CBO9780511550683

[28] KOZUBOWSKI, T. J., PODGÓRSKI, K. and RYCHLIK, I. (2013). Multivariate generalized Laplace distribution and related random fields. *J. Multivariate Anal.* **113** 59–72. MR2984356 https://doi.org/10.1016/j.jmva.2012.02.010

[29] KRUPSKII, P. and GENTON, M. G. (2019). A copula model for non-Gaussian multivariate spatial data. *J. Multivariate Anal.* **169** 264–277. MR3875599 https://doi.org/10.1016/j.jmva.2018.09.007

[30] KRUPSKII, P. and JOE, H. (2013). Factor copula models for multivariate data. *J. Multivariate Anal.* **120** 85–101. MR3072719 https://doi.org/10.1016/j.jmva.2013.05.001

[31] KRUPSKII, P. and JOE, H. (2015). Structured factor copula models: Theory, inference and computation. *J. Multivariate Anal.* **138** 53–73. MR3348833 https://doi.org/10.1016/j.jmva.2014.11.002

[32] MARSHALL, A. W. and OLKIN, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11** 168–177. MR0682807

[33] MASAROTTO, G. and VARIN, C. (2012). Gaussian copula marginal regression. *Electron. J. Stat.* **6** 1517–1549. MR2988457 https://doi.org/10.1214/12-EJS721

[34] MAZO, G., GIRARD, S. and FORBES, F. (2016). A flexible and tractable class of one-factor copulas. *Stat. Comput.* **26** 965–979. MR3535039 https://doi.org/10.1007/s11222-015-9580-7

[35] MCNEIL, A. J. and NEŠLEHOVÁ, J. (2010). From Archimedean to Liouville copulas. *J. Multivariate Anal.* **101** 1772–1790. MR2651954 https://doi.org/10.1016/j.jmva.2010.03.015

[36] NASRI, B. R. (2020). On non-central squared copulas. *Statist. Probab. Lett.* **161** 108704, 10. MR4065482 https://doi.org/10.1016/j.spl.2020.108704

[37] NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

[38] NOH, H., EL GHOUCH, A. and BOUEZMARNI, T. (2013). Copula-based regression estimation and inference. *J. Amer. Statist. Assoc.* **108** 676–688. MR3174651 https://doi.org/10.1080/01621459.2013.783842

[39] OH, D. H. and PATTON, A. J. (2013). Simulated method of moments estimation for copula-based multivariate models. *J. Amer. Statist. Assoc.* **108** 689–700. MR3174652 https://doi.org/10.1080/01621459.2013.785952

[40] PICKANDS, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the* 43*rd Session of the International Statistical Institute*, *Vol.* 2 (*Buenos Aires*, 1981) **49** 859–878, 894–902. With a discussion. MR0820979

[41] QUESSY, J.-F. (2024). Supplement to "A general construction of multivariate dependence structures with non-monotone mappings and its applications." https://doi.org/10.1214/23-STS916SUPP

[42] QUESSY, J.-F. and DUROCHER, M. (2019). The class of copulas arising from squared distributions: Properties and inference. *Econom. Stat.* **12** 148–166. MR4008632 https://doi.org/10.1016/j.ecosta.2019.02.002

[43] QUESSY, J.-F., RIVEST, L.-P. and TOUPIN, M.-H. (2015). Semi-parametric pairwise inference methods in spatial models based on copulas. *Spat. Stat.* **14** 472–490. MR3431052 https://doi.org/10.1016/j.spasta.2015.08.002

[44] QUESSY, J.-F., RIVEST, L.-P. and TOUPIN, M.-H. (2016). On the family of multivariate chi-square copulas. *J. Multivariate Anal.* **152** 40–60. MR3554778 https://doi.org/10.1016/j.jmva.2016.07.007

[45] RADHAKRISHNA RAO, C. (1983). Multivariate analysis: Some reminiscences on its origin and development. *Sankhya*, *Ser. B* **45** 284–299.

[46] RILEY, S. N. (2017). Investigating the multivariate nature of NHL player performance with structural equation modeling. *PLoS ONE* **12** e0184346. https://doi.org/10.1371/journal.pone.0184346

[47] RODRÍGUEZ-LALLENA, J. A. and ÚBEDA-FLORES, M. (2004). A new class of bivariate copulas. *Statist. Probab. Lett.* **66** 315–325. MR2045476 https://doi.org/10.1016/j.spl.2003.09.010

[48] SARMANOV, O. V. (1966). Generalized normal correlation and two-dimensional Fréchet classes. *Dokl. Akad. Nauk SSSR* **168** 32–35. MR0196783

[49] SCHWEIZER, B. and SKLAR, A. (1958). Espaces métriques aléatoires. *C. R. Math. Acad. Sci. Paris* **247** 2092–2094. MR0099068

[50] SKLAR, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika* (*Prague*) **9** 449–460. MR0345164

[51] SKLAR, A. (1996). Random variables, distribution functions, and copulas—a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics* (*Seattle*, *WA*, 1993). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **28** 1–14. IMS, Hayward. MR1485519 https://doi.org/10.1214/lnms/1215452606

[52] SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

[53] SUROSO, S. and BÁRDOSSY, A. (2018). Multisite daily precipitation simulation in Singapore. *MATEC Web Conf.* **195** 05007. EDP Sciences.

[54] ZHAO, Y. and GENEST, C. (2019). Inference for elliptical copula multivariate response regression models. *Electron. J. Stat.* **13** 911–984. MR3934620 https://doi.org/10.1214/19-EJS1534

Statistical Science
2024, Vol. 39, No. 3, 409–427
https://doi.org/10.1214/23-STS917

Studentization Versus Variance Stabilization: A Simple Way Out of an Old Dilemma

Dimitris N. Politis

Abstract. Assume $\hat{\theta}_n$ is a statistic used to estimate a parameter θ on the basis of data $X_1, \ldots, X_n$. Further assume that $\hat{\theta}_n$ is consistent and asymptotically normal, with asymptotic variance given by $\sigma^2(\theta)$. Even if the functional form of $\sigma^2(\cdot)$ is known, its dependence on the unknown parameter θ creates a dilemma as regards the construction of a confidence interval for θ. Should the interval be based on the normal quantiles with estimated variance, that is, studentization, or shall we transform the statistic $\hat{\theta}_n$ to $Y_n = g(\hat{\theta}_n)$ such that the asymptotic variance of Y_n does not depend on θ, that is, variance stabilization? We show how this dilemma can be bypassed by a straightforward construction that applies rather generally, and just hinges on solving simple algebraic equations. We illustrate the new approach on a host of numerical examples, including two examples in nonparametric function estimation. In the latter, a different sort of dilemma arises: employing undersmoothing versus an explicit bias correction. This paper is dedicated to the memory of Dr. Dimitrios Gatzouras (1962–2020).

Key words and phrases: Bias correction, confidence intervals, Edgeworth expansion, finite-sample coverage, probability density estimation, undersmoothing.

REFERENCES

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. (With discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252. MR0192611

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. With comments and a rejoinder by the authors. MR1861069 https://doi.org/10.1214/ss/1009213286

CASELLA, G. and ROBERT, C. (1989). Refining Poisson confidence intervals. *Canad. J. Statist.* **17** 45–57. MR1014090 https://doi.org/10.2307/3314762

CROW, E. L. and GARDNER, R. S. (1959). Confidence intervals for the expectation of a Poisson variable. *Biometrika* **46** 441–453. MR0125681 https://doi.org/10.1093/biomet/46.3-4.441

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. *Springer Texts in Statistics*. Springer, New York. MR2664452

DECROUEZ, G. and HALL, P. (2014). Split sample methods for constructing confidence intervals for binomial and Poisson parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 949–975. MR3271174 https://doi.org/10.1111/rssb.12051

DICICCIO, T. and TIBSHIRANI, R. (1987). Bootstrap confidence intervals and bootstrap approximations. *J. Amer. Statist. Assoc.* **82** 163–170. MR0883344

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. With a discussion and a reply by the author. MR0959185 https://doi.org/10.1214/aos/1176350933

HALL, P. (1992a). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7

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

HALL, P. (1992c). On the removal of skewness by transformation. *J. Roy. Statist. Soc. Ser. B* **54** 221–228. MR1157721

JOHNSON, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *J. Amer. Statist. Assoc.* **73** 536–544. MR0514154

KREISS, J.-P. and PAPARODITIS, E. (2023). *Bootstrap for Time Series*: *Theory and Methods*. Springer, Heidelberg.

LAHIRI, S. N. (1997). Variance stabilizing transformations, Studentization and the bootstrap. *J. Statist. Plann. Inference* **61** 105–123. MR1455804 https://doi.org/10.1016/S0378-3758(97)89715-7

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. *Springer Series in Statistics*. Springer, New York. MR2001447

Dimitris N. Politis is a Distinguished Professor at the Department of Mathematics and the Halicioglu Data Science Institute, University of California, San Diego, La Jolla, California 92093-0112, USA (e-mail: dpolitis@ucsd.edu).

https://doi.org/10.1007/978-1-4757-3803-2

MCELROY, T. S. and POLITIS, D. N. (2020). *Time Series*: *A First Course with Bootstrap Starter*. CRC Press, Boca Raton, FL.

POLITIS, D. N. (2001). On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and Statistical Models with Applications* (C. Charalambides et al., eds.). 469–483. CRC Press, Boca Raton, FL.

POLITIS, D. N. (2003a). Adaptive bandwidth choice. *J. Nonparametr. Stat*. **15** 517–533. MR2017485 https://doi.org/10.1080/10485250310001604659

POLITIS, D. N. (2003b). The impact of bootstrap methods on time series analysis. *Statist. Sci*. **18** 219–230. MR2026081 https://doi.org/10.1214/ss/1063994977

POLITIS, D. N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744.

POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal*. **16** 67–103. MR1323618 https://doi.org/10.1111/j.1467-9892.1995.tb00223.x

POLITIS, D. N. ROMANO, J. P. and LAI, T. L. (1992). Bootstrap confidence bands for spectra and cross-spectra. *IEEE Trans. Signal Process*. **40** 1206–1215. https://doi.org/10.1109/MDSP.1989.97045

ROSENBLATT, M. (1991). *Stochastic Curve Estimation*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **3**. IMS, Hayward, CA.

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist*. **35** 1773–1801. MR2351105 https://doi.org/10.1214/009053606000001479

SHORACK, G. R. (2000). *Probability for Statisticians*. *Springer Texts in Statistics*. Springer, New York. MR1762415

THULIN, M. (2014). On split sample and randomized confidence intervals for binomial proportions. *Statist. Probab. Lett*. **92** 65–71. MR3230474 https://doi.org/10.1016/j.spl.2014.05.005

WILSON, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc*. **22** 209–212.

Statistical Science
2024, Vol. 39, No. 3, 428–448
https://doi.org/10.1214/23-STS918

Nonparametric Quantile Regression for Time Series with Replicated Observations and Its Application to Climate Data

Soudeep Deb and Kaushik Jana

Abstract. This paper proposes a model-free nonparametric estimator of conditional quantile of a time-series regression model where the covariate vector is repeated many times for different values of the response. This type of data abounds in climate studies. Although the use of quantile regression is standard in such studies, the opportunity to improve the results using the replicated nature of data is increasingly realized. The proposed method exploits this feature of the data and improves on the restrictive linear model structure of conventional quantile regression. Relevant asymptotic theories for the nonparametric estimators of the mean and variance function of the model are derived under a very general framework. We conduct a detailed simulation study that demonstrates the gain in efficiency of the proposed method over other benchmark models, especially when the actual data-generating process entails a nonlinear mean function and heteroskedastic pattern with time-dependent covariates. The predictive accuracy of the nonparametric method is remarkably high compared to other approaches when attention is on the higher quantiles of the variable of interest. The usefulness of the proposed method is then illustrated with two climatological applications, one with a well-known tropical cyclone wind-speed data and the other with an air pollution data.

Key words and phrases: Air pollution data, cyclone data, Nadaraya–Watson estimators, asymptotic theory.

REFERENCES

ABHILASH, M., THAKUR, A., GUPTA, D. and SREEVIDYA, B. (2018). Time series analysis of air pollution in bengaluru using ARIMA model. In *Ambient Communications and Computer Systems* 413–426. Springer, Berlin.

BERCU, B., CAPDEROU, S. and DURRIEU, G. (2019). A nonparametric statistical procedure for the detection of marine pollution. *J. Appl. Stat.* **46** 119–140. MR3870215 https://doi.org/10.1080/02664763.2018.1458824

BHATTACHARYA, I. and GHOSAL, S. (2021). Bayesian multivariate quantile regression using Dependent Dirichlet Process prior. *J. Multivariate Anal.* **185** 104763. MR4260589 https://doi.org/10.1016/j.jmva.2021.104763

BONDELL, H. D., REICH, B. J. and WANG, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* **97** 825–838. MR2746154 https://doi.org/10.1093/biomet/asq048

BOUZEBDA, S. and SLAOUI, Y. (2020). Bandwidth selector for nonparametric recursive density estimation for spatial data defined by stochastic approximation method. *Comm. Statist. Theory Methods* **49** 2942–2963. MR4095600 https://doi.org/10.1080/03610926.2019.1584313

BROWN, B. G., KATZ, R. W. and MURPHY, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *J. Appl. Meteorol. Climatol.* **23** 1184–1195.

CAI, Z. (2002). Regression quantiles for time series. *Econometric Theory* **18** 169–192. MR1885356 https://doi.org/10.1017/S0266466602181096

CHEN, I.-C., BERTKE, S. J. and CURWIN, B. D. (2021). Quantile regression for exposure data with repeated measures in the presence of non-detects. *J. Expo. Sci. Environ. Epidemiol.* 1–10.

CHRISTOU, E. (2022). quantdr: Dimension Reduction Techniques for Conditional Quantiles. R package version 1.2.2.

DABO-NIANG, S. and LAKSACI, A. (2012). Nonparametric quantile regression estimation for functional dependent data. *Comm. Statist. Theory Methods* **41** 1254–1268. MR2891592 https://doi.org/10.1080/03610926.2010.542850

DEB, S. and JANA, K. (2024). Supplement to "Nonparametric quantile regression for time series with replicated observations and its application to climate data." https://doi.org/10.1214/23-STS918SUPP

Soudeep Deb is Assistant Professor, Indian Institute of Management Bangalore, Bannerghatta Main Road, Bangalore, Karnataka 560076, India (e-mail: soudeep@iimb.ac.in). Kaushik Jana is Assistant Professor, Mathematical and Physical Sciences Division, Ahmedabad University, Navrangpura, Ahmedabad, Gujarat 380009, India (e-mail: kaushikjana11@gmail.com).

DELAIGLE, A. and GIJBELS, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249–267. MR2045631 https://doi.org/10.1016/S0167-9473(02)00329-8

DING, Q., SCHWEIGER, A., L'HEUREUX, M., BATTISTI, D. S., PO-CHEDLEY, S., JOHNSON, N. C., BLANCHARD-WRIGGLESWORTH, E., HARNOS, K., ZHANG, Q. et al. (2017). Influence of high-latitude atmospheric circulation changes on summertime Arctic sea ice. *Nat. Clim. Change* **7** 289–295.

DUNSON, D. B. and TAYLOR, J. A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Stat.* **17** 385–400. MR2129840 https://doi.org/10.1080/10485250500039049

ELSNER, J. B., KOSSIN, J. P. and JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455** 92–95. https://doi.org/10.1038/nature07234

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events. Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

FAN, J. and GIJBELS, I. (2018). *Local Polynomial Modelling and Its Applications*: *Monographs on Statistics and Applied Probability* **66**. Routledge, London.

FENG, Y., CHEN, Y. and HE, X. (2015). Bayesian quantile regression with approximate likelihood. *Bernoulli* **21** 832–850. MR3338648 https://doi.org/10.3150/13-BEJ589

FERNÁNDEZ, J. R., REDDEN, D. T., PIETROBELLI, A. and ALLISON, D. B. (2004). Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J. Pediatr.* **145** 439–444. https://doi.org/10.1016/j.jpeds.2004.06.044

GELFAND, A. E. and KOTTAS, A. (2003). Bayesian semiparametric regression for median residual life. *Scand. J. Stat.* **30** 651–665. MR2155475 https://doi.org/10.1111/1467-9469.00356

GILLILAND, J. M. and KEIM, B. D. (2018). Surface wind speed: Trend and climatology of Brazil from 1980–2014. *Int. J. Climatol.* **38** 1060–1073.

GREGORY, K. B., LAHIRI, S. N. and NORDMAN, D. J. (2018). A smooth block bootstrap for quantile regression with time series. *Ann. Statist.* **46** 1138–1166. MR3797999 https://doi.org/10.1214/17-AOS1580

GUTTIKUNDA, S. K., NISHADH, K. and JAWAHAR, P. (2019). Air pollution knowledge assessments (APnA) for 20 Indian cities. *Urban Clim.* **27** 124–141.

HENRY, R. C., CHANG, Y.-S. and SPIEGELMAN, C. H. (2002). Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. *Atmos. Environ.* **36** 2237–2244.

HONDA, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470. MR1794246 https://doi.org/10.1023/A:1004113201457

HONDA, T. (2013). Nonparametric quantile regression with heavy-tailed and strongly dependent errors. *Ann. Inst. Statist. Math.* **65** 23–47. MR3011612 https://doi.org/10.1007/s10463-012-0359-8

HUANG, Q., ZHANG, H., CHEN, J. and HE, M. (2017). Quantile regression models and their applications: A review. *J. Biometr. Biostat.* **8** 2155–6180.

JAGGER, T. H. and ELSNER, J. B. (2006). Climatology models for extreme hurricane winds near the United States. *J. Climate* **19** 3220–3236.

JAGGER, T. H. and ELSNER, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. *Int. J. Climatol.* **29** 1351–1361.

JANA, K. and SENGUPTA, D. (2022). Improving linear quantile regression for replicated data. *Statistics* **56** 1193–1206. MR4522127 https://doi.org/10.1080/02331888.2022.2152451

JANA, K., SENGUPTA, D., KUNDU, S., CHAKRABORTY, A. and SHAW, P. (2020). The statistical face of a region under monsoon rainfall in eastern India. *J. Amer. Statist. Assoc.* **115** 1559–1573. MR4189739 https://doi.org/10.1080/01621459.2019.1681275

JERRETT, M., BURNETT, R. T., MA, R., POPE III, C. A., KREWSKI, D., NEWBOLD, K. B., THURSTON, G., SHI, Y., FINKELSTEIN, N. et al. (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 727–736.

KANG, N.-Y., KIM, D. and ELSNER, J. B. (2019). The contribution of super typhoons to tropical cyclone activity in response to ENSO. *Sci. Rep.* **9** 1–6.

KNIGHT, K. (2001). Comparing conditional quantile estimators: First and second order considerations. Technical Report Citeseer. University of Toronto working paper.

KOENKER, R. (2005a). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657 https://doi.org/10.1017/CBO9780511754098

KOENKER, R. (2019). quantreg: Quantile Regression. R package version 5.54.

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

KOENKER, R. and PARK, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *J. Econometrics* **71** 265–283. MR1381085 https://doi.org/10.1016/0304-4076(96)84507-6

KOSSIN, J. P., OLANDER, T. L. and KNAPP, K. R. (2013). Trend analysis with a new global record of tropical cyclone intensity. *J. Climate* **26** 9960–9976.

KOTTAS, A. and GELFAND, A. E. (2001). Bayesian semiparametric mdeian regression modeling. *J. Amer. Statist. Assoc.* **96** 1458–1468. MR1946590 https://doi.org/10.1198/016214501753382363

KOTTAS, A. and KRNJAJIĆ, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.* **36** 297–319. MR2528986 https://doi.org/10.1111/j.1467-9469.2008.00626.x

LANZANTE, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.* **16** 1197–1226.

LEIDER, J. (2012). *A Quantile Regression Study of Climate Change in Chicago*, 1960–2010. Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL.

LI, D., LI, Q. and LI, Z. (2021). Nonparametric quantile regression estimation with mixed discrete and continuous data. *J. Bus. Econom. Statist.* **39** 741–756. MR4272932 https://doi.org/10.1080/07350015.2020.1730856

LI, Y., LIU, Y. and ZHU, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *J. Amer. Statist. Assoc.* **102** 255–268. MR2293307 https://doi.org/10.1198/016214506000000979

LIPSITZ, M., BELLONI, A., CHERNOZHUKOV, V. and FERNANDEZ-VAL, I. (2016). quantreg.nonpar: Nonparametric Series Quantile Regression. R package version 1.0.

LU, X. and FAN, Z. (2015). Weighted quantile regression for longitudinal data. *Comput. Statist.* **30** 569–592. MR3357076 https://doi.org/10.1007/s00180-014-0550-x

NICHOLLS, N. and KARIKO, A. (1993). East Australian rainfall events: Interannual variations, trends, and relationships with the southern oscillation. *J. Climate* **6** 1141–1152.

NICHOLLS, N., LANDSEA, C. and GILL, J. (1998). Recent trends in Australian region tropical cyclone activity. *Meteorol. Atmos. Phys.* **65** 197–205.

LONDON & PARTNERS (2021). London: Europe's global tech city. *DealRoom.co*.

REDDEN, D. T., FERNÁNDEZ, J. R. and ALLISON, D. B. (2004). A simple significance test for quantile regression. *Stat. Med.* **23** 2587–2597. https://doi.org/10.1002/sim.1839

REICH, B. J. and SMITH, L. B. (2013). Bayesian quantile regression for censored data. *Biometrics* **69** 651–660. MR3106593 https://doi.org/10.1111/biom.12053

RIVAS, M. D. G. and GONZALO, J. (2020). Trends in distributional characteristics: Existence of global warming. *J. Econometrics* **214** 153–174. MR4038227 https://doi.org/10.1016/j.jeconom.2019.05.009

RODRIGUES, T. and FAN, Y. (2017). Regression adjustment for noncrossing Bayesian quantile regression. *J. Comput. Graph. Statist.* **26** 275–284. MR3640185 https://doi.org/10.1080/10618600.2016.1172016

ROPELEWSKI, C. F. and BELL, M. A. (2008). Shifts in the statistics of daily rainfall in South America conditional on ENSO phase. *J. Climate* **21** 849–865.

SMITH, R. L., KOLENIKOV, S. and COX, L. H. (2003). Spatiotemporal modeling of PM2.5 data with missing values. *J. Geophys. Res., Atmos.* **108**.

TADDY, M. A. and KOTTAS, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *J. Bus. Econom. Statist.* **28** 357–369. MR2723605 https://doi.org/10.1198/jbes.2009.07331

THOMPSON, P., CAI, Y., MOYEED, R., REEVE, D. and STANDER, J. (2010). Bayesian nonparametric quantile regression using splines. *Comput. Statist. Data Anal.* **54** 1138–1150. MR2580945 https://doi.org/10.1016/j.csda.2009.09.004

THURSTON, G. D., BURNETT, R. T., TURNER, M. C., SHI, Y., KREWSKI, D., LALL, R., ITO, K., JERRETT, M., GAPSTUR, S. M. et al. (2016). Ischemic heart disease mortality and long-term exposure to source-related components of US fine particle air pollution. *Environ. Health Perspect.* **124** 785–794.

TOKDAR, S. T. and KADANE, J. B. (2012). Simultaneous linear quantile regression: A semiparametric Bayesian approach. *Bayesian Anal.* **7** 51–72. MR2896712 https://doi.org/10.1214/12-BA702

TORRES, J. L., GARCIA, A., DE BLAS, M. and DE FRANCISCO, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Sol. Energy* **79** 65–77.

VASSEUR, S. P. and AZNARTE, J. L. (2021). Comparing quantile regression methods for probabilistic forecasting of NO2 pollution levels. *Sci. Rep.* **11** 1–8.

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WATSON-PARRIS, D. (2021). Machine learning for weather and climate are worlds apart. *Philos. Trans. R. Soc. A* **379** 20200098. MR4236150 https://doi.org/10.1098/rsta.2020.0098

WEI, Y., KEHM, R. D., GOLDBERG, M. and TERRY, M. B. (2019). Applications for quantile regression in epidemiology. *Curr. Epidemiol. Rep.* **6** 191–199.

WU, T. and NARISETTY, N. N. (2021). Bayesian multiple quantile regression for linear models using a score likelihood. *Bayesian Anal.* **16** 875–903. MR4303872 https://doi.org/10.1214/20-BA1217

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

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

YANG, Y. and TOKDAR, S. T. (2017). Joint estimation of quantile planes over arbitrary predictor spaces. *J. Amer. Statist. Assoc.* **112** 1107–1120. MR3735363 https://doi.org/10.1080/01621459.2016.1192545

YANG, Y., WANG, H. J. and HE, X. (2016). Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *Int. Stat. Rev.* **84** 327–344. MR3580414 https://doi.org/10.1111/insr.12114

YING, M., CHEN, B. and WU, G. (2011). Climate trends in tropical cyclone-induced wind and precipitation over mainland China. *Geophys. Res. Lett.* **38**.

YU, K. and JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628 https://doi.org/10.2307/2669619

YU, K. and MOYEED, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54** 437–447. MR1861390 https://doi.org/10.1016/S0167-7152(01)00124-9

ZHAO, Z. and WU, W. B. (2006). Kernel quantile regression for nonlinear stochastic models. *Technical ReportNo* **572**.

ZHAO, Z. and WU, W. B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36** 1854–1878. MR2435458 https://doi.org/10.1214/07-AOS533

ZIEGELMANN, F. A. (2005). A nonparametric least-absolute-deviations estimator of volatility functions. *XXVII Encontro Brasileiro de Econometria, Natal* **10**.

Statistical Science
2024, Vol. 39, No. 3, 449–468
https://doi.org/10.1214/24-STS922

Protecting Classifiers from Attacks

Víctor Gallego, Roi Naveiro, Alberto Redondo, David Ríos Insua and Fabrizio Ruggeri

Abstract. In multiple domains such as malware detection, automated driving systems, or fraud detection, classification algorithms are susceptible to being attacked by malicious agents willing to perturb the value of instance covariates to pursue certain goals. Such problems pertain to the field of adversarial machine learning and have been mainly dealt with, perhaps implicitly, through game-theoretic ideas with strong underlying common knowledge assumptions. These are not realistic in numerous application domains in relation to security and business competition. We present an alternative Bayesian decision theoretic framework that accounts for the uncertainty about the attacker's behavior using adversarial risk analysis concepts. In doing so, we also present core ideas in adversarial machine learning to a statistical audience. A key ingredient in our framework is the ability to sample from the distribution of originating instances given the, possibly attacked, observed ones. We propose an initial procedure based on approximate Bayesian computation usable during operations; within it, we simulate the attacker's problem taking into account our uncertainty about his elements. Large-scale problems require an alternative scalable approach implementable during the training stage. Globally, we are able to robustify statistical classification algorithms against malicious attacks.

Key words and phrases: Classification, Bayesian methods, adversarial machine learning, adversarial risk analysis, deep models.

REFERENCES

BANKS, D. L., RIOS, J. and RÍOS INSUA, D. (2016). *Adversarial Risk Analysis*. CRC Press, Boca Raton, FL. MR3409217

BARRENO, M., NELSON, B., SEARS, R., JOSEPH, A. D. and TYGAR, J. D. (2006). Can machine learning be secure? In *Proceedings of the* 2006 *ACM Symposium on Information, Computer and Communications Security* 16–25. ACM, New York.

BIGGIO, B. and ROLI, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognit*. **84** 317–331.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

BORTOT, P., COLES, S. G. and SISSON, S. A. (2007). Inference for stereological extremes. *J. Amer. Statist. Assoc*. **102** 84–92. MR2345549 https://doi.org/10.1198/016214506000000988

BOTTOU, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'*2010 177–186. Physica-Verlag, Heidelberg. MR3362066

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

CABALLERO, W. N., RIOS INSUA, D. and NAVEIRO, R. (2023). Some statistical challenges in automated driving systems. *Appl. Stoch. Models Bus. Ind*. **39** 629–652. MR4662205

CARLINI, N. and WAGNER, D. (2017). Towards evaluating the robustness of neural networks. In 2017 *IEEE Symposium on Security and Privacy* (*SP*) 39–57. IEEE Press, New York.

CASELLA, G., ROBERT, C. P. and WELLS, M. T. (2004). Generalized accept–reject sampling schemes. In *A Festschrift for Herman Rubin*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 342–347. IMS, Beachwood, OH. MR2126910 https://doi.org/10.1214/lnms/1196285403

CHAI, X., DENG, L., YANG, Q. and LING, C. X. (2004). Test-cost sensitive naive Bayes classification. In *Fourth IEEE International Conference on Data Mining* (*ICDM'*04) 51–58. IEEE Press, New York.

CHUNG, K. L. (2001). *A Course in Probability Theory*, 3rd ed. Academic Press, San Diego, CA. MR1796326

COMITER, M. (2019). Attacking artificial intelligence. Belfer Center paper.

EUROPEAN COMMISSION (2022). Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union Legislative acts. European Commission.

CSILLÉRY, K., BLUM, M. G. B., GAGGIOTTI, O. E. and FRANÇOIS, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol*. **25** 410–418. https://doi.org/10.1016/j.tree.2010.04.001

Víctor Gallego is PhD, Komorebi AI Technologies, Madrid, Spain. Roi Naveiro is Assistant Professor, CUNEF Universidad, Madrid, Spain (e-mail: roi.naveiro@cunef.edu). Alberto Redondo is PhD, Institute of Mathematical Sciences (ICMAT-CSIC), Madrid, Spain. David Ríos Insua is Research Professor, Institute of Mathematical Sciences (ICMAT-CSIC), Madrid, Spain. Fabrizio Ruggeri is Senior Fellow, CNR-IMATI, Milano, Italy.

DALVI, N., DOMINGOS, P., MAUSAM, S. S. and VERMA, D. (2004). Adversarial classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. *KDD* '04 99–108.

EKIN, T., NAVEIRO, R., RÍOS INSUA, D. and TORRES-BARRÁN, A. (2023). Augmented probability simulation methods for sequential games. *European J. Oper. Res.* **306** 418–430. MR4522449 https://doi.org/10.1016/j.ejor.2022.06.042

FEARNHEAD, P. and PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 419–474. MR2925370 https://doi.org/10.1111/j.1467-9868.2011.01010.x

FRENCH, S. and RIOS INSUA, D. (2000). *Statistical Decision Theory*. Wiley, New York.

GALLEGO, V. and INSUA, D. R. (2018). Stochastic gradient MCMC with repulsive forces. Bayesian Deep Learning Workshop, Neural Information and Processing Systems (NIPS). arXiv preprint, available at arXiv:1812.00071.

GALLEGO, V. and RÍOS INSUA, D. (2022). Current advances in neural networks. *Annu. Rev. Stat. Appl.* **9** 197–222. MR4394906 https://doi.org/10.1146/annurev-statistics-040220-112019

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* 27 (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) 2672–2680. Curran Associates, Red Hook.

GOODFELLOW, I., SHLENS, J. and SZEGEDY, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint, available at arXiv:1412.6572.

GRATHWOHL, W., WANG, K.-C., JACOBSEN, J.-H., DUVENAUD, D., NOROUZI, M. and SWERSKY, K. (2019). Your classifier is secretly an energy based model and you should treat it like one. In *International Conference on Learning Representations*.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HANEA, A. M., NANE, G. F., BEDFORD, T. and FRENCH, S., eds. (2021). *Expert Judgement in Risk and Decision Analysis*. *International Series in Operations Research & Management Science* **293**. Springer, Cham. MR4238538 https://doi.org/10.1007/978-3-030-46474-5

HARGREAVES-HEAP, S. and VAROUFAKIS, Y. (2004). *Game Theory*: *A Critical Introduction*. Routledge, London.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

HE, K., ZHANG, X., REN, S. and SUN, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778.

HUANG, L., JOSEPH, A. D., NELSON, B., RUBINSTEIN, B. I. P. and TYGAR, J. D. (2011). Adversarial machine learning. In *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*. *AISec* '11 43–58.

JOSEPH, A., MELSON, B., RUBISNTEIN, B. and TYGAR, J. D. (2019). *Adversarial Machine Learning*. Cambridge Univ. Press, Cambridge.

JOSHI, C., ALIAGA, J. R. and INSUA, D. R. (2021). Insider threat modeling: An adversarial risk analysis approach. *IEEE Trans. Inform. Forensics Secur.* **16** 1131–1142. https://doi.org/10.1109/TIFS.2020.3029898

KADANE, J. B. and LARKEY, P. D. (1982). Subjective probability and the theory of games. *Manage. Sci.* **28** 113–125. MR0652016 https://doi.org/10.1287/mnsc.28.2.113

KANNAN, H., KURAKIN, A. and GOODFELLOW, I. (2018). Adversarial logit pairing. arXiv preprint, available at arXiv:1803.06373.

KATZ, G., BARRETT, C., DILL, D. L., JULIAN, K. and KOCHENDERFER, M. J. (2017). Towards proving the adversarial robustness of deep neural networks. arXiv preprint, available at arXiv:1709.02802.

KIM, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput. Statist. Data Anal.* **53** 3735–3745. MR2749918 https://doi.org/10.1016/j.csda.2009.04.009

KINGMA, D. P., MOHAMED, S., REZENDE, D. J. and WELLING, M. (2014). Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems* 3581–3589.

KOTZIAS, D., DENIL, M., DE FREITAS, N. and SMYTH, P. (2015). From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 597–606.

KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 1097–1105.

LAKSHMINARAYANAN, B., PRITZEL, A. and BLUNDELL, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* **30** 6402–6413.

LECUN, Y., CORTES, C. and BURGES, C. (1998). THE MNIST DATABASE of handwritten digits. Available at http://yann.lecun.com/exdb/mnist/.

MA, Y.-A., CHEN, T. and FOX, E. (2015). A complete recipe for stochastic gradient MCMC. *Adv. Neural Inf. Process. Syst.* 2917–2925.

MADRY, A., MAKELOV, A., SCHMIDT, L., TSIPRAS, D. and VLADU, A. (2018). Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*.

MARTIN, G. M., FRAZIER, D. T. and ROBERT, C. P. (2024). Approximating Bayes in the 21st century. *Statist. Sci.* **39** 20–45. MR4718525 https://doi.org/10.1214/22-sts875

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. Second edition of [MR0727836]. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

MICRO, T. (2020). Exploiting AI: How cybercriminals misuse and abuse AI and ML. Trend micro.

MILLER, J. W. and DUNSON, D. B. (2019). Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* **114** 1113–1125. MR4011766 https://doi.org/10.1080/01621459.2018.1469995

MORRIS, D., OAKLEY, J. and CROWE, J. (2014). A web-based tool for eliciting probability distributions from experts. *Environ. Model. Softw.* **52** 1–4.

MÜLLER, P. and INSUA, D. R. (1998). Issues in Bayesian analysis of neural network models. *Neural Comput.* **10** 749–770. https://doi.org/10.1162/089976698300017737

NAVEIRO, R., REDONDO, A., RÍOS INSUA, D. and RUGGERI, F. (2019). Adversarial classification: An adversarial risk analysis approach. *Internat. J. Approx. Reason.* **113** 133–148. MR3983223 https://doi.org/10.1016/j.ijar.2019.07.003

NIST (2022). AI risk management framework: Second draft. NIST.

PAPAMAKARIOS, G. and MURRAY, I. (2016). Fast ε-free inference of simulation models with Bayesian conditional density estimation. *Adv. Neural Inf. Process. Syst.* **29**.

PLATT, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Larg. Margin Classif.* **10** 61–74.

RAIFFA, H. (1982). *The Art and Science of Negotiation*. Harvard Univ. Press, Cambridge.

RÍOS INSUA, D., GONZÁLEZ-ORTEGA, J., BANKS, D. and RÍOS, J. (2018). Concept uncertainty in adversarial statistical decision theory. In *The Mathematics of the Uncertain. Stud. Syst. Decis. Control* **142** 527–542. Springer, Cham. MR3793229

RÍOS INSUA, D. and RUGGERI, F., eds. (2000). *Robust Bayesian Analysis. Lecture Notes in Statistics* **152**. Springer, New York. MR1795206 https://doi.org/10.1007/978-1-4612-1306-2

RIOS INSUA, I., RIOS, J. and BANKS, D. (2009). Adversarial risk analysis. *J. Amer. Statist. Assoc.* **104** 841–854. MR2541597 https://doi.org/10.1198/jasa.2009.0155

RIOS, J. and RIOS INSUA, D. (2012). Adversarial risk analysis for counterterrorism modeling. *Risk Anal.* **32** 894–915. MR2946741 https://doi.org/10.1287/deca.1110.0225

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

STAHL, D. O. and WILSON, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games Econom. Behav.* **10** 218–254. MR1347938 https://doi.org/10.1006/game.1995.1031

SU, J., VARGAS, V. and KOUICHI, S. (2017). One pixel attack for fooling deep neural networks. Available at arXiv:1710.08864.

SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I. and FERGUS, R. (2014). Intriguing properties of neural networks. arXiv preprint, available at arXiv:1312.6199.

TABASSI, E., BURNS, K., HADJIMICHAEL, M., MOLINA-MARKHAM, A. and SEXTON, J. (2020). A taxonomy and terminology of adversarial machine learning. NIST-MITRE.

THE WHITE HOUSE (2023). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Briefing Room, Presidential Actions.

VOROBEICHYK, Y. and KANTARCIOGLU, M. (2019). *Adversarial Machine Learning*. Morgan & Claypool, CA.

WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the* 28*th International Conference on Machine Learning* (*ICML*-11) 681–688.

WIPER, M., RIOS INSUA, D. and RUGGERI, F. (2001). Mixtures of gamma distributions with applications. *J. Comput. Graph. Statist.* **10** 440–454. MR1939034 https://doi.org/10.1198/106186001317115054

YANG, R. and BERGER, J. (1997). A catalog of noninformative priors. ISDS Discussion paper 97-42.

ZHOU, Y., KANTARCIOGLU, M. and XI, B. (2018). A survey of game theoretic approach for adversarial machine learning. Wiley interdisciplinary reviews: Data mining and knowledge discovery e1259.

Statistical Science
2024, Vol. 39, No. 3, 469–492
https://doi.org/10.1214/24-STS923

The Matérn Model: A Journey Through Statistics, Numerical Analysis and Machine Learning

Emilio Porcu, Moreno Bevilacqua, Robert Schaback and Chris J. Oates

Abstract. The Matérn model has been a cornerstone of spatial statistics for more than half a century. More recently, the Matérn model has been exploited in disciplines as diverse as numerical analysis, approximation theory, computational statistics, machine learning, and probability theory. In this article, we take a Matérn-based journey across these disciplines. First, we reflect on the importance of the Matérn model for estimation and prediction in spatial statistics, establishing also connections to other disciplines in which the Matérn model has been influential. Then, we position the Matérn model within the literature on big data and scalable computation: the SPDE approach, the Vecchia likelihood approximation, and recent applications in Bayesian computation are all discussed. Finally, we review recent devlopments, including flexible alternatives to the Matérn model, whose performance we compare in terms of estimation, prediction, screening effect, computation, and Sobolev regularity properties.

Key words and phrases: Approximation theory, compact support, covariance, kernel, kriging, machine learning, maximum likelihood, reproducing kernel Hilbert spaces, spatial statistics, Sobolev spaces.

REFERENCES

[1] Åberg, S. and Podgórski, K. (2011). A class of non-Gaussian second order random fields. *Extremes* **14** 187–222. MR2794324 https://doi.org/10.1007/s10687-010-0119-1

[2] Abramowitz, M. and Stegun, I. A., eds. (1970). *Handbook of Mathematical Functions* Dover, New York.

[3] Alegría, A., Cuevas-Pacheco, F., Diggle, P. and Porcu, E. (2021). The $\mathcal{F}$-family of covariance functions: A Matérn analogue for modeling random fields on spheres. *Spat. Stat.* **43** Paper No. 100512, 25 pp. MR4253853 https://doi.org/10.1016/j.spasta.2021.100512

[4] Alegría, A., Emery, X. and Porcu, E. (2021). Bivariate Matérn covariances with cross-dimple for modeling coregionalized variables. *Spat. Stat.* **41** Paper No. 100491, 16 pp. MR4205104 https://doi.org/10.1016/j.spasta.2021.100491

[5] Allard, D., Clarotto, L. and Emery, X. (2022). Fully nonseparable Gneiting covariance functions for multivariate space-time data. *Spat. Stat.* **52** Paper No. 100706, 24 pp. MR4496718 https://doi.org/10.1016/j.spasta.2022.100706

[6] Allard, D., Senoussi, R. and Porcu, E. (2016). Anisotropy models for spatial data. *Math. Geosci.* **48** 305–328. MR3476704 https://doi.org/10.1007/s11004-015-9594-x

[7] Anderes, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700 https://doi.org/10.1214/09-AOS725

[8] 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

[9] Angulo, J. M., Kelbert, M. Y., Leonenko, N. N. and Ruiz-Medina, M. D. (2008). Spatiotemporal random fields associated with stochastic fractional Helmholtz and heat equations. *Stoch. Environ. Res. Risk Assess.* **22** 3–13. MR2416949 https://doi.org/10.1007/s00477-007-0157-7

[10] Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Amer. Statist. Assoc.* **107** 180–193. MR2949350 https://doi.org/10.1080/01621459.2011.643197

[11] Askey, R. (1973). Radial characteristic functions. Technical report, Research Center, Univ. Wisconsin.

Emilio Porcu is Professor, Department of Mathematics, Khalifa University, Abu Dhabi, The United Arab Emirates and Research Fellow, ADIA Lab, Abu Dhabi, The United Arab Emirates (e-mail: emilio.porcu@bku.ac.ae). Moreno Bevilacqua is Professor, Department of Statistics, Universidad Adolfo Ibanez, Santiago, Chile (e-mail: emilio.porcu@bku.ac.ae). Robert Schaback is Professor, Department of Mathematics, University of Göttingen, Göttingen, Germany (e-mail: emilio.porcu@bku.ac.ae). Chris J. Oates is Professor, School of Mathematics, Statistics & Physics, Newcastle University, Newcastle, UK (e-mail: emilio.porcu@bku.ac.ae).

[12] Bachoc, F., Bevilacqua, M. and Velandia, D. (2019). Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. *J. Multivariate Anal.* **174** 104534, 17 pp. MR3985049 https://doi.org/10.1016/j.jmva.2019.104534

[13] Bachoc, F., Porcu, E., Bevilacqua, M., Furrer, R. and Faouzi, T. (2022). Asymptotically equivalent prediction in multivariate geostatistics. *Bernoulli* **28** 2518–2545. MR4474552 https://doi.org/10.3150/21-bej1427

[14] Bakka, H., Krainski, E., Bolin, D., Rue, H. and Lindgren, F. (2020). The diffusion-based extension of the Matérn field to space-time. Preprint. Available at arXiv:2006.04917.

[15] Banerjee, S. and Gelfand, A. E. (2006). Bayesian wombling: Curvilinear gradient assessment under spatial process models. *J. Amer. Statist. Assoc.* **101** 1487–1501. MR2279474 https://doi.org/10.1198/016214506000000041

[16] Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 https://doi.org/10.1111/j.1467-9868.2008.00663.x

[17] Banerjee, S., Gelfand, A. E. and Sirmans, C. F. (2003). Directional rates of change under spatial process models. *J. Amer. Statist. Assoc.* **98** 946–954. MR2041483 https://doi.org/10.1198/C16214503000000909

[18] Barp, A., Oates, C. J., Porcu, E. and Girolami, M. (2022). A Riemann–Stein kernel method. *Bernoulli* **28** 2181–2208. MR4474540 https://doi.org/10.3150/21-bej1415

[19] Benzi, M. (2016). Localization in matrix computations: Theory and applications. In *Exploiting Hidden Structure in Matrix Computations*: *Algorithms and Applications*. *Lecture Notes in Math.* **2173** 211–317. Springer, Cham. MR3618088

[20] Berg, C., Mateu, J. and Porcu, E. (2008). The Dagum family of isotropic correlation functions. *Bernoulli* **14** 1134–1149. MR2543589 https://doi.org/10.3150/08-BEJ139

[21] Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B. and Morales-Oñate, V. (2021). Non-Gaussian geostatistical modeling using (skew) t processes. *Scand. J. Stat.* **48** 212–245. MR4233171 https://doi.org/10.1111/sjos.12447

[22] Bevilacqua, M., Caamaño-Carrillo, C. and Porcu, E. (2022). Unifying compactly supported and Matérn covariance functions in spatial statistics. *J. Multivariate Anal.* **189** Paper No. 104949, 17 pp. MR4375672 https://doi.org/10.1016/j.jmva.2022.104949

[23] Bevilacqua, M., Faouzi, T., Furrer, R. and Porcu, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Ann. Statist.* **47** 828–856. MR3909952 https://doi.org/10.1214/17-AOS1652

[24] Bevilacqua, M. and Gaetan, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Stat. Comput.* **25** 877–892. MR3375623 https://doi.org/10.1007/s11222-014-9460-6

[25] Bevilacqua, M., Morales-Oñate, V. and Caamaño-Carrillo, C. (2023). GeoModels: Procedures for Gaussian and non Gaussian geostatistical (large) data analysis. R package version 1.1.5.

[26] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR1015093

[27] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Univ. California Press, Berkeley-Los Angeles, CA. MR0072370

[28] Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scand. J. Stat.* **41** 557–579. MR3249417 https://doi.org/10.1111/sjos.12046

[29] Bolin, D. and Kirchner, K. (2020). The rational SPDE approach for Gaussian random fields with general smoothness. *J. Comput. Graph. Statist.* **29** 274–285. MR4116041 https://doi.org/10.1080/10618600.2019.1665537

[30] Bolin, D. and Kirchner, K. (2023). Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. *Bernoulli* **29** 1476–1504. MR4550232 https://doi.org/10.3150/22-bej1507

[31] 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

[32] Bolin, D. and Wallin, J. (2020). Multivariate type G Matérn stochastic partial differential equation random fields. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 215–239. MR4060983

[33] Borovitskiy, V., Azangulov, I., Terenin, A., Mostowsky, P., Deisenroth, M. and Durrande, N. (2021). Matérn Gaussian processes on graphs. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 2593–2601. PMLR.

[34] Borovitskiy, V., Terenin, A., Mostowsky, P. et al. (2020). Matérn Gaussian processes on Riemannian manifolds. In *Advances in Neural Information Processing Systems* **33** 12426–12437.

[35] Brown, R. D., Bardsley, J. M. and Cui, T. (2020). Semivariogram methods for modeling Whittle–Matérn priors in Bayesian inverse problems. *Inverse Probl.* **36** 055006, 27 pp. MR4105319 https://doi.org/10.1088/1361-6420/ab762e

[36] Buhmann, M. D. (2001). A new class of radial basis functions with compact support. *Math. Comp.* **70** 307–318. MR1803129 https://doi.org/10.1090/S0025-5718-00-01251-5

[37] Cameletti, M., Lindgren, F., Simpson, D. and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Adv. Stat. Anal.* **97** 109–131. MR3045763 https://doi.org/10.1007/s10182-012-0196-3

[38] Chernih, A. and Hubbert, S. (2014). Closed form representations and properties of the generalised Wendland functions. *J. Approx. Theory* **177** 17–33. MR3128401 https://doi.org/10.1016/j.jat.2013.09.005

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

[40] Clarotto, L., Allard, D., Romary, T. and Desassis, N. (2022). The SPDE approach for spatio-temporal datasets with advection and diffusion: A matrix-free approach. In *15th International Conference of the ERCIM WG on Com-Putational and Methodological Statistics* (*CMStatistics* 2022).

[41] Cockayne, J., Oates, C. J., Sullivan, T. J. and Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM Rev.* **61** 756–789. MR4027836 https://doi.org/10.1137/17M1139357

[42] Cressie, N. (1990). The origins of kriging. *Math. Geol.* **22** 239–252. MR1047810 https://doi.org/10.1007/BF00889887

[43] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639 https://doi.org/10.1111/j.1467-9868.2007.00633.x

[44] DALEY, D. J., PORCU, E. and BEVILACQUA, M. (2015). Classes of compactly supported covariance functions for multivariate random fields. *Stoch. Environ. Res. Risk Assess.* **29** 1249–1263.

[45] DATTA, A. (2022). Nearest-neighbor sparse Cholesky matrices in spatial statistics. *Wiley Interdiscip. Rev.: Comput. Stat.* **14** Paper No. e1574, 21 pp. MR4483687

[46] DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706 https://doi.org/10.1080/01621459.2015.1044091

[47] DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Ann. Appl. Stat.* **10** 1286–1316. MR3553225 https://doi.org/10.1214/16-AOAS931

[48] DAVYDOV, O. and SCHABACK, R. (2019). Optimal stencils in Sobolev spaces. *IMA J. Numer. Anal.* **39** 398–422. MR4023749 https://doi.org/10.1093/imanum/drx076

[49] DEY, D., DATTA, A. and BANERJEE, S. (2022). Graphical Gaussian process models for highly multivariate spatial data. *Biometrika* **109** 993–1014. MR4519112 https://doi.org/10.1093/biomet/asab061

[50] EIDSVIK, J., SHABY, B. A., REICH, B. J., WHEELER, M. and NIEMI, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.* **23** 295–315. MR3215812 https://doi.org/10.1080/10618600.2012.760460

[51] EMERY, X. and ALEGRÍA, A. (2022). The Gauss hypergeometric covariance kernel for modeling second-order stationary random fields in Euclidean spaces: Its compact support, properties and spectral representation. *Stoch. Environ. Res. Risk Assess.* **36** 2819–2834.

[52] EMERY, X., PORCU, E. and WHITE, P. (2022). New validity conditions for the multivariate Matérn coregionalization model, with an application to exploration geochemistry. *Math. Geosci.* **54** 1043–1068. MR4455143 https://doi.org/10.1007/s11004-022-10000-6

[53] FAOUZI, T., PORCU, E. and BEVILACQUA, M. (2022). Space-time estimation and prediction under fixed-domain asymptotics with compactly supported covariance functions. *Statist. Sinica* **32** 1187–1203. MR4449890 https://doi.org/10.5705/ss.202020.0010

[54] FAOUZI, T., PORCU, E., BEVILACQUA, M. and KONDRASHUK, I. (2020). Zastavnyi operators and positive definite radial functions. *Statist. Probab. Lett.* **157** 108620, 7 pp. MR4016601 https://doi.org/10.1016/j.spl.2019.108620

[55] FASSHAUER, G. E. (1997). Solving partial differential equations by collocation with radial basis functions. In *Surface Fitting and Multiresolution Methods* (A. LeMéhauté, C. Rabut and L. Schumaker, eds.). Vanderbilt University Press, Nashville, TN.

[56] FASSHAUER, G. E. and MCCOURT, M. (2015). *Kernel-Based Approximation Methods Using MATLAB. Interdisciplinary Mathematical Sciences* **19**. World Scientific, Singapore.

[57] FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261 https://doi.org/10.1198/106186006X132178

[58] GENTON, M. G. and KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist. Sci.* **30** 147–163. MR3353096 https://doi.org/10.1214/14-STS487

[59] GNEITING, T. (2002). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508. MR1945966 https://doi.org/10.1006/jmva.2001.2056

[60] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

[61] GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

[62] GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

[63] GNEITING, T. and SCHLATHER, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Rev.* **46** 269–282. MR2114455 https://doi.org/10.1137/S0036144501394387

[64] GNEITING, T., ŠEVČÍKOVÁ, H. and PERCIVAL, D. B. (2012). Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statist. Sci.* **27** 247–277. MR2963995 https://doi.org/10.1214/11-STS370

[65] GOUGH, B. (2009). GNU scientific library reference manual. Network Theory Ltd.

[66] GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60** 415–429. MR3878098 https://doi.org/10.1080/00401706.2018.1437476

[67] GUINNESS, J. (2022). Inverses of Matérn covariances on grids. *Biometrika* **109** 535–541. MR4430973 https://doi.org/10.1093/biomet/asab017

[68] GUINNESS, J. and FUENTES, M. (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. *J. Multivariate Anal.* **143** 143–152. MR3431424 https://doi.org/10.1016/j.jmva.2015.08.018

[69] GUTTORP, P. and GNEITING, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika* **93** 989–995. MR2285084 https://doi.org/10.1093/biomet/93.4.989

[70] HALDER, A., BANERJEE, S. and DEY, D. K. (2023). Bayesian modeling with spatial curvature processes. *J. Amer. Statist. Assoc.* 1–27.

[71] HARRAR, S. W., SENETA, E. and GUPTA, A. K. (2006). Duality between matrix variate t and matrix variate V.G. distributions. *J. Multivariate Anal.* **97** 1467–1475. MR2256162 https://doi.org/10.1016/j.jmva.2005.09.002

[72] HARTIKAINEN, J. and SÄRKKÄ, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In 2010 *IEEE International Workshop on Machine Learning for Signal Processing* 379–384. IEEE.

[73] HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

[74] HENNIG, P., OSBORNE, M. A. and KERSTING, H. P. (2022). *Probabilistic Numerics—Computation as Machine Learning*. Cambridge Univ. Press, Cambridge. MR4420061 https://doi.org/10.1017/9781316681411

[75] HUBBERT, S. (2012). Closed form representations for a class of compactly supported radial basis functions. *Adv. Comput. Math.* **36** 115–136. MR2886183 https://doi.org/10.1007/s10444-011-9184-5

[76] IP, R. H. L. and LI, W. K. (2017). On some Matérn covariance functions for spatio-temporal random fields. *Statist. Sinica* **27** 805–822. MR3674697

[77] JANSSON, E., KOVÁCS, M. and LANG, A. (2022). Surface finite element approximation of spherical Whittle–Matérn Gaussian random fields. *SIAM J. Sci. Comput.* **44** A825–A842. MR4405199 https://doi.org/10.1137/21M1400717

[78] JAQUIER, N., BOROVITSKIY, V., SMOLENSKY, A., TERENIN, A., ASFOUR, T. and ROZO, L. (2022). Geometry-aware Bayesian optimization in robotics using Riemannian Matérn kernels. In *Conference on Robot Learning* 794–805. PMLR.

[79] JENSEN, K., KAO, T.-C., TRIPODI, M. and HENNEQUIN, G. (2020). Manifold gplvms for discovering non-Euclidean latent structure in neural data. *Adv. Neural Inf. Process. Syst.* **33** 22580–22592.

[80] KARVONEN, T. (2022). Asymptotic bounds for smoothness parameter estimates in Gaussian process regression. Preprint. Available at arXiv:2203.05400.

[81] KARVONEN, T. and OATES, C. J. (2023). Maximum likelihood estimation in Gaussian process regression is ill-posed. *J. Mach. Learn. Res.* **24** Paper No. 120, 47 pp. MR4583281

[82] KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. MR4194207 https://doi.org/10.1214/19-STS755

[83] KAUFMAN, C. G. and SHABY, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100** 473–484. MR3068447 https://doi.org/10.1093/biomet/ass079

[84] KELBERT, M. Y., LEONENKO, N. N. and RUIZ-MEDINA, M. D. (2005). Fractional random fields associated with stochastic fractional heat equations. *Adv. in Appl. Probab.* **37** 108–133. MR2135156 https://doi.org/10.1239/aap/1113402402

[85] KENT, J. T. (1989). Continuity properties for random fields. *Ann. Probab.* **17** 1432–1440. MR1048935

[86] KHRISTENKO, U., SCARABOSIO, L., SWIERCZYNSKI, P., ULLMANN, E. and WOHLMUTH, B. (2019). Analysis of boundary effects on PDE-based sampling of Whittle–Matérn random fields. *SIAM/ASA J. Uncertain. Quantificat.* **7** 948–974. MR3984316 https://doi.org/10.1137/18M1215700

[87] KLEIBER, W. and NYCHKA, D. (2012). Nonstationary modeling for multivariate spatial processes. *J. Multivariate Anal.* **112** 76–91. MR2957287 https://doi.org/10.1016/j.jmva.2012.05.011

[88] KLEIBER, W. and PORCU, E. (2015). Nonstationary matrix covariances: Compact support, long range dependence and quasi-arithmetic constructions. *Stoch. Environ. Res. Risk Assess.* **29** 193–204.

[89] KOEPERNIK, P. and PFAFF, F. (2021). Consistency of Gaussian process regression in metric spaces. *J. Mach. Learn. Res.* **22** Paper No. 244, 27 pp. MR4329823 https://doi.org/10.1080/15502287.2021.1916700

[90] LANG, A. and PEREIRA, M. (2023). Galerkin–Chebyshev approximation of Gaussian random fields on compact Riemannian manifolds. *BIT* **63** Paper No. 51, 44 pp. MR4653424 https://doi.org/10.1007/s10543-023-00986-8

[91] LAVIN, A., ZENIL, H., PAIGE, B., KRAKAUER, D., GOTTSCHLICH, J., MATTSON, T., ANANDKUMAR, A., CHOUDRY, S., ROCKI, K. et al. (2021). Simulation intelligence: Towards a new generation of scientific methods. Preprint. Available at arXiv:2112.03235.

[92] LEONENKO, N. and MALYARENKO, A. (2017). Matérn class tensor-valued random fields and beyond. *J. Stat. Phys.* **168** 1276–1301. MR3691251 https://doi.org/10.1007/s10955-017-1847-2

[93] LEONENKO, N., MALYARENKO, A. and OLENKO, A. (2022). On spectral theory of random fields in the ball. *Theory Probab. Math. Statist.* **107** 61–76. MR4511144 https://doi.org/10.1090/tpms/1175

[94] LEONENKO, N. N., RUIZ-MEDINA, M. D. and TAQQU, M. S. (2011). Fractional elliptic, hyperbolic and parabolic random fields. *Electron. J. Probab.* **16** 1134–1172. MR2820073 https://doi.org/10.1214/EJP.v16-891

[95] LI, C. (2022). Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model. *Ann. Statist.* **50** 3334–3363. MR4524499 https://doi.org/10.1214/22-aos2230

[96] LI, C., SUN, S. and ZHU, Y. (2023). Fixed-domain posterior contraction rates for spatial Gaussian process model with nugget. *J. Amer. Statist. Assoc.* 1–25.

[97] LI, D., JONES, A., BANERJEE, S. and ENGELHARDT, B. E. (2021). Multi-group Gaussian processes. Preprint. Available at arXiv:2110.08411.

[98] LI, D., TANG, W. and BANERJEE, S. (2023). Inference for Gaussian processes with Matérn covariogram on compact Riemannian manifolds. *J. Mach. Learn. Res.* **24** Paper No. 101, 26 pp. MR4583262

[99] LILLY, J. M., SYKULSKI, A. M., EARLY, J. J. and OLHEDE, S. C. (2017). Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Process. Geophys.* **24** 481–514.

[100] LIM, S. C. and TEO, L. P. (2009). Generalized Whittle–Matérn random field as a model of correlated fluctuations. *J. Phys. A* **42** 105202, 21 pp. MR2485858 https://doi.org/10.1088/1751-8113/42/10/105202

[101] LINDGREN, F., BOLIN, D. and RUE, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spat. Stat.* **50** Paper No. 100599, 29 pp. MR4439328 https://doi.org/10.1016/j.spasta.2022.100599

[102] 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

[103] LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (*Ithaca, NY*, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014 https://doi.org/10.1090/conm/080/999014

[104] LODHIA, A., SHEFFIELD, S., SUN, X. and WATSON, S. S. (2016). Fractional Gaussian fields: A survey. *Probab. Surv.* **13** 1–56. MR3466837 https://doi.org/10.1214/14-PS243

[105] LOH, W.-L. and SUN, S. (2023). Estimating the parameters of some common Gaussian random fields with nugget under fixed-domain asymptotics. *Bernoulli* **29** 2519–2543. MR4580926 https://doi.org/10.3150/22-bej1551

[106] LOH, W.-L., SUN, S. and WEN, J. (2021). On fixed-domain asymptotics, parameter estimation and isotropic Gaussian random fields with Matérn covariance functions. *Ann. Statist.* **49** 3127–3152. MR4352525 https://doi.org/10.1214/21-aos2077

[107] MA, P. and BHADRA, A. (2023). Beyond Matérn: On a class of interpretable confluent hypergeometric covariance functions. *J. Amer. Statist. Assoc.* **118** 2045–2058. MR4646625 https://doi.org/10.1080/01621459.2022.2027775

[108] MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance

in spatial regression. *Biometrika* **71** 135–146. MR0738334 https://doi.org/10.1093/biomet/71.1.135

[109] MATÉRN, B. (1986). *Spatial Variation*: *Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*, 2nd ed. Springer, Heidelberg. MR0169346

[110] MATHERON, G. (1963). *Traité de Géostatistique Appliquée. Tome II*: *Le Krigeage*. *Mémoires du Bureau de Recherches Géologiques et Minières* **24**. Editions BRGM, Paris.

[111] MATHERON, G. (1965). *Les Variables Régionalisées et Leur Estimation*. Masson, Paris.

[112] MATHERON, G. (1971). *The Theory of Regionalized Variables and Its Applications*. Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, France.

[113] MENEGATTO, V. A., OLIVEIRA, C. P. and PORCU, E. (2020). Gneiting class, semi-metric spaces and isometric embeddings. *Constr. Math. Anal.* **3** 85–95. MR4188612 https://doi.org/10.33205/cma.712049

[114] MORALES-NAVARRETE, D., BEVILACQUA, M., CAAMAÑO-CARRILLO, C. and CASTRO, L. M. (2024). Modeling point referenced spatial count data: A Poisson process approach. *J. Amer. Statist. Assoc.* **119** 664–677. MR4713921 https://doi.org/10.1080/01621459.2022.2140053

[115] NARCOWICH, F. J., WARD, J. D. and WENDLAND, H. (2006). Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr. Approx.* **24** 175–186. MR2239119 https://doi.org/10.1007/s00365-005-0624-7

[116] NIKITIN, A. V., JOHN, S., SOLIN, A. and KASKI, S. (2022). Non-separable spatio-temporal graph kernels via SPDEs. In *International Conference on Artificial Intelligence and Statistics* 10640–10660. PMLR.

[117] NOVAK, E. and WOŹNIAKOWSKI, H. (2010). *Tractability of Multivariate Problems*. *Volume II*: *Standard Information for Functionals*. *EMS Tracts in Mathematics* **12**. European Mathematical Society (EMS), Zürich. MR2676032 https://doi.org/10.4171/084

[118] OATES, C. J., GIROLAMI, M. and CHOPIN, N. (2017). Control functionals for Monte Carlo integration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 695–718. MR3641403 https://doi.org/10.1111/rssb.12185

[119] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W., eds. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge. MR2723248

[120] PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17** 483–506. MR2240939 https://doi.org/10.1002/env.785

[121] PALACIOS, M. B. and STEEL, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *J. Amer. Statist. Assoc.* **101** 604–618. MR2281244 https://doi.org/10.1198/016214505000001195

[122] PEREIRA, M., DESASSIS, N. and ALLARD, D. (2022). Geostatistics for large datasets on Riemannian manifolds: A matrix-free approach. Preprint. Available at arXiv:2208.12501.

[123] PORCU, E., ALEGRIA, A. and FURRER, R. (2018). Modelling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.* **86** 344–377. MR3852415 https://doi.org/10.1111/insr.12266

[124] PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2020). Nonseparable, space-time covariance functions with dynamical compact supports. *Statist. Sinica* **30** 719–739. MR4214159

[125] PORCU, E., BEVILACQUA, M., SCHABACK, R. and OATES, C. J. (2024). Supplement to "The Matérn model: A journey through statistics, numerical analysis and machine learning." https://doi.org/10.1214/24-STS923SUPP

[126] PORCU, E., DALEY, D. J., BUHMANN, M. and BEVILACQUA, M. (2013). Radial basis functions with compact support for multivariate geostatistics. *Stoch. Environ. Res. Risk Assess.* **27** 909–922.

[127] PORCU, E., EMERY, X. and MERY, N. (2022). Criteria and characterizations for spatially isotropic and temporally symmetric matrix-valued covariance functions. *Comput. Appl. Math.* **41** Paper No. 223, 14 pp. MR4446820 https://doi.org/10.1007/s40314-022-01930-2

[128] PORCU, E., FURRER, R. and NYCHKA, D. (2021). 30 years of space-time covariance functions. *Wiley Interdiscip. Rev.*: *Comput. Stat.* **13** Paper No. e1512, 24 pp. MR4218945 https://doi.org/10.1002/wics.1512

[129] PORCU, E., WHITE, P. A. and GENTON, M. G. (2022). Nonseparable space-time stationary covariance functions on networks cross time. Preprint. Available at arXiv:2208.03359.

[130] PORCU, E., ZASTAVNYI, V., BEVILACQUA, M. and EMERY, X. (2020). Stein hypothesis and screening effect for covariances with compact support. *Electron. J. Stat.* **14** 2510–2528. MR4119265 https://doi.org/10.1214/20-EJS1719

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

[132] QUICK, H., BANERJEE, S. and CARLIN, B. P. (2013). Modeling temporal gradients in regionally aggregated California asthma hospitalization data. *Ann. Appl. Stat.* **7** 154–176. MR3086414 https://doi.org/10.1214/12-AOAS600

[133] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[134] ROININEN, L., GIROLAMI, M., LASANEN, S. and MARKKANEN, M. (2019). Hyperpriors for Matérn fields with applications in Bayesian inversion. *Inverse Probl. Imaging* **13** 1–29. MR3917848 https://doi.org/10.3934/ipi.2019001

[135] SANZ-ALONSO, D. and YANG, R. (2022). Finite element representations of Gaussian processes: Balancing numerical and statistical accuracy. *SIAM/ASA J. Uncertain. Quantificat.* **10** 1323–1349. MR4499223 https://doi.org/10.1137/21M144788X

[136] SANZ-ALONSO, D. and YANG, R. (2022). The SPDE approach to Matérn fields: Graph representations. *Statist. Sci.* **37** 519–540. MR4497230 https://doi.org/10.1214/21-sts838

[137] SÄRKKÄ, S., SOLIN, A. and HARTIKAINEN, J. (2013). Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering. *IEEE Signal Process. Mag.* **30** 51–61.

[138] SCHABACK, R. (2011). The missing Wendland functions. *Adv. Comput. Math.* **34** 67–81. MR2783302 https://doi.org/10.1007/s10444-009-9142-7

[139] SCHABACK, R. (2015). A computational tool for comparing all linear PDE solvers: Error-optimal methods are meshless. *Adv. Comput. Math.* **41** 333–355. MR3337495 https://doi.org/10.1007/s10444-014-9360-5

[140] SCHABACK, R. (2017). Error analysis of nodal meshless methods. In *Meshfree Methods for Partial Differential Equations VIII*. *Lecture Notes in Computational Science and Engineering* **115** 117–143. Springer, Cham. MR3718331 https://doi.org/10.1007/978-3-319-51954-8_7

[141] SCHABACK, R. and WENDLAND, H. (2006). Kernel techniques: From machine learning to meshless methods. *Acta Numer.* **15** 543–639. MR2269747 https://doi.org/10.1017/S0962492906270016

[142] SCHEUERER, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Process. Appl.* **120** 1879–1897. MR2673978 https://doi.org/10.1016/j.spa.2010.05.009

[143] SCHEUERER, M., SCHABACK, R. and SCHLATHER, M. (2013). Interpolation of spatial data—A stochastic or a deterministic problem? *European J. Appl. Math.* **24** 601–629. MR3082868 https://doi.org/10.1017/S0956792513000016

[144] SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.* (2) **39** 811–841. MR1503439 https://doi.org/10.2307/1968466

[145] SCHOENBERG, I. J. (1942). Positive definite functions on spheres. *Duke Math. J.* **9** 96–108. MR0005922

[146] SHABY, B. and RUPPERT, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *J. Comput. Graph. Statist.* **21** 433–452. MR2945475 https://doi.org/10.1080/10618600.2012.680819

[147] SOUTH, L. F., KARVONEN, T., NEMETH, C., GIROLAMI, M. and OATES, C. J. (2022). Semi-exact control functionals from Sard's method. *Biometrika* **109** 351–367. MR4430962 https://doi.org/10.1093/biomet/asab036

[148] STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856 https://doi.org/10.1214/aos/1176350690

[149] STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

[150] STEIN, M. L. (2002). The screening effect in kriging. *Ann. Statist.* **30** 298–323. MR1892665 https://doi.org/10.1214/aos/1015362194

[151] STEIN, M. L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *J. Statist. Plann. Inference* **123** 1–11. MR2058118 https://doi.org/10.1016/S0378-3758(03)00144-7

[152] STEIN, M. L. (2005). Statistical methods for regular monitoring data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 667–687. MR2210686 https://doi.org/10.1111/j.1467-9868.2005.00520.x

[153] STEIN, M. L. (2011). 2010 Rietz Lecture: When does the screening effect hold? *Ann. Statist.* **39** 2795–2819. MR3012392 https://doi.org/10.1214/11-AOS909

[154] STEIN, M. L. (2015). When does the screening effect not hold? *Spat. Stat.* **11** 65–80. MR3311857 https://doi.org/10.1016/j.spasta.2014.12.003

[155] STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376 https://doi.org/10.1046/j.1369-7412.2003.05512.x

[156] SUN, Y., LI, B. and GENTON, M. G. (2012). Geostatistics for large datasets. In *Advances and Challenges in Space-Time Modelling of Natural Events* 55–77. Springer, Berlin.

[157] TANG, J. and ZIMMERMAN, D. (2024). Space-time covariance models on networks. *Electron. J. Stat.* **18** 490–514. MR4705177 https://doi.org/10.1214/23-ejs2206

[158] TANG, W., ZHANG, L. and BANERJEE, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 1044–1070. MR4349127 https://doi.org/10.1111/rssb.12472

[159] TERDIK, G. (2015). Angular spectra for non-Gaussian isotropic fields. *Braz. J. Probab. Stat.* **29** 833–865. MR3397396 https://doi.org/10.1214/14-BJPS249

[160] TERDIK, G. H. (2022). Spatiotemporal covariance functions for Laplacian ARMA fields in higher dimensions. *Theory Probab. Math. Statist.* **107** 111–132. MR4511146 https://doi.org/10.1090/tpms/1173

[161] TUO, R. and WANG, W. (2020). Kriging prediction with isotropic Matérn correlations: Robustness and experimental designs. *J. Mach. Learn. Res.* **21** Paper No. 187, 38 pp. MR4209473

[162] TUO, R., WANG, Y. and WU, C. F. J. (2020). On the improved rates of convergence for Matérn-type kernel ridge regression with application to calibration of computer models. *SIAM/ASA J. Uncertain. Quantificat.* **8** 1522–1547. MR4184289 https://doi.org/10.1137/19M1304222

[163] VAKILI, S., MOSS, H., ARTEMEV, A., DUTORDOIR, V. and PICHENY, V. (2021). Scalable Thompson sampling using sparse Gaussian process models. *Adv. Neural Inf. Process. Syst.* **34** 5631–5643.

[164] VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852

[165] VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. MR0964183

[166] WALLIN, J. and BOLIN, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scand. J. Stat.* **42** 872–890. MR3391697 https://doi.org/10.1111/sjos.12141

[167] WENDLAND, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4** 389–396. MR1366510 https://doi.org/10.1007/BF02123482

[168] WENDLAND, H. (2005). *Scattered Data Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724

[169] WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. Roy. Statist. Soc. Ser. B* **15** 125–139. MR0056902

[170] WHITTLE, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Int. Stat.* **40** 974–994. MR0173287

[171] WILSON, A. G., HU, Z., SALAKHUTDINOV, R. and XING, E. P. (2016). Deep kernel learning. In *Artificial Intelligence and Statistics* 370–378. PMLR.

[172] WILSON, J. T., BOROVITSKIY, V., TERENIN, A., MOSTOWSKY, P. and DEISENROTH, M. P. (2021). Pathwise conditioning of Gaussian processes. *J. Mach. Learn. Res.* **22** Paper No. 105, 47 pp. MR4279756

[173] WU, Z. M. (1992). Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approx. Theory Appl.* **8** 1–10. MR1192639

[174] XU, G. and GENTON, M. G. (2017). Tukey g-and-h random fields. *J. Amer. Statist. Assoc.* **112** 1236–1249. MR3735373 https://doi.org/10.1080/01621459.2016.1205501

[175] YAN, Y. and GENTON, M. G. (2018). Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics* **29** e2458, 15 pp. MR3830874 https://doi.org/10.1002/env.2458

[176] ZASTAVNYI, V. P. (2002). Positive-definite radial functions and splines. *Dokl. Akad. Nauk* **386** 446–449. MR2006036

[177] ZASTAVNYI, V. P. (2006). On some properties of the Buhmann functions. *Ukrainian Math. J.* **58** 1045–1067. MR2345078 https://doi.org/10.1007/s11253-006-0128-z

[178] ZASTAVNYI, V. P. and PORCU, E. (2011). Characterization theorems for the Gneiting class of space-time covariances. *Bernoulli* **17** 456–465. MR2797999 https://doi.org/10.3150/10-BEJ278

[179] ZASTAVNYI, V. P. and PORCU, E. (2017). On positive definiteness of some radial functions. *Lobachevskii J. Math.* **38** 386–394. MR3636153 https://doi.org/10.1134/S1995080217020226

[180] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

[181] ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195 https://doi.org/10.1093/biomet/92.4.921

Statistical Science
2024, Vol. 39, No. 3, 493–507
https://doi.org/10.1214/24-STS926

Antoine Gombaud, Chevalier de Méré

David R. Bellhouse and Christian Genest

Abstract. In 1654, Antoine Gombaud, Chevalier de Méré, approached Blaise Pascal with a question about the throw of dice and drew his attention to the problem of points, which had been around for 250 years or more in the Italian "abbaco" literature. A correspondence ensued between Pascal and Pierre de Fermat which is widely regarded as marking the birth of probability calculus. While historians of probability have rightfully focused on Pascal and Fermat, they have generally ignored Gombaud's part, portraying him only as a gambler, sometimes an avid one. Through a careful examination of Gombaud's life and philosophy, it is argued here that his role was more important than has usually been attributed to him. In addition, a review of the historical background to the problem of points shows that gambling was not as central to the early development of probability theory as has often been assumed.

Key words and phrases: Antoine Gombaud, Blaise Pascal, Chevalier de Méré, Pierre de Fermat, problem of points.

REFERENCES

[1] ARBUTHNOT, J. *Of the Laws of Chance*, *or, a Method of Calculating the Hazards of Game*. Benjamin Motte, London.

[2] BELLHOUSE, D. R. and FILLION, N. (2015). Le Her and other problems in probability discussed by Bernoulli, Montmort and Waldegrave. *Statist. Sci.* **30** 26–39. MR3317752 https://doi.org/10.1214/14-STS469

[3] BERNOULLI, J. (1690). Quæstiones nonnullæ de usuris, cum solutione problematis de sorte alearum, propositi in Ephem. Gall. A. 1685, Art. 25. *Acta Erud.* **9** 219–223.

[4] BERNOULLI, J. (1713). *Ars conjectandi*. Thursinius, Basel, Switzerland.

[5] BERNOULLI, J. (2006). *The Art of Conjecturing*: *Together with "Letter to a Friend on Sets in Court Tennis."* Johns Hopkins Univ. Press, Baltimore, MD. Translated from the Latin and with an introduction and notes by Edith Dudley Sylla. MR2195221

[6] BERNOULLI, N. (1709). *De usu artis conjectandi in jure*. Johannis Conradi, Basel, Switzerland.

[7] BOYER, A. (1702). *Dictionnaire royal*, *françois et anglois*. *Le françois tiré des dictionnaires de Richelet*, *Furetiere*, *Tachard*, *de l'Academie françoise*, *& des remarques de Vaugelas*, *Menage & Bouhours*. *Divisé en deux parties*. Adrian Moetjens, The Hague, The Netherlands.

[8] CARDANO, G. (1539). *Practica arithmetica et mensurandi singularis*. *Milan*. *Reprinted* 1966 *in Hieronymus Cardanus Omnia Opera*: *Faksimile-Neudruck der Ausgabe Lyon* 1663 *mit einer Einleitung von August Buck* **4**. Friedrich Frommann Verlag, Stuttgart-Bad Cannstatt.

[9] CASTIGLIONE, B. (1967). *The Book of the Courtier* (G. Bull, ed.). Penguin, London.

[10] CHAMAILLARD, E. (1921). *Le Chevalier de Méré*, *rival de Voiture*, *ami de Pascal*, *précepteur de Mme de Maintenon*. G. Clouzot, Niort, France.

[11] CONLEY, J. (2019). Madeleine de Scudéry. In *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.).

[12] DARLING, D. (2004). *The Universal Book of Mathematics*: *From Abracadabra to Zeno's Paradoxes*. Wiley, Hoboken, NJ. MR2078978

[13] DAVID, F. N. (1962). *Gods*, *Games and Gambling*: *The Origins and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian Era*. Hafner, New York. MR1638639

[14] DE COSTE, H. (1649). *La vie du R. P. Marin Mersenne*: *Theologien*, *philosophe et mathematicien de l'Ordre des Peres Minimes*. Sebastien & Gabriel Cramoisy, Paris.

[15] DE FERMAT, P. (1679). *Varia opera mathematica*. Joannis Pech, Toulouse.

[16] DE FERMAT, P. (1894). *Œuvres de Fermat* 2. (P. Tannery and C. Henry, ed.). Gauthier-Villars, Paris.

[17] DE MONTMORT, P. R. (1980). *Essay d'analyse sur les jeux de hazard*, Jacques Quillau, Paris. MR0605303

[18] DUCLAUX, M. (1927). *Portrait of Pascal*. Harper & Brothers, New York.

[19] EDWARDS, A. W. F. (1982). Pascal and the problem of points. *Int. Stat. Rev.* **50** 259–266. MR0693220 https://doi.org/10.2307/1402496

[20] EPSTEIN, S. A. (1996). *Genoa and the Genoese* 958–1528. Univ. North Carolina Press Chapel Hill, NC.

[21] FISHER, S. (2014). Pierre Gassendi. In *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.).

[22] FORESTANI, L. (1603). *Pratica d'arithmetica*, *e geometria*. Giorgio, Varisco, Venezia.

[23] FRANCE, P. (2005). Honnêteté. In *The New Oxford Companion to Literature in French*. Clarendon, Oxford.

David R. Bellhouse is Professor Emeritus, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada (e-mail: drbell@uwo.ca). Christian Genest is Professor and Canada Research Chair, Department of Mathematics and Statistics, McGill University, Montréal (Québec) Canada (e-mail: christian.genest@mcgill.ca).

[24] FRANCI, R. (2002). Una soluzione esatta del problema delle parti in un manoscritto della prima metà del Quattrocento. *Boll. Stor. Sci. Mat.* **22** 253–265.

[25] FRANKLIN, J. (2001). *The Science of Conjecture*: *Evidence and Probability Before Pascal*. Johns Hopkins Univ. Press, Baltimore, MD. MR1893301

[26] GASSENDI, P. (1976). The reality of infinite void according to Aristotle. In *The Concepts of Space and Time*: *Their Structure and Their Development* (M. Čapek, ed.). 91–95. Springer, Dordrecht. MR0462420

[27] GOMBAUD, A. (1674). *Le jeu de l'hombre*. Claude Barbin, Paris.

[28] GOMBAUD, A. (1682). *Lettres de Monsieur le Chevalier de Méré. Premiere partie*. Denis Thierry & Claude Barbin, Paris.

[29] GOMBAUD, A. (1700). *Œuvres posthumes de M. le Chevalier de Méré*. Jean & Michel Guignard, Paris.

[30] GOMBAUD, A. (1930). *Œuvres complètes du Chevalier de Méré*: *Texte établi et présenté par Charles-H. Boudhors. Tomes* 1–3 (F. Roches, ed.). Les Belles Lettres, Paris.

[31] GOURAUD, C. (1848). *Histoire du calcul des probabilités depuis ses origines jusqu'à nos jours*. Auguste Durand, Paris.

[32] GRUBBS, H. A. (1932). *Damien Mitton* (1618–1690): *Bourgeois honnête homme*. Princeton Univ. Press, Princeton, NJ. Reprinted 1965 by Kraus Reprint Corporation, New York.

[33] HALD, A. (1990). *A History of Probability and Statistics and Their Applications Before* 1750. Wiley, New York. MR1029276 https://doi.org/10.1002/0471725161

[34] HUYGENS, C. (1657). De ratiociniis in ludo aleæ. In *Exercitationum mathematicarum in libri quinque* (F. van Schooten, ed.) 517–534. Elsevier, Leiden, The Netherlands.

[35] KAVANAGH, T. M. (2005). *Dice*, *Cards*, *Wheels*: *A Different History of French Culture*. Univ. Pennsylvania Press, Philadelphia, PA.

[36] KENDALL, M. G. (1956). Studies in the history of probability and statistics. II. The beginnings of a probability calculus. *Biometrika* **43** 1–14 (2 plates). MR0076678 https://doi.org/10.2307/2333573

[37] LIEBER, A. E. (1968). Eastern business practices and medieval European commerce. *Econ. Hist. Rev.* **21** 230–243.

[38] LUBBOCK, J. W. and DRINKWATER BETHUNE, J. E. (1830). *On Probability*. Library of Useful Knowledge, London.

[39] MACK, R. E. (2002). *Bazaar to Piazza*: *Islamic Trade and Italian Art*, 1300–1600. Univ. California Press, Berkeley, CA.

[40] MALET, A. (1996). From indivisibles to infinitesimals: Studies on seventeenth century mathematizations of infinitely small quantities. Departament de Filosofia (BeIlaterra), Universitat Autónoma de Barcelona, Barcelona. MR1658963

[41] MARQUIS DE DANGEAU, P. C. (1854). Journal du Marquis de Dangeau, 1. Fermin Didot Frères, Paris.

[42] MARQUISE DE SÉVIGNÉ, M. R.-C. (1974). *Correspondance de Mme de Sévigné* (R. Duchêne, ed.). **2**. Gallimard, Paris.

[43] MESNARD, J. (1965). *Pascal et les Roannez*. Desclée de Brouwer, Paris.

[44] MEUSNIER, N. (2007). Le problème des partis peut-il être d'origine arabo-musulmane? *J. Électron. Hist. Probab. Stat.* **3** 14. MR2393128

[45] MICHAUT, G. (1902). *Les époques de la pensée de Pascal*, 2nd ed. Albert Fontemoing, Paris.

[46] MITTON, D. (2013). *Pensieri sull'onestà decorosa e altri scritti* (M. Lanterna, ed.). Clinamen, Firenze.

[47] MONTUCLA, J.-É. (1960). *Histoire des mathématiques*. Librairie scientifique et technique Albert Blanchard, Paris. 4 tomes, nouveau tirage augmenté d'un avant-propos par M. Ch. Naux. MR0124168

[48] MOTLEY, M. E. (1987). Aristocratic education in seventeenth century France, 1580–1715. Doctoral dissertation, Dept. of History, Princeton, NJ.

[49] PACIOLI, L. (1494). *Su*[*m*]*ma de arithmetica*, *geometria*, *proportioni et proportionalita*. Paganinus de Paganinis, Venezia.

[50] PETERSON, M. A. (1997). The geometry of Piero della Francesca. *Math. Intelligencer* **19** 33–40. MR1475147 https://doi.org/10.1007/BF03025346

[51] PEVERONE, G. F. (1558). *Due brevi e facili trattati*, *il primo d'arithmetica*, *l'altro di geometria*: *Ne i quali si contengono alcune cose nuove piaceuoli è utili*, *si à gentilhuomini come artegiani*. Per Gio di Tournes, Lyon.

[52] PISANO, L. (2002). *Fibonacci's liber abaci. Sources and Studies in the History of Mathematics and Physical Sciences*. Springer, New York. A translation into modern English of Leonardo Pisano's book of calculation, translated from the Latin and with an introduction, notes and bibliography by L. E. Sigler. MR1923794

[53] REVILLOUT, C. J. (1887). *Antoine Gombaud*, *Chevalier de Méré*, *sa famille*, *son frère et ses amis illustres*. Charles Boehm, Montpellier.

[54] RIGATELLI, L. T. (1985). Il problema dell parti in manoscritti del XIV e XV secolo. In *Mathemata*: *Festscrift für Helmuth Gericke* (M. Folkerts and U. Lindgren, eds.). Steiner, Wisebaden, Germany. MR0799739

[55] SANCHEZ, C. (2014). Les livres de jeux aux XVIIe et XVIIIe siècles: Une typologie des lecteurs-joueurs. Mémoire de master, Université Lumière Lyon 2, Lyon, France.

[56] SCHNEIDER, I. (1988). The marketplace and games of chance in the fifteenth and sixteenth centuries. In *Mathematics from Manuscript to Print*, 1300–1600 (*Oxford*, 1984) (C. Hay, ed.). *Oxford Sci. Publ.* 220–235. Oxford Univ. Press, New York. MR0961299

[57] SEIFERT, L. C. (2009). *Manning the Margins*: *Masculinity and Writing in Seventeenth-Century France*. Univ. Michigan Press, Ann Arbor, MI.

[58] SMITH, D. E. (1970). *Rara Arithmetica. A Catalogue of the Arithmetics Written Before the Year MDCI with a Description of Those in the Library of George Arthur Plimpton of New York*. Chelsea Publishing Co., New York. Fourth edition, including A. De Morgan's Arithmetical books. MR0284307

[59] SPIEGELHALTER, D. (2019). *The Art of Statistics*: *Learning from Data*. Basic Books, New York. MR3967688

[60] STEINMETZ, A. (1870). The gaming table: Its votaries and victims. In *All Times and Countries*, *Especially in England and in France* **1**. Tinsley Brothers, London.

[61] TAILLÉ, J.-P. and FOISSEAU, R. (2012). *Le Chevalier de Méré*: *Un gentilhomme poitevin*, *cultivé*, *précieux et parfois arrogant*. Mon Petit Éditeur, Paris.

[62] TARTAGLIA, N. (1556). *General trattato di numeri e misure*. Curzio Troiano dei Nauo, Venezia.

[63] TURNER, A. M. (1966). In search of the honnête femme: The character of Pauline in Polyeucte. *Romance Notes* **7** 165–170.

[64] VAN EGMOND, W. (1981). *Practical Mathematics in the Italian Renaissance*: *A Catalog of Italian Abbacus Manuscripts and Printed Books to* 1600. Editrice Giunti Barbèra, Firenze.

[65] WETSEL, D. (1993). Pascal and Mitton: Theological objections to 'l'honnêteté' in the Pensées. *Fr. Stud.* **47** 404–410.

Statistical Science
2024, Vol. 39, No. 3, 508–517
https://doi.org/10.1214/24-STS931

A Conversation with Raymond J. Carroll

Xihong Lin and Nilanjan Chatterjee

Abstract. Raymond J. Carroll is Distinguished Professor of Statistics, Nutrition and Toxicology at Texas A& M University, USA. He has made fundamental contributions to numerous statistical and health science areas, including measurement error models, data transformation and weighting in regression, nonparametric and semiparametric regression, longitudinal data analysis, and statistical methods and applications in nutrition, epidemiology and molecular biology. Carroll has received many distinguished honors. Some highlights are the 1988 Presidents' Award of the Committee of Presidents' of Statistical Societies (COPSS), the 2002 COPSS Fisher Lectureship Award, American Statistial Association and Institute of Mathematical Statistics fellows. He was the first statistician given a Method to Extend Research in Time (MERIT) Award from the US National Cancer Institute. Carroll has provided outstanding professional services, including editor of Biometrics, the Journal of the American Statistical Association (Theory and Methods), and founding chair of the Biostatistical Methods and Research Design study section of the US National Institutes of Health. Carroll is an inspirational and successful teacher and mentor. He has won a College of Science Teaching Award from Texas A&M University, and has trained over 45 Ph.D. students. He has also been an outstanding mentor and supporter to many junior researchers in the statistical community, including the two authors of this article. In this interview, Carroll talks about his career, including his passion for mentoring junior researchers, and offers some helpful advice.

Key words and phrases: Epidemiology, longitudinal data analysis, measurement error, nonparametric and semiparametric regression, nutrition, transformation.

REFERENCES

CARROLL, R. J. (2003). Variances are not always nuisance parameters. *Biometrics* **59** 211–220. MR1987387 https://doi.org/10.1111/1541-0420.t01-1-00027

CARROLL, R. J. (2014). Personal reflections on the COPSS presidents' award. In *Past*, *Present*, *and Future of Statistical Science*. 595–604. CRC Press, Boca Raton, FL.

WU, M. C. and CARROLL, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44** 175–188. MR0931633 https://doi.org/10.2307/2531905

Xihong Lin is Professor of Biostatistics at Harvard T.H. Chan School of Public Health and Professor of Statistics at Harvard University, Boston, Massachusetts 02115, USA (e-mail: xlin@hsph.harvard.edu). Nilanjan Chatterjee is Bloomberg Distinguished Professor of Biostatistics and Genetic Epidemiology at Johns Hopkins School of Public Health, Baltimore, Maryland 21205, USA (e-mail: nchatte2@jhu.edu).

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

President: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

President-Elect: Tony Cai, Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104-6304, USA

Past President: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

Executive Secretary: Peter Hoff, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Enno Mammen, Institute for Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

The Annals of Applied Statistics. *Editor-in-Chief*: Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. *Editors*: Paul Bourgade, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012-1185, USA. Julien Dubedat, Department of Mathematics, Columbia University, New York, NY 10027, USA

The Annals of Applied Probability. *Editors*: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. *Editor*: Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The IMS Bulletin. *Editor*: Tati Howell, bulletin@imstat.org