Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics Volume 39 • Number 3 • September 2025

Brazilian Journal of Probability and Statistics

Volume 39 • Number 3 • September 2025

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 39, Number 3, September 2025. Published quarterly by the Brazilian Statistical Association.

POSTMASTER:

Send address changes to Brazilian Journal of Probability and Statistics, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Brazilian Statistical Association members should send address changes to Rua do Matão, 1010 sala 250A, 05508-090 São Paulo/SP Brazil (address of the BSA office).

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2025, Vol. 39, No. 3, 287–320
https://doi.org/10.1214/25-BJPS636

Least squares estimation for path-distribution dependent stochastic differential equations driven by fractional Brownian motions associated with the interacting particle systems

Huan Zhou[1,2,a] **and Guangjun Shen**[2,b]

[1]*College of Mathematics and Computer Science, Tongling University, Tongling, 244061, China,* [a]*zhouhuan_1997@163.com*
[2]*Department of Mathematics, Anhui Normal University, Wuhu, 241002, China,* [b]*gjshen@ahnu.edu.cn*

Abstract. In this paper, we focus on least squares estimator for an unknown parameter in the drift coefficient of path-distribution dependent stochastic differential equation driven by fractional Brownian motions with Hurst parameter $H \in (1/2, 1)$. Based on the time-discretized interacting particle systems of the stochastic differential equation combined with the contrast function, we discuss the consistency and asymptotic distribution of the estimator with non-Lipschitz conditions.

References

Bao, J. and Huang, X. (2021). Approximations of McKean-Vlasov stochastic differential equations with irregular coefficients. *Journal of Theoretical Probability* **35**, 1187–1215. MR4414415 https://doi.org/10.1007/s10959-021-01082-9

Bao, J., Ren, P. and Wang, F. (2021). Bismut formula for Lions derivative of distribution-path dependent SDEs. *Journal of Differential Equations* **282**, 285–329. MR4219322 https://doi.org/10.1016/j.jde.2021.02.019

Biagini, F., Hu, Y., Øksendal, B. and Zhang, T. (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. London: Springer. MR2387368 https://doi.org/10.1007/978-1-84628-797-8

Bossy, M. and Talay, D. (1997). A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation* **66**, 157–192. MR1370849 https://doi.org/10.1090/S0025-5718-97-00776-X

Buckdahn, R., Li, J., Peng, S. and Rainer, C. (2017). Mean-field stochastic differential equations and associated PDEs. *Annals of Probability* **2**, 824–878. MR3630288 https://doi.org/10.1214/15-AOP1076

Carmona, R. and Delarue, F. (2013). Probabilistic analysis of mean-field games. *SIAM Journal on Control and Optimization* **51**, 2705–2734. MR3072222 https://doi.org/10.1137/120883499

Chaudru, P. and Raynal, D. (2020). Strong well-posedness of McKean-Vlasov stochastic differential equations with Hölder drift. *Stochastic Processes and Their Applications* **130**, 79–107. MR4035024 https://doi.org/10.1016/j.spa.2019.01.006

Decreusefond, L. and Üstünel, A. (1998). Stochastic analysis of the fractional Brownian motion. *Potential Analysis* **10**, 177–214. MR1677455 https://doi.org/10.1023/A:1008634027843

Fan, X., Huang, X., Suo, Y. and Yuan, C. (2022). Distribution dependent SDEs driven by fractional Brownian motions. *Stochastic Processes and Their Applications* **151**, 23–67. MR4441502 https://doi.org/10.1016/j.spa.2022.05.007

Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **3**, 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

Garsia, A., Rodemich, E., Rumsey, H. and Rosenblatt, M. (1978). A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana University Mathematics Journal* **20**, 565–578. MR0267632 https://doi.org/10.1512/iumj.1970.20.20046

Hammersley, W., Šiška, D. and Szpruch, L. (2021). McKean-Vlasov SDEs under measure dependent Lyapunov conditions. *Annales de L'IHP Probabilités et Statistiques* **57**, 1032–1057. MR4260494 https://doi.org/10.1214/20-aihp1106

Hu, Y., Xi, F. and Zhu, M. (2022). Least squares estimation for distribution-dependent stochastic differential delay equations. *Communications on Pure and Applied Analysis* **21**, 1505–1536. MR4405320 https://doi.org/10.3934/cpaa.2022027

Key words and phrases. Path-distribution dependent stochastic differential equation, fractional Brownian motion, least squares estimator, consistency, asymptotic distribution.

Hu, Y., Xi, F. and Zhu, M. (2024). Asymptotic properties for the parameter estimation in stochastic (functional) differential equations with Hölder drift. *Stochastic* **96**, 766–798. MR4716851 https://doi.org/10.1080/17442508.2024.2320397

Huang, X., Röckner, M. and Wang, F. (2019). Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs. *Discrete and Continuous Dynamical Systems* **39**, 3017–3035. MR3959419 https://doi.org/10.3934/dcds.2019125

Huang, X. and Wang, F. (2019). Distribution dependent SDEs with singular coefficients. *Stochastic Processes and Their Applications* **129**, 4747–4770. MR4013879 https://doi.org/10.1016/j.spa.2018.12.012

Liu, M. and Qiao, H. (2022). Parameter estimation of path-dependent McKean-Vlasov stochastic differential equations. *Acta Mathematica Scientia* **42**, 876–886. MR4411008 https://doi.org/10.1007/s10473-022-0304-8

Long, H., Ma, C. and Yasutaka, S. (2017). Least squares estimators for stochastic differential equations driven by small Lévy noises. *Stochastic Processes and Their Applications* **127**, 1475–1495. MR3630232 https://doi.org/10.1016/j.spa.2016.08.006

Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian motion, fractional noises and applications. *SIAM Review* **10**, 422–437. MR0242239 https://doi.org/10.1137/1010093

Mehri, S. and Stannat, W. (2019). Weak solutions to Vlasov-McKean equations under Lyapunov-type conditions. *Stochastics and Dynamics* **19**, 1950042. MR4033001 https://doi.org/10.1142/S0219493719500424

Mishura, Y. and Veretennikov, A. (2021). Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations. *Theory of Probability and Mathematical Statistics* **103**, 59–101. MR4421344 https://doi.org/10.1090/tpms/1135

Ren, J. and Zhang, X. (2008). Freidlin-Wentzell's large deviations for stochastic evolution equations. *Journal of Functional Analysis* **254**, 3148–3172. MR2418622 https://doi.org/10.1016/j.jfa.2008.02.010

Ren, P. and Wang, F. (2021). Donsker-Varadhan large deviations for path-distribution dependent SPDEs. *Journal of Mathematical Analysis and Applications* **499**, 125000. MR4207324 https://doi.org/10.1016/j.jmaa.2021.125000

Ren, P. and Wu, J. (2021). Least squares estimation for path-distribution dependent stochastic differential equations. *Applied Mathematics and Computation* **410**, 126457. MR4278032 https://doi.org/10.1016/j.amc.2021.126457

Röckner, M. and Zhang, X. (2021). Well-posedness of distribution dependent SDEs with singular drifts. *Bernoulli* **27**, 1131–1158. MR4255229 https://doi.org/10.3150/20-bej1268

Shen, G., Xiang, J. and Wu, J. (2022). Averaging principle for distribution dependent stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *Journal of Differential Equations* **321**, 381–414. MR4396023 https://doi.org/10.1016/j.jde.2022.03.015

Sobczyk, K. (1990). *Stochastic Differential Equations: With Applications to Physics and. Engineering. Math. Appl. (East European Series)*. Dordrecht: Kluwer Academic Publishers. MR1135326 https://doi.org/10.1007/978-94-011-3712-6

Van der Vaart, A. (2000). *Asymptotic Statistics (Vol. 3)*. Cambridge: Cambridge University Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Wang, F. (2018). Distribution-dependent SDEs for Landau type equations. *Stochastic Processes and Their Applications* **128**, 595–621. MR3739509 https://doi.org/10.1016/j.spa.2017.05.006

Young, L. (1936). An inequality of Hölder type, connected with Stieltjes integration. *Acta Mathematica* **67**, 251–282. MR1555421 https://doi.org/10.1007/BF02401743

Zhang, X. and Shu, H. (2022). Trajectory fitting estimation for a class of SDEs with small Lévy noises. *Brazilian Journal of Probability and Statistics* **36**, 579–592. MR4489182 https://doi.org/10.1214/22-bjps541

Zhu, M. and Hu, Y. (2022). Least squares estimation for delay McKean-Vlasov stochastic differential equations and interacting particle systems. *Communications in Mathematical Sciences* **20**, 265–296. MR4359256 https://doi.org/10.4310/CMS.2022.v20.n1.a8

Brazilian Journal of Probability and Statistics
2025, Vol. 39, No. 3, 321–347
https://doi.org/10.1214/25-BJPS637

A geometric framework for multivariate jump locations estimation

Hugo Henneuse[a]

Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, Orsay, France,
[a]*hugo.henneuse@universite-paris-saclay.fr*

Abstract. Our study addresses the estimation of the locations of discontinuities (or jumps) within multivariate signals from noisy observations in the nonparametric regression setting. Departing from standard analytical approaches, we propose a new framework, based on geometric control over the jump locations. This allows us to consider larger classes of signals, of any dimension, with potentially wild discontinuity set (exhibiting, typically, self-intersections and corners). We study a simple estimation procedure relying on histogram differences and show its consistency and near-optimality for the Hausdorff distance over these new classes. Furthermore, exploiting this new geometric framework, we design procedures to estimate consistently several topological descriptors of the jump locations.

References

Balakrishnan, S., Rinaldo, A., Sheehy, D., Singh, A. and Wasserman, L. (2012). Minimax rates for homology inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. *Proceedings of Machine Learning Research* **22**, 64–72.

Balakrishnan, S., Rinaldo, A., Singh, A. and Wasserman, L. A. (2013). Tight Lower Bounds for Homology Inference.

Bengs, V., Eulert, M. and Holzmann, H. (2019). Asymptotic confidence sets for the jump curve in bivariate regression problems. *Journal of Multivariate Analysis* **173**, 291–312. MR3926257 https://doi.org/10.1016/j.jmva.2019.02.017

Bobrowski, O., Mukherjee, S. and Taylor, J. E. (2017). Topological consistency via kernel estimation. *Bernoulli* **23**, 288–328. MR3556774 https://doi.org/10.3150/15-BEJ744

Bradski, G. (2000). The OpenCV library. *Dr Dobb's Journal*.

Bubenik, P., Carlsson, G., Kim, P. and Luo, Z. (2009). *Statistical Topology via Morse Theory Persistence and Nonparametric Estimation*. *Contemporary Mathematics* **516**, 75–92. MR2730741 https://doi.org/10.1090/conm/516/10167

Bubenik, P. and Kim, P. (2006). A statistical approach to persistent homology. *Homology, Homotopy, and Applications* **9**. MR2366953 https://doi.org/10.4310/hha.2007.v9.n2.a12

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, 679–698.

Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L. J. and Oudot, S. Y. (2009). Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry. SCG '09*, 237–246. New York, NY, USA: Association for Computing Machinery.

Chazal, F., Cohen-Steiner, D. and Lieutier, A. (2006). A sampling theory for compact sets in Euclidean space. *Discrete & Computational Geometry* **41**, 461–479. MR2486371 https://doi.org/10.1007/s00454-009-9144-8

Chazal, F., Cohen-Steiner, D., Lieutier, A. and Thibert, B. (2007). Shape smoothing using double offset. In *Proceedings—SPM 2007: ACM Symposium on Solid and Physical Modeling*.

Chazal, F., Glisse, M., Labruère Chazal, C. and Michel, B. (2014). Convergence rates for persistence diagram estimation in topological data analysis. In *31st International Conference on Machine Learning, ICML 2014, Vol. 1*. MR3450548

Chazal, F., Guibas, L., Oudot, S. and Skraba, P. (2011). Persistence-based clustering in Riemannian manifolds. *Journal of the ACM* **60**. MR3144911 https://doi.org/10.1145/2535927

Chazal, F. and Lieutier, A. (2005). Weak feature size and persistent homology: Computing homology of solids in Rn from noisy data samples. In *Proceedings of the Twenty-First Annual Symposium on Computational Geometry. SCG '05*, 255–262. MR2460371 https://doi.org/10.1145/1064092.1064132

Key words and phrases. Jumps detection, nonparametric statistics, geometric inference, topological data analysis.

Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence* **4**.

Chazal, F., Oudot, S., Glisse, M. and de Silva, V. (2016). *The Structure and Stability of Persistence Modules*. *SpringerBriefs in Mathematics*. Berlin: Springer. MR3524869 https://doi.org/10.1007/978-3-319-42545-0

Chu, C. K. and Cheng, P. E. (1996). Estimation of jump points and jump values of a density function. *Statistica Sinica* **6**, 79–95. MR1379050

Clough, J. R., Byrne, N., Oksuz, I., Zimmer, V. A., Schnabel, J. A. and King, A. P. (2022). A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**, 8766–8778.

Commaret, A. (2024). Generalized Morse theory for tubular neighborhoods.

Duda, R. and Hart, P. (1974). *Pattern Classification and Scene Analysis*, 271–272. A Wiley-Interscience publication.

Dunaeva, O., Edelsbrunner, H., Lukyanov, A., Machin, M., Malkova, D., Kuvaev, R. and Kashin, S. (2016). The classification of endoscopy images with persistent homology. *Pattern Recognition Letters* **83**, 13–22. Geometric, topological and harmonic trends to image processing.

Edelsbrunner, H. (2013). Persistent homology in image processing. In *Graph-Based Representations in Pattern Recognition 182–183*. Berlin: Springer.

Edelsbrunner, H. and Harer, J. (2009). *Computational Topology*. Providence: American Mathematical Society. MR2572029 https://doi.org/10.1090/mbk/069

Fasy, B., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S. and Singh, A. (2014). Confidence sets for persistence diagrams. *The Annals of Statistics* **42**, 2301–2339. MR3269981 https://doi.org/10.1214/14-AOS1252

Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society*. MR0110078 https://doi.org/10.2307/1993504

Garlipp, T. and Müller, C. H. (2007). Robust jump detection in regression surface. *Sankhyā: The Indian Journal of Statistics (2003–2007)* **69**, 55–86. MR2385278

Gayraud, G. (2002). Minimax estimation of a discontinuity for the density. *Journal of Nonparametric Statistics* **14**, 59–66. MR1905585 https://doi.org/10.1080/10485250211390

Hall, P. and Titterington, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34**, 429–440. MR1190262 https://doi.org/10.2307/1268942

Hatcher, A. (2000). *Algebraic Topology*. Cambridge: Cambridge Univ. Press. MR1867354

Henneuse, H. (2024a). Persistent Diagram Estimation of Multivariate Piecewise Hölder-continuous Signals.

Henneuse, H. (2024b). Persistent Diagram Estimation of Irregular Signals: Beyond “Plug-in” Approaches.

Henneuse, H. (2024c). Persistence-based modes inference.

Islambekov, U. and Gel, Y. R. (2018). Unsupervised space–time clustering using persistent homology. *Environ-Metrics* **30**. MR3948462 https://doi.org/10.1002/env.2539

Kang, Y. and Qiu, P. (2014). Jump detection in blurred regression surfaces. *Technometrics* **56**, 539–550. MR3290623 https://doi.org/10.1080/00401706.2013.844732

Khasawneh, F. and Munch, E. (2015). Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing* **70**.

Kim, J., Shin, J., Chazal, F., Rinaldo, A. and Wasserman, L. (2020). Homotopy Reconstruction via the Cech Complex and the Vietoris–Rips Complex. MR4117767

Korostelev, A. and Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics*. Berlin: Springer. MR1226450 https://doi.org/10.1007/978-1-4612-2712-0

McDonald, J. A. and Owen, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28**, 195–208. MR0853113 https://doi.org/10.2307/1269075

Meng, Z., Anand, D. V., Lu, Y., Wu, J. and Xia, K. (2019). Weighted persistent homology for biomolecular data analysis. *Scientific Reports* **10**.

Muller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics* **20**, 737–761. MR1165590 https://doi.org/10.1214/aos/1176348654

Muller, H. G. and Song, K. S. (1994). Maximin estimation of multidimensional boundaries. *Journal of Multivariate Analysis* **50**, 265–281. MR1293046 https://doi.org/10.1006/jmva.1994.1042

Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry* **39**, 419–441. MR2383768 https://doi.org/10.1007/s00454-008-9053-2

O'Sullivan, F. and Qian, M. (1994). A regularized contrast statistic for object boundary estimation-implementation and statistical evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**, 561–570.

Pascucci, V., Tricoche, X., Hagen, H. and Tierny, J. (2011). *Topological Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications*. Berlin: Springer. MR2849158 https://doi.org/10.1007/978-3-642-15014-2

Pokorny, F. T., Hawasly, M. and Ramamoorthy, S. (2016). Topological trajectory classification with filtrations of simplicial complexes and persistent homology. *The International Journal of Robotics Research* **35**, 204–223.

Prewitt, J. (1970). Object Enhancement and Extraction. Picture processing and Psychopictorics.

Pun, C. S., Xia, K. and Lee, S. (2018). Persistent-homology-based machine learning and its applications—a survey. *SSRN Electronic Journal*.

Qaiser, T., Tsang, Y.-W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D. and Rajpoot, N. (2019). Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis* **55**, 1–14.

Qiu, P. (1992). Estimation of the number of jumps of the jump regression functions. *Communications in Statistics - Theory and Methods* **23**. MR1293176 https://doi.org/10.1080/03610929408831378

Qiu, P. (2002). A nonparametric procedure to detect jumps in regression surfaces. *Journal of Computational and Graphical Statistics* **11**, 799–822. MR1944264 https://doi.org/10.1198/106186002321018795

Qiu, P. (2005). *Image Processing and Jump Regression Analysis*. New York: Wiley. MR2111430 https://doi.org/10.1002/0471733156

Qiu, P. and Yandell, B. (1997). Jump detection in regression surfaces. *Journal of Computational and Graphical Statistics* **6**, 332–354. MR1466871 https://doi.org/10.2307/1390737

Ravishanker, N. and Chen, R. (2021). An introduction to persistent homology for time series. *Wiley Interdisciplinary Reviews Computational Statistics* **13**. MR4242810 https://doi.org/10.1002/wics.1548

Rieck, B. and Leitte, H. (2016). Exploring and comparing clusterings of multivariate data sets using persistent homology. *Computer Graphics Forum* **35**, 81–90.

Roberts, L. (1965). Machine Perception of 3-D Solids. Optical and Electro-optical Information Processing.

Scholtes, S. (2013). On hypersurfaces of positive reach, alternating Steiner formulæ and Hadwiger's Problem.

Seversky, L. M., Davis, S. and Berger, M. (2016). On time-series topological data analysis: New data and opportunities. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1014–1022.

Skraba, P., Ovsjanikov, M., Chazal, F. and Guibas, L. (2010). Persistence-based segmentation of deformable shapes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops*, 45–52.

Sun, R., Lei, T., Chen, Q., Wang, Z., Du, X., Zhao, W. and Nandi, A. K. (2022). Survey of image edge detection. *Frontiers in Signal Processing* **2**.

Suzuki, A., Miyazawa, M., Minto, J., Tsuji, T., Obayashi, I., Hiraoka, Y. and Ito, T. (2021). Flow estimation solely from image data through persistent homology analysis. *Scientific Reports* **11**.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*. Berlin: Springer. MR2724359 https://doi.org/10.1007/b13794

Turner, K., Mukherjee, S. and Boyer, D. M. (2014). Persistent homology transform for modeling shapes and surfaces. *Information and Inference* **3**, 310–344. MR3311455 https://doi.org/10.1093/imaiai/iau011

Umeda, Y. (2017). Time series classification via topological data analysis. *Information Media & Technology* **12**, 228–239.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–397. MR1354236 https://doi.org/10.1093/biomet/82.2.385

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association* **93**, 163–172. MR1614616 https://doi.org/10.2307/2669613

Wong, C.-C. and Vong, C.-M. (2021). Persistent homology based graph convolution network for fine-grained 3D shape segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7098–7107.

Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics* **21**, 1545–1566. MR1241278 https://doi.org/10.1214/aos/1176349271

Zomorodian, A. J. (2005). *Topology for Computing*. *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge: Cambridge University Press. MR2111929 https://doi.org/10.1017/CBO9780511546945

Brazilian Journal of Probability and Statistics
2025, Vol. 39, No. 3, 348–365
https://doi.org/10.1214/25-BJPS638

Quantile importance sampling

Jyotishka Datta[1,a] **and Nicholas G. Polson**[2,b]

[1] *Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA,* [a]*jyotishka@vt.edu*
[2] *Booth School of Business, University of Chicago, Chicago, Illinois, USA,* [b]*ngp@chicagobooth.edu*

Abstract. In Bayesian inference, the approximation of integrals of the form $\psi = \mathbb{E}_F l(X) = \int_\chi l(\mathbf{x})\, dF(\mathbf{x})$ is a fundamental challenge. Such integrals are crucial for evidence estimation, which is important for various purposes, including model selection and numerical analysis. The existing strategies for evidence estimation are classified into four categories: deterministic approximation, density estimation, importance sampling and vertical representation (*SIAM Review* **65** (2023) 3–58). In this paper, we show that the Riemann sum estimator due to (*SIAM Journal on Numerical Analysis* **15** (1978) 1289–1300) can be used in the context of nested sampling (*Bayesian Analysis* **1** (2006) 833–859) to achieve a $O(n^{-4})$ rate of convergence, faster than the usual ergodic central limit theorem, under certain regularity conditions. We provide a brief overview of the literature on the Riemann sum estimators and the nested sampling algorithm and its connections to vertical likelihood Monte Carlo. We provide theoretical and numerical arguments to show how merging these two ideas may result in improved and more robust estimators for evidence estimation, especially in higher-dimensional spaces. We also briefly discuss the idea of simulating the Lorenz curve that avoids the problem of intractable Λ functions, essential for the vertical representation and nested sampling.

References

Ashton, G., Bernstein, N., Buchner, J., Chen, X., Csányi, G., Fowlie, A., Feroz, F., Griffiths, M., Handley, W., Habeck, M., et al. (2022). Nested sampling for physical scientists. *Nature Reviews Methods Primers* **2**, 39.

Chen, X., Feroz, F. and Hobson, M. (2023). Bayesian posterior repartitioning for nested sampling. *Bayesian Analysis* **18**, 695–721. MR4626354 https://doi.org/10.1214/22-ba1323

Chen, X., Hobson, M., Das, S. and Gelderblom, P. (2019). Improving the efficiency and robustness of nested sampling using posterior repartitioning. *Statistics and Computing* **29**, 835–850. MR3955288 https://doi.org/10.1007/s11222-018-9841-3

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321. MR1379473

Chopin, N. and Robert, C. P. (2010). Properties of nested sampling. *Biometrika* **97**, 741–755. MR2672495 https://doi.org/10.1093/biomet/asq021

Datta, J. and Polson, N. G. (2025). Supplement to "Quantile importance sampling." https://doi.org/10.1214/25-BJPS638SUPP

Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method. *Los Alamos Science* **15**, 131–137. MR0935772

Firth, D. (2011). On improved estimation for importance sampling. *Brazilian Journal of Probability and Statistics* **25**, 437–443. MR2832895 https://doi.org/10.1214/11-BJPS155

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models. Analytical Methods for Social Research*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Goldie, C. M. (1977). Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability* **9**, 765–791. MR0478267 https://doi.org/10.2307/1426700

Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37**, 185–194.

Hesterberg, T. C. (1988). Advances in importance sampling. PhD thesis, Stanford University. MR2637036

Key words and phrases. Importance sampling, nested sampling, vertical-likelihood, evidence estimation, Monte Carlo.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. New York: Springer. MR1842342

Llorente, F., Martino, L., Delgado, D. and López-Santiago, J. (2023). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. *SIAM Review* **65**, 3–58. MR4545927 https://doi.org/10.1137/20M1310849

Martin, G. M., Frazier, D. T. and Robert, C. P. (2024). Computing Bayes: From then 'til now. *Statistical Science* **39**, 3–19. MR4718524 https://doi.org/10.1214/22-sts876

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association* **44**, 335–341. MR0031341

Neal, R. (1994). Contribution to the discussion of "Approximate Bayesian inference with the weighted likelihood bootstrap" by Michael A. Newton and Adrian E. Raftery. *Journal of the Royal Statistical Society, Series B, Methodological* **56**, 41–42. MR1257793

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B, Methodological* **56**, 1–48. MR1257793

Philippe, A. (1997). Processing simulation output by Riemann sums. *Journal of Statistical Computation and Simulation* **59**, 295–314. MR1623186 https://doi.org/10.1080/00949659708811863

Philippe, A. and Robert, C. P. (2001). Riemann sums for MCMC estimation and convergence monitoring. *Statistics and Computing* **11**, 103–115. MR1837130 https://doi.org/10.1023/A:1008926514119

Polson, N. G. and Scott, J. G. (2014). Vertical-likelihood Monte Carlo. arXiv preprint. Available at arXiv:1409.3601.

Raftery, A. E., Newton, M. A., Satagopan, J. M. and Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics* **8**, 1–45. MR2433201

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods 2*. New York, NY: Springer. MR1707311 https://doi.org/10.1007/978-1-4757-3071-5

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**, 833–859. MR2282208 https://doi.org/10.1214/06-BA127

Skilling, J. (2008). Tests of Nested Sampling. Unpublished manuscript.

Yakowitz, S., Krimmel, J. and Szidarovszky, F. (1978). Weighted Monte Carlo integration. *SIAM Journal on Numerical Analysis* **15**, 1289–1300. MR0512700 https://doi.org/10.1137/0715088

Brazilian Journal of Probability and Statistics
2025, Vol. 39, No. 3, 366–376
https://doi.org/10.1214/25-BJPS640

Variance estimation in the presence of scrambled response using ranked set sampling

Arvind Kumar Alok[1,a], Arnab Bandyopadhyay[2] and Garib Nath Singh[1]

[1]*Department of Mathematics and Computing, Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand – 826004, India,* [a]*arvindkralok@gmail.com*

[2]*Department of Basic Science and Humanities, Dr. B. C. Roy Engineering College, Durgapur – 713206, West Bengal, India*

Abstract. This study addresses a significant statistical challenge: estimating population variance in the context of ranked set sampling (RSS). The scenario is particularly complex due to the presence of both study and auxiliary variables that exhibit multiplicative and additive scrambled responses. This intricacy necessitates a robust approach to estimate the population variance accurately. To assess the performance and effectiveness of the proposed estimator, both simulation and empirical studies have been conducted. These studies were designed to rigorously evaluate the efficacy of the new estimator in comparison to existing methods. The results obtained from these investigations indicate that the proposed estimator demonstrates significant dominance over the natural estimator of population variance, highlighting its reliability and accuracy in practical applications.

References

Ahmed, S., Albalawi, O. and Shabbir, J. (2024). On indirect estimation of small area parameters under ranked set sampling. *Alexandria Engineering Journal* **107**, 690–697.

Ahmed, S. and Shabbir, J. (2017). Ranked set sampling with scrambled response model to subsample non-respondents.

Ahmed, S. and Shabbir, J. (2019). Extreme-cum-median ranked set sampling.

Al-Odat, M. T. and Al-Saleh, M. F. (2001). A variation of ranked set sampling. *Journal of Applied Statistical Science* **10**, 137–146. MR1834451

Al-Omari, A. and Bouza, C. (2015). Ratio estimators of the population mean with missing values using ranked set sampling. *EnvironMetrics* **26**, 67–76. MR3324902 https://doi.org/10.1002/env.2286

Biradar, B., Santosha, C., et al (2014). Estimation of the mean of the exponential distribution using maximum ranked set sampling with unequal samples. *Open Journal of Statistics* **4**, 641.

Bouza, C. N. (2002). Ranked set sub sampling the non response strata for estimating the difference of means. *Biometrical Journal* **44**, 903–915. MR1934971 https://doi.org/10.1002/1521-4036(200210)44:7<903::AID-BIMJ903>3.0.CO;2-A

Cochran, W. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *Journal of Agricultural Science* **30**, 262–275.

Cochran, W. G. (2007). *Sampling Techniques*. New York: Wiley. MR0054199

Das, A. K. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhyà, Series C* **40**, 139–148.

Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 545–555.

Gupta, S., Mehta, S., Shabbir, J. and Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative rrt models. *Journal of Statistical Theory and Practice* **12**, 506–511. MR3836435 https://doi.org/10.1080/15598608.2017.1415175

Gupta, S., Qureshi, M. N. and Khalil, S. (2020). Variance estimation using randomized response technique. *REVSTAT Statistical Journal* **18**, 165–176. MR4105234

Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association* **78**, 117–123. MR0696855

Key words and phrases. Ranked set sampling, study variable, auxiliary variable, scrambled responses, bias, optimum MSEs, percentage relative efficiency (PRE), simulation studies.

Kadılar, C. and Cingi, H. (2006). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics* **35**, 111–115. MR2267274

McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research* **3**, 385–390.

Muttlak, H. (1998). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *EnvironMetrics* **9**, 255–267. MR1858331

Perron, F. and Sinha, B. K. (2004). Estimation of variance based on a ranked set sample. *Journal of Statistical Planning and Inference* **120**, 21–28. MR2026480 https://doi.org/10.1016/S0378-3758(02)00497-4

Saleem, I., Sanaullah, A., Al-Essa, L. A., Bashir, S. and Al Mutairi, A. (2023). Efficient estimation of population variance of a sensitive variable using a new scrambling response model. *Scientific Reports* **13**, 19913.

Samawi, H. M. and Tawalbeh, E. M. (2002). Double median ranked set sample: Comparing to other double ranked samples for mean and ratio estimators. *Journal of Modern Applied Statistical Methods* **1**, 428–442. MR1769826 https://doi.org/10.1016/S0167-7152(99)00206-0

Singh, G. N., Alok, A. K. and Chandra, P. (2024). Analysis of stress-strength modeling utilizing median-ranked set sampling with its illustration. *International Journal of Reliability, Quality, and Safety Engineering*.

Singh, S., Horn, S., Chowdhury, S. and Yu, F. (1999). Theory & methods: Calibration of the estimators of variance. *Australian & New Zealand Journal of Statistics* **41**, 199–212. MR1705413 https://doi.org/10.1111/1467-842X.00074

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics* **20**, 1–31. MR0228143 https://doi.org/10.1007/BF02911622

Wolfe, D. A. (2012). Ranked set sampling: Its relevance and impact on statistical inference. *International Scholarly Research Notices* **2012**, 568385.

Yan, Z., Wang, J. and Lai, J. (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics - Theory and Methods* **38**, 400–408. MR2510792 https://doi.org/10.1080/03610920802220785

Brazilian Journal of Probability and Statistics
2025, Vol. 39, No. 3, 377–407
https://doi.org/10.1214/25-BJPS642

Novel exponential-type imputation models for handling missing data

Anjali Gupta[a] **and R. R. Sinha**[b]

Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, 144008, India, [a]*anjaligupta07@yahoo.com,* [b]*raghawraman@gmail.com*

Abstract. Missing data is a common issue across fields such as engineering, finance, healthcare and social sciences. If not addressed appropriately, it can lead to biased analyses and reduced statistical power. This paper introduces a family of novel Exponential-Type Imputation Models (NEtIMs) designed under three distinct strategies to handle missing data effectively. These models exploit the properties of newly formulated mean estimators by incorporating measures such as absolute relative bias and mean squared error to better capture uncertainty. The performance of NEtIMs is evaluated through extensive simulations on both symmetric and asymmetric datasets, as well as real-world data. Results show that NEtIMs consistently outperform conventional imputation methods in terms of accuracy, efficiency and robustness across different missing data mechanisms. Additionally, optimality constraints are established to demonstrate the broader applicability and reliability of the proposed mean estimators formulated through NEtIMs.

References

Bahl, S. and Tuteja, R. (1991). Ratio and product type exponential estimators. *Journal of Information & Optimization Sciences* **12**, 159–164. MR1108297 https://doi.org/10.1080/02522667.1991.10699058

Bhushan, S., Kumar, A., Pandey, A. P. and Singh, S. (2023). Estimation of population mean in presence of missing data under simple random sampling. *Communications in Statistics Simulation and Computation* **52**, 6048–6069. MR4672535 https://doi.org/10.1080/03610918.2021.2006713

Bhushan, S. and Pandey, A. P. (2016). Optimal imputation of missing data for estimation of population mean. *Journal of Statistics & Management Systems* **19**, 755–769.

Bhushan, S. and Pandey, A. P. (2018). Optimality of ratio type estimation methods for population mean in the presence of missing data. *Communications in Statistics—Theory and Methods* **47**, 2576–2589. MR3780252 https://doi.org/10.1080/03610926.2016.1167906

Bhushan, S. and Pandey, A. P. (2023). New chain imputation methods for estimating population mean in the presence of missing data using two auxiliary variables. *Communications in Mathematics and Statistics* **11**, 325–340. MR4599600 https://doi.org/10.1007/s40304-021-00251-w

Diana, G. and Perri, F. P. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics—Theory and Methods* **39**, 3245–3251. MR2747581 https://doi.org/10.1080/03610920903009400

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2012). *Applied Longitudinal Analysis*. New York: Wiley & Sons. MR2830137

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology* **60**, 549–576.

Grover, L. K. and Kaur, A. (2022). Additive ratio type exponential estimator of finite population mean of sensitive variable using non-sensitive auxiliary information based on optional randomized response model. *Brazilian Journal of Probability and Statistics* **36**, 463–481. MR4489176 https://doi.org/10.1214/22-bjps535

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *International Journal of Public Opinion Quarterly* **70**, 646–675.

Gupta, A. and Sinha, R. R. (2025). Supplement to “Novel exponential-type imputation models for handling missing data.” https://doi.org/10.1214/25-BJPS642SUPP

Heitjan, D. F. and Basu, S. (1996). Distinguishing “missing at random” and “missing completely at random”. *American Statistician* **50**, 207–213. MR1422070 https://doi.org/10.2307/2684656

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity* **34**, 331–351.

Key words and phrases. Imputation, mean, mean square error, coverage probability, simulation.

Kalton, G., Kasprzyk, D. and Santos, R. (1981). Issues of nonresponse and imputation in the survey of income and program participation. In *Current Topics in Survey Sampling*, 455–480. San Diego: Academic Press.

Lee, H., Rancourt, E. and Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics* **10**, 231–231.

Lee, H., Rancourt, E. and Särndal, C. E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. In *Proc. of the American Statistical Association (Social Survey Research Methods Section)*, 384–389.

Little, M., Mcsharry, P., Roberts, S., Costello, D. and Moroz, I. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nature Precedings*, 1–1.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons. MR1925014 https://doi.org/10.1002/9781119013563

R Core Team (2024). A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Reddy, V. N. (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya C* **40**, 29–37.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1979). Illustrating the use of multiple imputations to handle nonresponse in sample surveys. *Bulletin of the International Statistical Institute* **48**, 517–532. MR0731562

Santos, R. (1981). Effects of imputation on regression coefficients. In *Proceedings of the Section on Survey Research Methods*, 140–145. Alexandria: American Statistical Association.

Singh, G. N., Pandey, A. K. and Sharma, A. K. (2021). Some improved and alternative imputation methods for finite population mean in presence of missing information. *Communications in Statistics—Theory and Methods* **50**, 4401–4427. MR4305818 https://doi.org/10.1080/03610926.2020.1713375

Singh, G. N., Priyanka, K., Kim, J. M. and Singh, S. (2010). Estimation of population mean using imputation techniques in sample surveys. *Journal of the Korean Statistical Society* **39**, 67–74. MR2655812 https://doi.org/10.1016/j.jkss.2009.04.002

Singh, S. (2009). A new method of imputation in survey sampling. *Statistics* **43**, 499–511. MR2566697 https://doi.org/10.1080/02331880802605114

Singh, S. and Deo, B. (2003). Imputation by power transformation. *Statistical Papers* **44**, 555–579. MR2008052 https://doi.org/10.1007/BF02926010

Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. *Metrika* **51**, 267–276. MR1795374 https://doi.org/10.1007/s001840000054

Srivastava, S. K. and Jhajj, H. S. (1983). *Calcutta Statistical Association Bulletin* **32**, 47–56. MR0733537 https://doi.org/10.1177/0008068319830104

AIMS AND SCOPE

The Brazilian Journal of Probability and Statistics aims to publish high quality research papers in applied probability, applied statistics, computational statistics, mathematical statistics, probability theory and stochastic processes.

More specifically, the following types of contributions will be considered:

(i) Original articles dealing with methodological developments, comparison of competing techniques or their computational aspects;
(ii) Original articles developing theoretical results;
(iii) Articles that contain novel applications of existing methodologies to practical problems. For these papers the focus is in the importance and originality of the applied problem, as well as, applications of the best available methodologies to solve it.
(iv) Survey articles containing a thorough coverage of topics of broad interest to probability and statistics. The journal will occasionally publish book reviews, invited papers and essays on the teaching of statistics.

GENERAL INFORMATION

Journal Homepage: https://www.imstat.org/journals-and-publications/brazilian-journal-of-probability-and-statistics/

Submissions: Manuscripts for *Brazilian Journal of Probability and Statistics* should be submitted online. Authors may access the Electronic Journals Management System (EJMS) at https://www.e-publications.org/ims/submission/.

Preparation of Manuscripts: https://imstat.org/journals-and-publications/brazilian-journal-of-probability-and-statistics/brazilian-journal-of-probability-and-statistics-manuscript-submission/

Correspondence. Mail concerning membership, subscriptions, nonreceipt claims, copyright permissions, advertising or back issues should be sent to the IMS Dues and Subscription Office, PO Box 729, Middletown, Maryland 21769, USA. Mail concerning submissions or editorial content should be sent to the Editor at nancy@ime.unicamp.br. Mail concerning the production of this journal should be sent to: Geri Mattson at bjps@mattsonpublishing.com.

Individual and Organizational Memberships:
https://www.imstat.org/individual-membership/

Individual and General Subscriptions:
https://imstat.org/journals-and-publications/subscriptions-and-orders/

Electronic Access to IMS Journals:
https://www.imstat.org/journals-and-publications/electronic-access/

The *Brazilian Journal of Probability and Statistics* is an IMS supported journal:
https://www.imstat.org/journals-and-publications/ims-supported-journals/