Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 34 • Number 2 • May 2020

Brazilian Journal of Probability and Statistics

Volume 34 • Number 2 • May 2020

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 34, Number 2, May 2020. 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, 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
2020, Vol. 34, No. 2, 203
https://doi.org/10.1214/20-BJPS342ED

A message from the editorial board

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 204–229
https://doi.org/10.1214/20-BJPS466

Recent developments in complex and spatially correlated functional data

Israel Martínez-Hernández and Marc G. Genton

King Abdullah University of Science and Technology

Abstract. As high-dimensional and high-frequency data are being collected on a large scale, the development of new statistical models is being pushed forward. Functional data analysis provides the required statistical methods to deal with large-scale and complex data by assuming that data are continuous functions, for example, realizations of a continuous process (curves) or continuous random field (surfaces), and that each curve or surface is considered as a single observation. Here, we provide an overview of functional data analysis when data are complex and spatially correlated. We provide definitions and estimators of the first and second moments of the corresponding functional random variable. We present two main approaches: The first assumes that data are realizations of a functional random field, that is, each observation is a curve with a spatial component. We call them *spatial functional data*. The second approach assumes that data are continuous deterministic fields observed over time. In this case, one observation is a surface or manifold, and we call them *surface time series*. For these two approaches, we describe software available for the statistical analysis. We also present a data illustration, using a high-resolution wind speed simulated dataset, as an example of the two approaches. The functional data approach offers a new paradigm of data analysis, where the continuous processes or random fields are considered as a single entity. We consider this approach to be very valuable in the context of big data.

References

Abdulah, S., Li, Y., Cao, J., Ltaief, H., Keyes, D. E., Genton, M. G. and Sun, Y. (2019). ExaGeoStatR: A package for large-scale geostatistics in R. Available at arXiv:1908.06936.

Aguilera-Morillo, M. C., Durbán, M. and Aguilera, A. M. (2017). Prediction of functional data with spatial dependence: A penalized approach. *Stochastic Environmental Research and Risk Assessment* **31**, 7–22.

Alfeld, P., Neamtu, M. and Schumaker, L. L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *Journal of Computational and Applied Mathematics* **73**, 5–43. MR1424867 https://doi.org/10.1016/0377-0427(96)00034-9

Arnone, E., Azzimonti, L., Nobile, F. and Sangalli, L. M. (2019). Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate Analysis* **170**, 275–295. Special Issue on Functional Data Analysis and Related Topics. MR3913041 https://doi.org/10.1016/j.jmva.2018.09.006

Aston, J. A. D., Pigoli, D. and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**, 1431–1461. MR3670184 https://doi.org/10.1214/16-AOS1495

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110**, 378–392. MR3338510 https://doi.org/10.1080/01621459.2014.909317

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M. and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with PDE penalization. *Journal of the American Statistical Association* **110**, 1057–1071. MR3420684 https://doi.org/10.1080/01621459.2014.946036

Baladandayuthapani, V., Mallick, B. K., Hong, M. Y., Lupton, J. R., Turner, N. D. and Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* **64**, 64–73. MR2422820 https://doi.org/10.1111/j.1541-0420.2007.00846.x

Key words and phrases. Functional data, functional random field, manifold data, spatial functional data, spatial statistics, spatio-temporal statistics, surface data.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*. *Monographs on Statistics and Applied Probability.* **135**. Boca Raton, FL: CRC Press. MR3362184

Bel, L., Bar-Hen, A., Petit, R. and Cheddadi, R. (2011). Spatio-temporal functional regression on paleoecological data. *Journal of Applied Statistics* **38**, 695–704. MR2773575 https://doi.org/10.1080/02664760903563650

Bernardi, M. S., Sangalli, L. M., Mazza, G. and Ramsay, J. O. (2017). A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic Environmental Research and Risk Assessment* **31**, 23–38.

Bohorquez, M., Giraldo, R. and Mateu, J. (2017). Multivariate functional random fields: Prediction and optimal sampling. *Stochastic Environmental Research and Risk Assessment* **31**, 53–70.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. *Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

Caballero, W., Giraldo, R. and Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment* **27**, 1553–1563.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22. MR1718346 https://doi.org/10.1016/S0167-7152(99)00036-X

Chen, K., Zhang, X., Petersen, A. and Müller, H.-G. (2017). Quantifying infinite-dimensional data: Functional data analysis in action. *Statistics in Biosciences* **9**, 582–604.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. M. and Punjabi, N. M. (2011). Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association* **106**, 775–790. MR2894733 https://doi.org/10.1198/jasa.2011.ap10089

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc. MR2848400

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: John Wiley & Sons, Inc. Paperback edition of the 1993 edition. MR3559472

Dassi, F., Ettinger, B., Perotto, S. and Sangalli, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics* **90**, 111–131. MR3300898 https://doi.org/10.1016/j.apnum.2014.10.007

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *EnvironMetrics* **21**, 224–239. MR2842240 https://doi.org/10.1002/env.1003

Diggle, P. J. and Ribeiro, P. J. Jr. (2007). *Model-Based Geostatistics*. *Springer Series in Statistics*. New York: Springer. MR2293378

Duchamp, T. and Stuetzle, W. (2003). Spline smoothing on surfaces. *Journal of Computational and Graphical Statistics* **12**, 354–381. MR1983159 https://doi.org/10.1198/1061860031743

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables* (W. Schempp and K. Zeller, eds.) 85–100. Berlin, Heidelberg: Springer. MR0493110

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121. MR1435485 https://doi.org/10.1214/ss/1038425655

Ettinger, B., Perotto, S. and Sangalli, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika* **103**, 71–88. MR3465822 https://doi.org/10.1093/biomet/asv069

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. New York: Springer. MR2229687

Finley, A. O., Banerjee, S. and Gelfand, A. E. (2015). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software* **63**, 1–28.

Galeano, P. and Peña, D. (2019). Data science, big data and statistics. *TEST* **28**, 289–329. MR3962048 https://doi.org/10.1007/s11749-019-00651-9

Genton, M. G., Johnson, C., Potter, K., Stenchikov, G. and Sun, Y. (2014). Surface boxplots. *Stat* **3**, 1–11. MR4027321 https://doi.org/10.1002/sta4.39

Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**, 147–163. MR3353096 https://doi.org/10.1214/14-STS487

Giraldo, R., Dabo-Niang, S. and Martínez, S. (2018). Statistical modeling of spatial big data: An approach from a functional data analysis perspective. *Statistics & Probability Letters* **136**, 126–129. MR3806852 https://doi.org/10.1016/j.spl.2018.02.025

Giraldo, R., Delicado, P. and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 66–82. MR2755385 https://doi.org/10.1007/s13253-009-0012-z

Giraldo, R., Delicado, P. and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics* **18**, 411–426. MR2832903 https://doi.org/10.1007/s10651-010-0143-y

Giraldo, R., Delicado, P. and Mateu, J. (2012). Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica* **66**, 403–421. MR2983302 https://doi.org/10.1111/j.1467-9574.2012.00522.x

Giraldo, R., Delicado, P. and Mateu, J. (2015). geofd: Spatial prediction for function value data. R package version 1.0.

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

Goulard, M. and Voltz, M. (1993) *Geostatistical Interpolation of Curves: A Case Study in Soil Science*, 805–816. Dordrecht: Springer.

Greco, F., Ventrucci, M. and Castelli, E. (2018). P-spline smoothing for spatial data collected worldwide. *Spatial Statistics* **27**, 1–17. MR3868196 https://doi.org/10.1016/j.spasta.2018.08.008

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Annals of Applied Statistics* **6**, 669–696. MR2976487 https://doi.org/10.1214/11-AOAS524

Grujic, O. and Menafoglio, A. (2017). fdagstat, an R package. R package version 1.0.

Grujic, O., Menafoglio, A., Yang, G. and Caers, J. (2018). Cokriging for multivariate Hilbert space valued random fields: Application to multi-fidelity computer code emulation. *Stochastic Environmental Research and Risk Assessment* **32**, 1955–1971.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Hall, P., Fisher, N. I. and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics* **22**, 2115–2134. MR1329185 https://doi.org/10.1214/aos/1176325774

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields* **99**, 399–424. MR1283119 https://doi.org/10.1007/BF01199899

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

Hörmann, S. and Kokoszka, P. (2012). Chapter 7—functional time series. In *Time Series Analysis: Methods and Applications* (T. S. Rao, S. S. Rao and C. Rao, eds.), *Handbook of Statistics* **30**, 157–186. Amsterdam: Elsevier. MR3295420

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

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **75**, 103–122. MR3008273 https://doi.org/10.1111/j.1467-9868.2012.01032.x

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179**, 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**, 4942–4956. MR2364551 https://doi.org/10.1016/j.csda.2006.07.028

Ignaccolo, R., Mateu, J. and Giraldo, R. (2014). Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment* **28**, 1171–1186.

Jiang, H. and Serban, N. (2012). Clustering random curves under spatial interdependence with application to service accessibility. *Technometrics* **54**, 108–119. MR2929427 https://doi.org/10.1080/00401706.2012.657106

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis* **34**, 116–129. MR3008019 https://doi.org/10.1111/j.1467-9892.2012.00816.x

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

Lai, M.-J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. *Encyclopedia of Mathematics and Its Applications.* Cambridge: Cambridge University Press. MR2355272 https://doi.org/10.1017/CBO9780511721588

Lee, D.-J., Zhu, Z. and Toscas, P. (2015). Spatio-temporal functional data analysis for wireless sensor networks data. *EnvironMetrics* **26**, 354–362. MR3366969 https://doi.org/10.1002/env.2344

Lila, E., Aston, J. A. D. and Sangalli, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *Annals of Applied Statistics* **10**, 1854–1879. MR3592040 https://doi.org/10.1214/16-AOAS975

Lila, E., Sangalli, L. M., Ramsay, J. and Formaggia, L. (2019). fdaPDE: Functional data analysis and partial differential equations; statistical analysis of functional and spatial data, based on regression with partial differential regularizations. R package version 0.1-6.

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. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **73**, 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

Liu, C., Ray, S. and Hooker, G. (2017). Functional principal component analysis of spatially correlated data. *Statistics and Computing* **27**, 1639–1654. MR3687330 https://doi.org/10.1007/s11222-016-9708-4

Martínez-Hernández, I. and Genton, M. G. (2020). Nonparametric trend estimation in functional time series with application to annual mortality rates. Available at arXiv:2001.04660.

Martínez-Hernández, I., Genton, M. G. and González-Farías, G. (2019). Robust depth-based estimation of the functional autoregressive model. *Computational Statistics & Data Analysis* **131**, 66–79. MR3906795 https://doi.org/10.1016/j.csda.2018.06.003

Mateu, J. and Romano, E. (2017). Advances in spatial functional statistics. *Stochastic Environmental Research and Risk Assessment* **31**, 1–6.

Menafoglio, A., Gaetani, G. and Secchi, P. (2018). Random domain decompositions for object-oriented kriging over complex domains. *Stochastic Environmental Research and Risk Assessment* **32**, 3421–3437.

Menafoglio, A., Grujic, O. and Caers, J. (2016). Universal kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales. *Spatial Statistics* **15**, 39–55. MR3457667 https://doi.org/10.1016/j.spasta.2015.12.003

Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research* **258**, 401–410. MR3582352 https://doi.org/10.1016/j.ejor.2016.09.061

Menafoglio, A., Secchi, P. and Dalla Rosa, M. (2013). A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electronic Journal of Statistics* **7**, 2209–2240. MR3108813 https://doi.org/10.1214/13-EJS843

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P. and Gutstein, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Annals of Applied Statistics* **5**, 894–923. MR2840180 https://doi.org/10.1214/10-AOAS407

Nerini, D., Monestiez, P. and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis* **101**, 409–418. MR2564350 https://doi.org/10.1016/j.jmva.2009.03.005

Nychka, D., Furrer, R., Paige, J. and Sain, S. (2017). fields: Tools for spatial data. R package version 9.9.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences* **30**, 683–691.

Qingguo, T. and Longsheng, C. (2010). B-spline estimation for spatial data. *Journal of Nonparametric Statistics* **22**, 197–217. MR2598962 https://doi.org/10.1080/10485250903272569

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2018). fda: Functional data analysis. R package version 2.4.8. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society, Series B* **64**, 307–319. MR1904707 https://doi.org/10.1111/1467-9868.00339

Rekabdarkolaee, H. M., Krut, C., Fuentes, M. and Reich, B. J. (2019). A Bayesian multivariate functional model with spatially varying coefficient approach for modeling hurricane track data. *Spatial Statistics* **29**, 351–365. MR3907916 https://doi.org/10.1016/j.spasta.2018.12.006

Reyes, A., Giraldo, R. and Mateu, J. (2015). Residual kriging for functional spatial prediction of salinity curves. *Communications in Statistics Theory and Methods* **44**, 798–809. MR3315788 https://doi.org/10.1080/03610926.2012.753087

Ribeiro, P. J. Jr. and Diggle, P. J. (2018). geoR: Analysis of geostatistical data. R package version 1.7-5.2.1.

Romano, E., Balzanella, A. and Verde, R. (2017). Spatial variability clustering for spatially dependent functional data. *Statistics and Computing* **27**, 645–658. MR3613590 https://doi.org/10.1007/s11222-016-9645-2

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **71**, 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

Ruiz-Medina, M. (2011). Spatial autoregressive and moving average Hilbertian processes. *Journal of Multivariate Analysis* **102**, 292–305. MR2739116 https://doi.org/10.1016/j.jmva.2010.09.005

Ruiz-Medina, M. (2012). New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics* **1**, 82–91.

Ruiz-Medina, M. D., Salmerón, R. and Angulo, J. M. (2007). Kalman filtering from POP-based diagonalization of ARH(1). *Computational Statistics & Data Analysis* **51**, 4994–5008. MR2361366 https://doi.org/10.1016/j.csda.2006.07.013

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge: Cambridge University Press. MR1998720 https://doi.org/10.1017/CBO9780511755453

Sangalli, L. M. (2020). A novel approach to the analysis of spatial and functional data over complex domains. *Quality Engineering* **32**, 181–190. https://doi.org/10.1080/08982112.2019.1659357

Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society, Series B* **75**, 681–703. MR3091654 https://doi.org/10.1111/rssb.12009

Sartori, I. and Torriani, L. (2019). Manifoldgstat, an R package. R package.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. *Texts in Statistical Science Series*. London: Chapman & Hall/CRC. MR2134116

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Auel, J., Menck, P. J., Gross, S., Ober, U., Ribeiro, P., Ripley, B. D., Singleton, R. and Pfaff, B. (R Core Team) (2019). RandomFields: Simulation and analysis of random fields. R package version 3.3.6.

Scott-Hayward, L. A. S., Mackenzie, M. L., Donovan, C. R., Walker, C. G. and Ashe, E. (2014). Complex region spatial smoother (CReSS). *Journal of Computational and Graphical Statistics* **23**, 340–360. MR3215814 https://doi.org/10.1080/10618600.2012.762920

Song, J. J. and Mallick, B. (2019). Hierarchical Bayesian models for predicting spatially correlated curves. *Statistics* **53**, 196–209. MR3900086 https://doi.org/10.1080/02331888.2018.1547905

Staicu, A.-M., Crainiceanu, C. M. and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11**, 177–194.

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

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20**, 316–334. MR2847798 https://doi.org/10.1198/jcgs.2011.09224

Sun, Y. and Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *EnvironMetrics* **23**, 54–64. MR2873783 https://doi.org/10.1002/env.1136

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer. ISBN 0-387-95457-0. MR1337030 https://doi.org/10.1007/978-1-4899-2819-1

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing* **2**, 5–16. MR0618629 https://doi.org/10.1137/0902002

Wang, H. and Ranalli, M. G. (2007). Low-rank smoothing splines on complicated domains. *Biometrics* **63**, 209–217. MR2345591 https://doi.org/10.1111/j.1541-0420.2006.00674.x

Wilhelm, M., Dedè, L., Sangalli, L. M. and Wilhelm, P. (2016). IGS: An IsoGeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering* **302**, 70–89. MR3461105 https://doi.org/10.1016/j.cma.2015.12.028

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: John Wiley & Sons, Inc. Paperback edition of the 1993 edition. MR3559472

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. London: Chapman & Hall. MR2206355

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62**, 1025–1036. MR2297673 https://doi.org/10.1111/j.1541-0420.2006.00574.x

Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society, Series B* **70**, 931–955. MR2530324 https://doi.org/10.1111/j.1467-9868.2008.00665.x

Xiao, L., Li, Y. and Ruppert, D. (2013). Fast bivariate P-splines: The sandwich smoother. *Journal of the Royal Statistical Society, Series B* **75**, 577–599. MR3065480 https://doi.org/10.1111/rssb.12007

Yip, C. M. A. (2018). Statistical characteristics and mapping of near-surface and elevated wind resources in the Middle East. Ph.D. Thesis, King Abdullah University of Science and Technology.

Yue, Y. and Speckman, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics* **19**, 96–116. MR2654402 https://doi.org/10.1198/jcgs.2009.08124

Zhang, J., Clayton, M. K. and Townsend, P. A. (2011). Functional concurrent linear regression model for spatial images. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 105–130. MR2767762 https://doi.org/10.1007/s13253-010-0047-1

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A. and Morris, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *Journal of the American Statistical Association* **111**, 772–786. MR3538704 https://doi.org/10.1080/01621459.2015.1042581

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association* **105**, 390–400. MR2757206 https://doi.org/10.1198/jasa.2010.tm08737

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 230–250
https://doi.org/10.1214/19-BJPS431

Agnostic tests can control the type I and type II errors simultaneously

Victor Coscrato, Rafael Izbicki and Rafael B. Stern

Federal University of São Carlos

Abstract. Despite its common practice, statistical hypothesis testing presents challenges in interpretation. For instance, in the standard frequentist framework there is no control of the type II error. As a result, the non-rejection of the null hypothesis (H_0) cannot reasonably be interpreted as its acceptance. We propose that this dilemma can be overcome by using agnostic hypothesis tests, since they can control the type I and II errors simultaneously. In order to make this idea operational, we show how to obtain agnostic hypothesis in typical models. For instance, we show how to build (unbiased) uniformly most powerful agnostic tests and how to obtain agnostic tests from standard p-values. Also, we present conditions such that the above tests can be made logically coherent. Finally, we present examples of consistent agnostic hypothesis tests.

References

Berg, N. (2004). No-decision classification: An alternative to testing for statistical significance. *The Journal of Socio-Economics* **33**, 631–650. https://doi.org/10.1016/j.socec.2004.09.036

Bickel, P. J. and Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume I*, Vol. 117. CRC Press. MR3445928

Casella, G. and Berger, R. L. (2002). *Statistical Inference, Vol. 2*. Pacific Grove, CA: Duxbury. MR1051420

Cohen, J. (1977). Chapter 9—F tests of variance proportions in multiple regression/correlation analysis. In *Statistical Power Analysis for the Behavioral Sciences*, Revised ed. (J. Cohen, ed.) 407–453. Academic Press.

DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics*. Addison-Wesley. MR0373075

Esteves, L. G., Izbicki, R., Stern, J. M. and Stern, R. B. (2016). The logical consistency of simultaneous agnostic hypothesis tests. *Entropy* **18**, 256. MR3550265 https://doi.org/10.3390/e18070256

Geisser, S. and Johnson, W. O. (2006). *Modes of Parametric Statistical Inference, Vol. 529*. John Wiley & Sons. MR2272721

Izbicki, R. and Esteves, L. G. (2015). Logical consistency in simultaneous statistical test procedures. *Logic Journal of the IGPL* **23**, 732–758. MR3403972 https://doi.org/10.1093/jigpal/jzv027

Jeske, D. R., Linehan, J. A., Wilson, T. G., Kawachi, M. H., Wittig, K., Lamparska, K., Amparo, C., Mejia, R., Lai, F., Georganopoulou, D. and Steven, S. S. (2017). Two-stage classifiers that minimize pca3 and the psa proteolytic activity testing in the prediction of prostate cancer recurrence after radical prostatectomy. *The Canadian Journal of Urology* **24**, 9089–9097.

Jeske, D. R. and Smith, S. (2017). Maximizing the usefulness of statistical classifiers for two populations with illustrative applications. In *Statistical Methods in Medical Research*. MR3825911 https://doi.org/10.1177/0962280216680244

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795. MR3363402 https://doi.org/10.1080/01621459.1995.10476572

Lehmann, E. L. and Romano, J. P. (2006). *Testing Statistical Hypotheses*. Springer. MR2135927

Lei, J. (2014). Classification with confidence. *Biometrika* **101**, 755–769. MR3286915 https://doi.org/10.1093/biomet/asu038

Levine, T. R., Weber, R., Park, H. S. and Hullett, C. R. (2008). A communication researchers' guide to null hypothesis significance testing and alternatives. *Human Communication Research* **34**, 188–209. https://doi.org/10.1111/j.1468-2958.2008.00318.x

Monahan, J. F. (2008). *A Primer on Linear Models*. CRC Press. MR2402599

Key words and phrases. Hypothesis test, uniformly most powerful tests, logical consistency, three-decision problem.

Morey, R. D. and Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor.

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*.

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996). *Applied Linear Statistical Models, Vol. 4*. Chicago: Irwin.

Neyman, J. (1976). Tests of statistical hypotheses and their use in studies of natural phenomena. *Communications in Statistics Theory and Methods* **5**, 737–751. MR0458772 https://doi.org/10.1080/03610927608827392

Sadinle, M., Lei, J. and Wasserman, L. (2019). Least ambiguous set-valued classifiers with bounded error levels. *Journal of the American Statistical Association*. **114**, 223–234. https://doi.org/10.1080/01621459.2017.1395341

Shao, J. (2003). *Mathematical Statistics, Vol. 2*. Springer. MR2002723 https://doi.org/10.1007/b97553

Stern, J. M. (2011). Symmetry, invariance and ontology in physics and statistics. *Symmetry* **3**, 611–635. MR2845301 https://doi.org/10.3390/sym3030611

Stern, J. M. (2017). Continuous versions of haack's puzzles: Equilibria, eigen-states and ontologies. *Logic Journal of the IGPL* **25**, 604–631. MR3685097 https://doi.org/10.1093/jigpal/jzx017

Stern, J. M., Esteves, L. G., Izbicki, R. and Stern, R. B. (2017). Logically-consistent hypothesis testing and the hexagon of oppositions. *Logic Journal of the IGPL* **25**, 741–757. https://doi.org/10.1093/jigpal/jzx024

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 251–272
https://doi.org/10.1214/18-BJPS421

Random environment binomial thinning integer-valued autoregressive process with Poisson or geometric marginal

Zhengwei Liu[a], Qi Li[b] and Fukang Zhu[a]

[a]*Jilin University*
[b]*Changchun Normal University*

Abstract. To predict time series of counts with small values and remarkable fluctuations, an available model is the r states random environment process based on the negative binomial thinning operator and the geometric marginal. However, we argue that the aforementioned model may suffer from the following two drawbacks. First, under the condition of no prior information, the overdispersed property of the geometric distribution may cause the predictions fluctuate greatly. Second, because of the constraints on the model parameters, some estimated parameters are close to zero in real-data examples, which may not objectively reveal the correlation relationship. For the first drawback, an r states random environment process based on the binomial thinning operator and the Poisson marginal is introduced. For the second drawback, we propose a generalized r states random environment integer-valued autoregressive model based on the binomial thinning operator to model fluctuations of data. Yule–Walker and conditional maximum likelihood estimates are considered and their performances are assessed via simulation studies. Two real-data sets are conducted to illustrate the better performances of the proposed models compared with some existing models.

References

Al-Osh, M. A. and Alzaid, A. A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *Journal of Time Series Analysis* **8**, 261–275. MR0903755

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics* **28**, 89–110. MR0084903

Barczy, M., Ispány, M. and Pap, G. (2011). Asymptotic behavior of unstable INAR(p) processes. *Stochastic Processes and Their Applications* **121**, 583–608. MR2763097

Bu, R., McCabe, B. P. M. and Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis* **29**, 973–994. MR2464949

Drost, F. C., van den Akker, R. and Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR(p) models. *Journal of Time Series Analysis* **29**, 783–801. MR2450896

Drost, F. C., van den Akker, R. and Werker, B. J. M. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **71**, 467–485. MR2649605

Jazi, M. A., Jones, G. and Lai, C. D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis* **33**, 954–963. MR2991911

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* **33**, 664–677. MR1401464

Laketa, P. N., Nastić, A. S. and Ristić, M. M. (2018). Generalized random environment INAR models of higher order. *Mediterranean Journal of Mathematics* **15**. MR3740339 https://doi.org/10.1007/s00009-017-1054-z

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive processes. *Journal of Time Series Analysis* **4**, 439–455. MR1652193

Li, H., Yang, K., Zhao, S. and Wang, D. (2018). First-order random coefficients integer-valued threshold autoregressive processes. *AStA Advances in Statistical Analysis* **102**, 305–331. MR3829551

McCabe, B. P. M., Martin, G. M. and Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **73**, 253–272. MR2814495

Key words and phrases. Binomial thinning, INAR, random environment, time series of counts.

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric distributions. *Advances in Applied Probability* **18**, 679–705. MR0857325

Nastić, A. S., Laketa, P. N. and Ristić, M. M. (2016). Random environment integer-valued autoregressive process. *Journal of Time Series Analysis* **37**, 267–287. MR3511585

Nastić, A. S., Laketa, P. N. and Ristić, M. M. (2018). Random environment INAR models of higher order. *REVSTAT Statistical Journal*. To appear.

Pedeli, X., Davison, A. C. and Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association* **110**, 1229–1238. MR3420697

Qi, X., Li, Q. and Zhu, F. (2019). Modeling time series of count with excess zeros and ones based on INAR(1) model with zero-and-one inflated Poisson innovations. *Journal of Computational and Applied Mathematics* **346**, 572–590. MR3864182 https://doi.org/10.1016/j.cam.2018.07.043

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**, 893–899. MR0542141

Tang, M. and Wang, Y. (2014). Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. *Advances in Difference Equations* **2014**, 99.

Weiß, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. Chichester: John Wiley & Sons.

Zheng, H., Basawa, I. V. and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis* **27**, 411–440. MR2328539

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 273–290
https://doi.org/10.1214/19-BJPS429

Symmetrical and asymmetrical mixture autoregressive processes

Mohsen Maleki[a], Arezo Hajrajabi[b] and Reinaldo B. Arellano-Valle[c]

[a]*Shiraz University*
[b]*Imam Khomeini International University*
[c]*Universidad Católica de Chile*

Abstract. In this paper, we study the finite mixtures of autoregressive processes assuming that the distribution of innovations (errors) belongs to the class of scale mixture of skew-normal (SMSN) distributions. The SMSN distributions allow a simultaneous modeling of the existence of outliers, heavy tails and asymmetries in the distribution of innovations. Therefore, a statistical methodology based on the SMSN family allows us to use a robust modeling on some non-linear time series with great flexibility, to accommodate skewness, heavy tails and heterogeneity simultaneously. The existence of convenient hierarchical representations of the SMSN distributions facilitates also the implementation of an ECME-type of algorithm to perform the likelihood inference in the considered model. Simulation studies and the application to a real data set are finally presented to illustrate the usefulness of the proposed model.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723. MR0423716 https://doi.org/10.1109/tac.1974.1100705

Akinyemi, M. I. (2013). Mixture autoregressive models: Asymptotic properties and application to financial risk. PhD thesis, University of Manchester. MR3052080

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**, 99–102. MR0359122

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science* **3**, 415–438. MR2410041 https://doi.org/10.1080/02664760701236905

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* **34**, 663–682. MR2410041 https://doi.org/10.1080/02664760701236905

Arellano-Valle, R. B., Castro, L. M., Genton, M. G. and Gómez, H. W. (2008). Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis* **3**, 513–539. MR2434401 https://doi.org/10.1214/08-BA320

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. *IMS Monographs Series*. Cambridge: Cambridge University Press. MR3468021

Basso, R. M., Lachos, V. H., Cabral, C. R. B. and Ghosh, P. (2010). Robust mixture modeling based on the scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis* **54**, 2926–2941. MR2727724 https://doi.org/10.1016/j.csda.2009.09.031

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others*. Boca Raton: Chapman & Hall/CRC. MR1684363

Böhning, D., Hennig, C., McLachlan, G. J. and McNicholas, P. D. (2014). Editorial: The 2nd special issue on advances in mixture models. *Computational Statistics & Data Analysis* **71**, 1–2. MR3131949 https://doi.org/10.1016/j.csda.2013.10.010

Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V. and Walther, G. (2007). Editorial: Advances in mixture models. *Computational Statistics & Data Analysis* **51**, 5205–5210. MR2370865 https://doi.org/10.1016/j.csda.2006.10.025

Key words and phrases. ECME algorithm, non-linear time series, finite mixtures of autoregressive models, scale mixtures of skew-normal distributions.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113. MR1867257 https://doi.org/10.1006/jmva.2000.1960

Cervone, D., Pillai, N. S., Pati, D., Berbeco, R. and Lewis, J. H. (2014). A location-mixture autoregressive model for online. Forecasting of lung tumor motion. *Annals of Applied Statistics* **8**, 1341–1371. MR3271335 https://doi.org/10.1214/14-AOAS744

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological* **39**, 1–22. MR0501537

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer. MR2265601

Garay, A. M., Lachos, V. H. and Abanto-Valle, C. A. (2011). Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society* **40**, 115–124. MR2752409 https://doi.org/10.1016/j.jkss.2010.08.003

Ghasami, S., Khodadadi, Z. and Maleki, M. (2019). Autoregressive processes with generalized hyperbolic innovations. *Communications in Statistics Simulation and Computation*. https://doi.org/10.1080/03610918.2018.1535066

Glasbey, C. A. (2001). Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *Journal of the Royal Statistical Society Series C* **50**, 143–154. MR1833269 https://doi.org/10.1111/1467-9876.00225

Goldin, D. Q., Mardales, R. and Nagy, G. (2006). In search of meaning for time series subsequence clustering: Matching algorithms based on a new distance measure. In *Proc. of the ACM International Conference on Information and Knowledge Management*, 347–356.

Hajrajabi, A. and Maleki, M. (2019). Nonlinear semiparametric autoregressive model with finite mixtures of scale mixtures of skew normal innovations. *Journal of Applied Statistics* **46**(11), 2010–2029. https://doi.org/10.1080/02664763.2019.1575953

Hoseinzadeh, A., Maleki, M., Khodadadi, Z. and Contreras-Reyes, J. E. (2018). The Skew-Reflected-Gompertz distribution for analyzing the symmetric and asymmetric data. *Journal of Computational and Applied Mathematics* **349**, 132–141. MR3886701 https://doi.org/10.1016/j.cam.2018.09.011

Jin, S. and Li, W. K. (2006). Modeling panel time series with mixture autoregressive model. *Journal of Data Science* **4**, 425–446.

Lanne, M. and Saikkonen, P. (2003). Modeling the U.S. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics* **1**, 96–125.

Lau, J. and So, M. K. P. (2008). Bayesian mixture of autoregressive models. *Computational Statistics & Data Analysis* **53**, 38–60. MR2528591 https://doi.org/10.1016/j.csda.2008.06.001

Lindsay, B. G. (1995). *Mixture Models: Theory Geometry and Applications*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **51**. Harward: Institute of Mathematical Statistics.

Liu, C. and Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **80**, 267–278. MR1326414 https://doi.org/10.1093/biomet/81.4.633

Liu, M. and Lin, T. I. (2014). A skew-normal mixture regression model. *Educational and Psychological Measurement* **74**, 139–162.

Mahmoudi, M. R., Maleki, M. and Pak, A. (2017). Testing the difference between two independent time series models. *Iranian Journal of Science and Technology, Transaction A, Science* **41**, 665–669. MR3714808 https://doi.org/10.1007/s40995-017-0288-8

Maleki, M. and Arellano-Valle, R. B. (2017). Maximum a-posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *Journal of Statistical Computation and Simulation* **87**, 1061–1083. MR3606839 https://doi.org/10.1080/00949655.2016.1245305

Maleki, M., Arellano-Valle, R. B., Dey, D. K., Mahmoudi, M. R. and Jalali, S. M. J. (2017). A Bayesian approach to robust skewed autoregressive processes. *Calcutta Statistical Association Bulletin* **69**, 165–182. MR3746983 https://doi.org/10.1177/0008068317732196

Maleki, M. and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics Theory and Methods* **46**, 12356–12369. MR3740871 https://doi.org/10.1080/03610926.2017.1295160

Maleki, M. and Nematollahi, A. R. (2017a). Bayesian approach to epsilon-skew-normal family. *Communications in Statistics Theory and Methods* **46**, 7546–7561. MR3648583 https://doi.org/10.1080/03610926.2016.1157186

Maleki, M. and Nematollahi, A. R. (2017b). Autoregressive models with mixture of scale mixtures of Gaussian innovations. *Iranian Journal of Science and Technology Transaction A, Science* **41**, 1099–1107. MR3736837 https://doi.org/10.1007/s40995-017-0237-6

Maleki, M. and Wraith, D. (2019). Mixtures of multivariate restricted skew-normal factor analyzer models in a Bayesian framework. *Computational Statistics* **34**, 1039–1053. https://doi.org/10.1007/s00180-019-00870-6

Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2018a). Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Statistics and Computing* Available at https://doi.org/10.1007/s11222-018-9815-5. https://doi.org/10.1007/s11222-018-9815-5

Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2018b). A flexible class of parametric distributions for Bayesian linear mixed models. *Test* Available at https://doi.org/10.1007/s11749-018-0590-6. https://doi.org/10.1007/s11749-018-0590-6

McCulloch, R. E. and Tsay, R. S. (1994). Statistical inference of macroeconomic time series via Markov switching models. *Journal of Time Series Analysis* **15**, 523–539. MR1263893 https://doi.org/10.1111/j.1467-9892.1994.tb00188.x

McLachlan, G. J. and Krishnan, T. (2008). *The eM Algorithm and Extensions*. New Jersey: Wiley. MR2392878 https://doi.org/10.1002/9780470191613

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley. MR1789474 https://doi.org/10.1002/0471721182

Mengersen, K., Robert, C. P. and Titterington, D. M. (2011). *Mixtures: Estimation and Applications*. Chichester: Wiley. MR2867716 https://doi.org/10.1002/9781119995678

Moravveji, B., Khodadai, Z. and Maleki, M. (2018). A Bayesian analysis of two-piece distributions based on the scale mixtures of normal family. *Iranian Journal of Science and Technology Transaction A, Science* Available at https://doi.org/10.1007/s40995-018-0541-9. https://doi.org/10.1007/s40995-018-0541-9

Ni, H. and Yin, H. (2008). A self-organizing mixture autoregressive network for time series modeling and prediction. *Neurocomputing* **72**, 3529–3537.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Van Wijk, J. J. and Van Selow, E. R. (1999). Cluster and calendar-based visualization of time series data. In *Proceedings of the IEEE Symposium on Information Visualization*, 4–9.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.

Wong, C. S. (1998). Statistical inference for some non-linear time series models. PhD thesis, University of Hong.Kong. Hong Kong.

Wong, C. S. and Chan, W. S. (2009). A Student-*t* mixture autoregressive model with applications to heavy tailed financial data. *Biometrika* **96**, 751–760. MR2538770 https://doi.org/10.1093/biomet/asp031

Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Series B* **62**, 95–115. MR1747398 https://doi.org/10.1111/1467-9868.00222

Wong, C. S. and Li, W. K. (2001a). On a mixture autoregressive conditional heteroscedastic model. *Journal of American Statistical Association, Series B* **96**, 982–995. MR1946366 https://doi.org/10.1198/016214501753208645

Wong, C. S. and Li, W. K. (2001b). On a logistic mixture autoregressive model. *Biometrika* **88**, 833–846. MR1859413 https://doi.org/10.1093/biomet/88.3.833

Wood, S., Rosen, O. and Kohn, R. (2011). Bayesian mixtures of autoregressive models. *Journal of Computational and Graphical Statistics* **20**, 174–195. MR2816544 https://doi.org/10.1198/jcgs.2010.09174

Xiong, Y. and Yeung, D. (2004). Time series clustering with ARMA mixtures. *Pattern Recognition* **37**, 1675–1689.

Zarrin, P., Maleki, M., Khodadadi, Z. and Arellano-Valle, R. B. (2018). Time series process based on the unrestricted skew normal process. *Journal of Statistical Computation and Simulation* **89**, 38–51. MR3871128 https://doi.org/10.1080/00949655.2018.1533962

Zeller, C. B., Lachos, V. H. and Vilca-Labra, F. E. (2011). Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics* **38**, 348–363. MR2749591 https://doi.org/10.1080/02664760903406504

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 291–303
https://doi.org/10.1214/18-BJPS427

Adaptive two-treatment three-period crossover design for normal responses

Uttam Bandyopadhyay[a], Shirsendu Mukherjee[b] and Atanu Biswas[c]

[a]*University of Calcutta*
[b]*Asutosh College*
[c]*Indian Statistical Institute*

Abstract. In adaptive crossover design, our goal is to allocate more patients to a promising treatment sequence. The present work contains a very simple three period crossover design for two competing treatments where the allocation in period 3 is done on the basis of the data obtained from the first two periods. Assuming normality of response variables we use a reliability functional for the choice between two treatments. We calculate the allocation proportions and their standard errors corresponding to the possible treatment combinations. We also derive some asymptotic results and provide solutions on related inferential problems. Moreover, the proposed procedure is compared with a possible competitor. Finally, we use a data set to illustrate the applicability of the proposed design.

References

Balaam, L. N. (1968). A two period design with t^2 experimental units. *Biometrics* **24**, 61–73. MR0221678 https://doi.org/10.2307/2528460

Bandyopadhyay, U., Biswas, A. and Mukherjee, S. (2009). Adaptive two-treatment two-period crossover design for binary treatment responses incorporating carry-over effects. *Statistical Methods and Applications* **18**, 13–33. MR2476476 https://doi.org/10.1007/s10260-007-0072-6

Bandyopadhyay, U., Biswas, A. and Mukherjee, S. (2012). A response-adaptive design in crossover trial. *Statistics* **46**, 645–661. MR2974644 https://doi.org/10.1080/02331888.2010.545211

Bandyopadhyay, U. and Mukherjee, S. (2015). Adaptive crossover design for normal responses. *Communications in Statistics Theory and Methods* **44**, 1466–1482. MR3328226 https://doi.org/10.1080/03610926.2013.763509

Ebbutt, A. F. (1984). Three-period crossover designs for two treatments. *Biometrics* **40**, 219–224.

Fleiss, J. L. (1986). On multiperiod crossover studies (Letter to the editor). *Biometrics* **42**, 449–450.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*. New York: Academic Press. MR0229351

Jones, B. and Kenward, M. G. (2015). *Design and Analysis of Cross-Over Trials*, 3rd ed. London: Chapman and Hall/CRC Press. MR1014893

Kabaila, P. and Vicendese, M. (2012). The performance of a two-stage analysis of ABAB/BABA crossover trials. *Biometrical Journal* **54**, 361–369. MR2931079 https://doi.org/10.1002/bimj.201100201

Kushner, H. B. (2003). Allocation rules for adaptive repeated measurement design. *Journal of Statistical Planning and Inference* **113**, 293–313. MR1963048 https://doi.org/10.1016/S0378-3758(01)00310-X

Liang, Y. and Carriere, K. C. (2009). Multiple-objective response adaptive repeated measurement designs for clinical trials. *Journal of Statistical Planning and Inference* **139**, 1134–1145. MR2479855 https://doi.org/10.1016/j.jspi.2008.07.003

Liang, Y., Li, Y., Wang, J. and Carriere, K. C. (2014). Multiple-objective response-adaptive repeated measurement designs in clinical trials for binary responses. *Statistics in Medicine* **33**, 607–617. MR3153541 https://doi.org/10.1002/sim.5951

Matthews, J. N. S. (1989). Estimating dispersion parameters in the analysis of data from crossover trials. *Biometrika* **76**, 239–244.

Patterson, H. D. and Lucas, H. L. (1959). Extra-period change-over designs. *Biometrics* **15**, 116–132.

Senn, S. J. (1994). The AB/BA crossover: Past, present and future? *Statistical Methods in Medical Research* **3**, 303–324.

Key words and phrases. Adaptive crossover design, Balaam's design, carryover effect, ethical allocation, skewed allocation.

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 304–322
https://doi.org/10.1214/18-BJPS423

Bayesian modeling and prior sensitivity analysis for zero–one augmented beta regression models with an application to psychometric data

Danilo Covaes Nogarotto[a], Caio Lucidius Naberezny Azevedo[a] and Jorge Luis Bazán[b]

[a]*Campinas State University*
[b]*University of São Paulo*

Abstract. The interest on the analysis of the zero–one augmented beta regression (ZOABR) model has been increasing over the last few years. In this work, we developed a Bayesian inference for the ZOABR model, providing some contributions, namely: we explored the use of Jeffreys-rule and independence Jeffreys prior for some of the parameters, performing a sensitivity study of prior choice, comparing the Bayesian estimates with the maximum likelihood ones and measuring the accuracy of the estimates under several scenarios of interest. The results indicate, in a general way, that: the Bayesian approach, under the Jeffreys-rule prior, was as accurate as the ML one. Also, different from other approaches, we use the predictive distribution of the response to implement Bayesian residuals. To further illustrate the advantages of our approach, we conduct an analysis of a real psychometric data set including a Bayesian residual analysis, where it is shown that misleading inference can be obtained when the data is transformed. That is, when the zeros and ones are transformed to suitable values and the usual beta regression model is considered, instead of the ZOABR model. Finally, future developments are discussed.

References

Bayes, C. L. and Valdivieso, L. (2016). A beta inflated mean regression model for fractional response variables. *Journal of Applied Statistics* **43**, 1814–1830. MR3492148 https://doi.org/10.1080/02664763.2015.1120711

Bazan, J. L., Romeo, J. R. and Rodrigues, J. (2014). Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics* **28**, 467–482. MR3263060 https://doi.org/10.1214/13-BJPS218

Branscum, A. J., Johnson, W. O. and Thurmond, M. C. (2007). Bayesian beta regression: Application to household data and genetic distance between foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics* **49**, 287–301. MR2405396 https://doi.org/10.1111/j.1467-842X.2007.00481.x

Buckley, J. (2003). Estimation of models with beta-distributed dependent variables: A replication and extension of Paolinos study. *Political Analysis* **11**, 204–205.

Carlstrom, L., Woodward, J. and Palmer, C. (2000). Evaluating the simplified conjoint expected risk model: Comparing the use of objective and subjective information. *Risk Analysis* **20**, 385–392.

Cepeda-Cuervo, E., Jaimes, D., Marín, M. and Rojas, J. (2016). Bayesian beta regression with Bayesian beta reg R-package. *Computational Statistics* **31**, 165–187. MR3481800 https://doi.org/10.1007/s00180-015-0591-9

Cribari-Neto, F. and Zeiles, A. (2010). Beta regression in R. *Journal of Statistical Software* **34**, 1–24.

Dey, D. K., Ghosh, S. K. and Mallick, B. N. (2000). *Generalized Linear Models: A Bayesian Perspective*. MR1893779

Espinheira, P., Ferrari, S. L. P. and Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics* **35**, 407–419. MR2420486 https://doi.org/10.1080/02664760701834931

Ferrari, S. L. P. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31**, 799–815. MR2095753 https://doi.org/10.1080/0266476042000214501

Figueroa-Zuñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. P. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis* **61**, 137–147. MR3063006 https://doi.org/10.1016/j.csda.2012.12.002

Key words and phrases. Augmented beta regression, Bayesian inference, Jeffreys prior, MCMC methods, residual analysis.

Galvis, D. M., Bandyopadhyay, D. and Lachos, V. H. (2014). Augmented mixed beta regression models for periodontal proportion data. *Statistics in Medicine* **33**, 3759–3771. MR3260658 https://doi.org/10.1002/sim.6179

Gamerman, D. and Lopes, H. (2006). *Stochastic Simulation for Bayesian Inference*, 2nd ed. New York: Chapman & Hall/CRC.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247–253. MR1689229 https://doi.org/10.2307/2669699

Homan, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623. MR3214779

Kieschnick, R. and McCullough, B. D. (2003). Regression analysis of variates observed on $(0, 1)$: Percentages, proportions, and fractions. *Statistical Modelling* **3**, 193–213. MR2005473 https://doi.org/10.1191/1471082X03st053oa

Kim, S., Chen, M.-H. and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika* **95**, 93–106. MR2409717 https://doi.org/10.1093/biomet/asm079

Lemonte, A. J. and Bazán, J. L. (2016). New class of Johnson SB distributions and its associated regression model for rates and proportions. *Biometrical Journal* **58**, 727–746. MR3527412 https://doi.org/10.1002/bimj.201500030

Liu, F. and Kong, Y. (2015). zoib: An R package for Bayesian inference for beta regression and zero/one inflated beta regression. *The R Journal* **7**, 34–51.

López, F. O. (2013). A Bayesian approach to parameter estimation in simplex regression model: A comparison with beta regression. *Revista Colombiana de Estadística* **36**, 1–21. MR3075189

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Economics*. New York: Cambridge University Press. MR0799154 https://doi.org/10.1017/CBO9780511810176

Nogarotto, D. C. (2013). Bayesian inference EM beta and inflated beta regression model. Master's dissertation (in Portuguese), IMECC-Unicamp.

Nogarotto, D. C., Azevedo, C. L. N. and Bazán, J. L. (2020). Supplement to "Bayesian modeling and prior sensitivity analysis for zero–one augmented beta regression models with an application to psychometric data." https://doi.org/10.1214/18-BJPS423SUPP

Oliveira, M. S. (2004). A beta regression model: Theory and application. Master's dissertation (in Portuguese), IME-USP.

Ospina, R. (2008). Inflated beta regression modelo. Doctoral's thesis (in Portuguese), IME-USP.

Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers* **51**, 111–126. MR2556590 https://doi.org/10.1007/s00362-008-0125-4

Ospina, R. and Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis* **56**, 1609–1623. MR2892364 https://doi.org/10.1016/j.csda.2011.10.005

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis* **9**, 325–346.

Papke, L. and Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics* **11**, 619–632.

Parker, A. J., Bandyopadhyay, D. and Slate, E. H. (2014). A spatial augmented beta regression model for periodontal proportion data. *Statistical Modelling* **14**, 503–521. MR3284553 https://doi.org/10.1177/1471082X14535515

Paulino, C. D., Turkman, M. A. A. and Murteira, B. (2003). *Bayesian Statistic* (in Portuguese). Lisboa: Fundação Calouste Gulbenkian.

Pereira, T. L. and Cribari-Neto, F. (2014). Detecting model misspecification in inflated beta regressions. *Communications in Statistics Simulation and Computation* **43**, 631–656. MR3200996 https://doi.org/10.1080/03610918.2012.712183

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at http://www.R-project.org.

Ramalho, J. J. S. and Silva, J. V. (2009). A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance* **9**, 621–636. MR2548100 https://doi.org/10.1080/14697680802448777

Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **71**, 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* **11**, 54–71.

Souza, T. C., Pereira, T. L., Cribari-Neto, F. and Lima, M. C. V. (2016). Testing inference in inflated beta regressions under model misspecification. *Communications in Statistics Simulation and Computation* **45**, 625–642. MR3457110 https://doi.org/10.1080/03610918.2013.867995

Stavrunova, O. and Yerokhin, O. (2012). Two-part fractional regression model for the demand for risky assets. *Applied Economics* **44**, 21–26.

Swearingen, C. J., Castro, M. S. M. and Bursac, Z. (2012). Inflated beta regression: Zero, one, and everything in between. In *SAS Global Forum 2012*. Available at http://support.sas.com/resources/papers/proceedings12/325-2012.pdf.

Wieczorek, J. and Hawala, S. (2011). A Bayesian zero–one inflated beta model for estimating poverty in U.S. counties. In *Proceedings of the American Statistical Association. Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 323–344
https://doi.org/10.1214/18-BJPS425

A Bayesian sparse finite mixture model for clustering data from a heterogeneous population

Erlandson F. Saraiva[a], **Adriano K. Suzuki**[b] **and Luís A. Milan**[c]

[a]*Universidade Federal de Mato Grosso do Sul*
[b]*Unversidade de São Paulo*
[c]*Universidade Federal de São Carlos*

Abstract. In this paper, we introduce a Bayesian approach for clustering data using a sparse finite mixture model (SFMM). The SFMM is a finite mixture model with a large number of components k previously fixed where many components can be empty. In this model, the number of components k can be interpreted as the maximum number of distinct mixture components. Then, we explore the use of a prior distribution for the weights of the mixture model that take into account the possibility that the number of clusters $k_{\mathbf{c}}$ (e.g., nonempty components) can be random and smaller than the number of components k of the finite mixture model. In order to determine clusters we develop a MCMC algorithm denominated Split-Merge allocation sampler. In this algorithm, the split-merge strategy is data-driven and was inserted within the algorithm in order to increase the mixing of the Markov chain in relation to the number of clusters. The performance of the method is verified using simulated datasets and three real datasets. The first real data set is the benchmark galaxy data, while second and third are the publicly available data set on Enzyme and Acidity, respectively.

References

Akaike, H. A. (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723. MR0423716 https://doi.org/10.1109/tac.1974.1100705

Anderson, J. J. (1985). Normal mixtures and the number of clusters problem. *Computational Statistics Quarterly* **2**, 3–14.

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803–821. MR1243494 https://doi.org/10.2307/2532201

Bensmail, H., Celeux, G., Raftery, A. E. and Robert, C. P. (1997). Inference in model-based cluster analysis. *Statistics and Computing* **7**, 1–10.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika* **65**, 31–38. MR0501592 https://doi.org/10.1093/biomet/65.1.31

Bouveyron, C. and Brunet, C. (2013). Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis* **71**, 52–78. MR3131954 https://doi.org/10.1016/j.csda.2012.12.008

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

Casella, G., Robert, C. and Wells, M. (2000). Mixture models, latent variables and partitioned importance sampling. Technical Report-2000-03, CREST, INSEE, Paris.

Celeux, G., Hurn, M. and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association* **95**, 957–970. MR1804450 https://doi.org/10.2307/2669477

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**, 327–335.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588. MR1340510

Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**. MR1951635 https://doi.org/10.1198/016214502760047131

Key words and phrases. Mixture model, Bayesian approach, Gibbs sampling, Metropolis–Hastings, split-merge update.

Fruhwirth-Schnatter, S. (2017). From here to infinity-sparse finite versus Dirichlet process mixture in model-based clustering. https://arxiv.org/abs/1706.07194.

Hartigan, J. A. and Wong, M. A. (1978). Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics* **28**, 100–108. MR0405726

Jasra, A., Holmes, C. C. and Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science* **20**, 50–67. MR2182987 https://doi.org/10.1214/088342305000000016

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281–297. Berkeley, CA: University of California Press. MR0214227

McLachlan, G. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker. MR0926484

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley Interscience. MR1789474 https://doi.org/10.1002/0471721182

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

Oh, M.-S. and Raftery, A. E. (2007). Model-based clustering with dissimilarities: A Bayesian approach. *Journal of Computational and Graphical Statistics* **16**, 559–585. MR2351080 https://doi.org/10.1198/106186007X236127

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **59**, 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixture of normals. *Journal of the American Statistical Association* **92**, 894–902. MR1482121 https://doi.org/10.2307/2965553

Saraiva, E. F., Louzada, F. and Milan, L. A. (2014). Mixture models with an unknown number of components via a new posterior split–merge MCMC algorithm. *Applied Mathematics and Computation* **244**, 959–975. MR3250635 https://doi.org/10.1016/j.amc.2014.07.032

Saraiva, E. F., Suzuki, A. K., Louzada, F. and Milan, L. A. (2016). Partitioning gene expression data by data-driven Markov chain Monte Carlo. *Journal of Applied Statistics* **43**, 1155–1173. MR3460559 https://doi.org/10.1080/02664763.2015.1092113

Saraiva, E. F., Suzuki, A. K. and Milan, L. A. (2019). Supplement to "A Bayesian sparse finite mixture model for clustering data from a heterogeneous population." https://doi.org/10.1214/18-BJPS425SUPP.

Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Sneath, P. H. A. (1957). The application of computers to taxonomy. *Journal of General Microbiology* **17**, 201–206.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* **38**, 1409–1438.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–616. MR1979380 https://doi.org/10.1111/1467-9868.00353

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **62**, 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

Walli, G. M., Frhwirth-Schnatter, S. and Grn, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing* **34**, 303–324. MR3439375 https://doi.org/10.1007/s11222-014-9500-2

Ward, J. H. (1963). Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association* **58**, 234–244. MR0148188

Witten, D. M. and Tibshirani, R. (2010). A framework for feature selection in clustering. *Journal of the American Statistical Association* **105**, 713–726. MR2724855 https://doi.org/10.1198/jasa.2010.tm09415

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 345–369
https://doi.org/10.1214/18-BJPS426

Reliability estimation in a multicomponent stress-strength model for Burr XII distribution under progressive censoring

Raj Kamal Maurya[a] **and Yogesh Mani Tripathi**[a]

[a]*Indian Institute of Technology Patna*

Abstract. We consider estimation of the multicomponent stress-strength reliability under progressive Type II censoring under the assumption that stress and strength variables follow Burr XII distributions with a common shape parameter. Maximum likelihood estimates of the reliability are obtained along with asymptotic intervals when common shape parameter may be known or unknown. Bayes estimates are also derived under the squared error loss function using different approximation methods. Further, we obtain exact Bayes and uniformly minimum variance unbiased estimates of the reliability for the case common shape parameter is known. The highest posterior density intervals are also obtained. We perform Monte Carlo simulations to compare the performance of proposed estimates and present a discussion based on this study. Finally, two real data sets are analyzed for illustration purposes.

References

Aslam, M., Azam, M. and Jun, C.-H. (2016). Multiple dependent state repetitive group sampling plan for Burr XII distribution. *Quality Engineering* **28**, 231–237.

Awad, A. M. and Gharraf, M. K. (1986). Estimation of $P(Y < X)$ in the Burr case: A comparative study. *Communications in Statistics Simulation and Computation* **15**, 389–403. MR0848181 https://doi.org/10.1080/03610918608812514

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Springer. MR1763020 https://doi.org/10.1007/978-1-4612-1334-5

Balakrishnan, N. and Sandhu, R. A. (1995). A simple simulational algorithm for generating progressive Type-II censored samples. *American Statistician* **49**, 229–230.

Basirat, M., Baratpour, S. and Ahmadi, J. (2015). Statistical inferences for stress–strength in the proportional hazard models based on progressive Type-II censored samples. *Journal of Statistical Computation and Simulation* **85**, 431–449. MR3275457 https://doi.org/10.1080/00949655.2013.824449

Bhattacharyya, G. K. and Johnson, R. A. (1974). Estimation of reliability in a multicomponent stress-strength model. *Journal of the American Statistical Association* **69**, 966–970. MR0443269

Birnbaum, Z. W., et al (1956). On a use of the Mann–Whitney statistic. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. MR0084909

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics* **13**, 215–232. MR0006644 https://doi.org/10.1214/aoms/1177731607

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics* **8**, 69–92. MR1705909 https://doi.org/10.2307/1390921

Church, J. D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships. *Technometrics* **12**, 49–54.

Dey, S., Mazucheli, J. and Anis, M. Z. (2017). Estimation of reliability of multicomponent stress–strength for a Kumaraswamy distribution. *Communications in Statistics Theory and Methods* **46**, 1560–1572. MR3574712 https://doi.org/10.1080/03610926.2015.1022457

Downtown, F. (1973). The estimation of $P(Y < X)$ in the normal case. *Journal of the American Statistical Association* **15**, 551–558. MR0334377

Draper, N. R. and Guttman, I. (1978). Bayesian analysis of reliability in multicomponent stress-strength models. *Communications in Statistics Theory and Methods* **7**, 441–451. MR0514232 https://doi.org/10.1080/03610927808827637

Key words and phrases. Bayes estimate, maximum likelihood estimate, multicomponent reliability, progressive censoring, uniformly minimum variance unbiased estimator.

Enis, P. and Geisser, S. (1971). Estimation of the probability that $Y < X$. *Journal of the American Statistical Association* **66**, 162–168.

Gunasekera, S. (2015). Generalized inferences of $R = \Pr(X > Y)$ for Pareto distribution. *Statistical Papers* **56**, 333–351. MR3331866 https://doi.org/10.1007/s00362-014-0584-8

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

Kızılaslan, F. (2017). Classical and Bayesian estimation of reliability in a multicomponent stress–strength model based on the proportional reversed hazard rate mode. *Mathematics and Computers in Simulation* **136**, 36–62. MR3610080 https://doi.org/10.1016/j.matcom.2016.10.011

Kızılaslan, F. and Nadar, M. (2016). Estimation of reliability in a multicomponent stress–strength model based on a bivariate Kumaraswamy distribution. *Statistical Papers*, 1–34. MR3765947 https://doi.org/10.1007/s00362-016-0765-8

Kohansal, A. (2017). On estimation of reliability in a multicomponent stress-strength model for a Kumaraswamy distribution based on progressively censored sample. *Statistical Papers*, 1–40. MR4027641 https://doi.org/10.1007/s00362-017-0916-6

Lindley, D. V. (1980). Approximate Bayesian methods. *Trabajos de Estadística Y de Investigación Operativa* **31**, 223–245.

Lyu, M. R., et al (1996). *Handbook of Software Reliability Engineering, Vol. 222*. IEEE Computer Society Press CA.

Maurya, R. K., Tripathi, Y. M., Rastogi, M. K. and Asgharzadeh, A. (2017). Parameter estimation for a Burr XII distribution under progressive censoring. *American Journal of Mathematical and Management Sciences* **36**, 259–276.

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.

Moore, D. and Papadopoulos, A. S. (2000). The Burr type XII distribution as a failure model under various loss functions. *Microelectronics and Reliability* **40**, 2117–2122.

Nadar, M. and Kızılaslan, F. (2016). Estimation of reliability in a multicomponent stress-strength model based on a Marshall–Olkin bivariate Weibull distribution. *IEEE Transactions on Reliability* **65**, 370–380. MR3374192 https://doi.org/10.15446/rce.v38n2.51674

Nadar, M., Kızılaslan, F. and Papadopoulos, A. (2014). Classical and Bayesian estimation of $P(Y < X)$ for Kumaraswamy's distribution. *Journal of Statistical Computation and Simulation* **84**, 1505–1529. MR3196018 https://doi.org/10.1080/00949655.2012.750658

Najarzadegan, H., Babaii, S., Rezaei, S. and Nadarajah, S. (2016). Estimation of $p(y < x)$ for the Levy distribution. *Hacettepe Bulletin of Natural Sciences and Engineering* **45**, 957–972. MR3588235 https://doi.org/10.15672/hjms.2015599877

Papadopoulos, A. S. (1978). The Burr distribution as a failure model from a Bayesian approach. *IEEE Transactions on Reliability* **27**, 369–371.

Rao, G. S., Aslam, M. and Kundu, D. (2015). Burr-XII distribution parametric estimation and estimation of reliability of multicomponent stress-strength. *Communications in Statistics Theory and Methods* **44**, 4953–4961. MR3429812 https://doi.org/10.1080/03610926.2013.821490

Raqab, M. Z. and Kundu, D. (2005). Comparison of different estimators of $P(Y < X)$ for a scaled Burr type X distribution. *Communications in Statistics Simulation and Computation* **34**, 465–483. MR2157322 https://doi.org/10.1081/SAC-200055741

Rastogi, M. K. and Tripathi, Y. M. (2012). Estimating the parameters of a Burr distribution under progressive type II censoring. *Statistical Methodology* **9**, 381–391. MR2871439 https://doi.org/10.1016/j.stamet.2011.10.002

Sinha, S. K. (1986). Bayesian estimation of the reliability function of the Inverse Gaussian distribution. *Statistics & Probability Letters* **4**, 319–323. MR0858325 https://doi.org/10.1016/0167-7152(86)90052-0

Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using Burr-XII model. *IEEE Transactions on Reliability* **54**, 34–42.

Surles, J. G. and Padgett, W. J. (1998). Inference for $P(Y < X)$ in the Burr type X model. *Journal of Applied Statistical Science* **7**, 225–238.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86. MR0830567 https://doi.org/10.1007/s00362-017-0916-6

Wingo, D. R. (1993). Maximum likelihood methods for fitting the Burr type XII distribution to multiply (progressively) censored life test data. *Metrika* **40**, 203–210. MR1235082 https://doi.org/10.1007/BF02613681

Zimmer, W. J., Keats, J. B. and Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology* **30**, 386.

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 370–393
https://doi.org/10.1214/19-BJPS432

Measuring symmetry and asymmetry of multiplicative distortion measurement errors data

Jun Zhang[a], Yujie Gai[b], Xia Cui[c] and Gaorong Li[d]

[a]*Shenzhen University*
[b]*Central University of Finance and Economics*
[c]*Guangzhou University*
[d]*Beijing Normal University*

Abstract. This paper studies the measure of symmetry or asymmetry of a continuous variable under the multiplicative distortion measurement errors setting. The unobservable variable is distorted in a multiplicative fashion by an observed confounding variable. First, two direct plug-in estimation procedures are proposed, and the empirical likelihood based confidence intervals are constructed to measure the symmetry or asymmetry of the unobserved variable. Next, we propose four test statistics for testing whether the unobserved variable is symmetric or not. The asymptotic properties of the proposed estimators and test statistics are examined. We conduct Monte Carlo simulation experiments to examine the performance of the proposed estimators and test statistics. These methods are applied to analyze a real dataset for an illustration.

References

Butler, C. C. (1969). A test for symmetry using the sample distribution function. *The Annals of Mathematical Statistics* **40**, 2209–2210.

Cabilio, J. and Cabilio, P. (1996). A simple test of symmetry about an unknown median. *Canadian Journal of Statistics* **24**, 349–361.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Nonlinear Measurement Error Models: A Modern Perspective*, 2nd ed. New York: Chapman and Hall. MR2243417 https://doi.org/10.1201/9781420010138

Chen, K., Guo, S., Lin, Y. and Ying, Z. (2010). Least absolute relative error estimation. *Journal of the American Statistical Association* **105**, 1104–1112. MR2752606 https://doi.org/10.1198/jasa.2010.tm09307

Chen, K., Lin, Y., Wang, Z. and Ying, Z. (2016). Least product relative error estimation. *Journal of Multivariate Analysis* **144**, 91–98. MR3434942 https://doi.org/10.1016/j.jmva.2015.10.017

Csörgő, S. and Heathcote, C. R. (1987). Testing for symmetry. *Biometrika* **74**, 177–184. MR0885930 https://doi.org/10.2307/2336032

Cui, X., Guo, W., Lin, L. and Zhu, L. (2009). Covariate-adjusted nonlinear regression. *The Annals of Statistics* **37**, 1839–1870. MR2533473 https://doi.org/10.1214/08-AOS627

Davis, C. E. and Quade, D. (1978). U-statistics for skewness or symmetry. *Communication in Statistics: Theory and Methods* **7**, 413–418.

Delaigle, A., Hall, P. and Zhou, W.-X. (2016). Nonparametric covariate-adjusted regression. *The Annals of Statistics* **44**, 2190–2220. MR3546448 https://doi.org/10.1214/16-AOS1442

Fuller, W. A. (1987). *Measurement Error Models*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: John Wiley & Sons Inc. MR0898653 https://doi.org/10.1002/9780470316665

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)* **33**, 391–399.

Gupta, M. K. (1967). An asymptotically nonparametric test of symmetry. *The Annals of Mathematical Statistics* **38**, 849–866. MR0208761 https://doi.org/10.1214/aoms/1177698879

Hill, D. L. and Rao, P. V. (1977). Tests of symmetry based on Cramér–von Mises statistics. *Biometrika* **64**, 489–494. MR0518893 https://doi.org/10.1093/biomet/64.3.489

Key words and phrases. Confounding variable, errors-in-variables, correlation coefficient, empirical likelihood, symmetry.

Hotelling, H. and Solomons, L. M. (1932). The limits of a measure of skewness. *The Annals of Mathematical Statistics* **3**, 141–142.

Hwang, J. T. (1986). Multiplicative errors-in-variables models with applications to recent data released by the U.S. Department of Energy. *Journal of the American Statistical Association* **81**, 680–688. MR0860499

Kiwitt, S. and Neumeyer, N. (2012). Estimating the conditional error distribution in non-parametric regression. *Scandinavian Journal of Statistics* **39**, 259–281. MR2927025 https://doi.org/10.1111/j.1467-9469.2011.00763.x

Kraft, C. H. and van Eeden, C. (1972). Linearized rank estimates and signed-rank estimates for the general linear hypothesis. *The Annals of Mathematical Statistics* **43**, 42–57. MR0301857 https://doi.org/10.1214/aoms/1177692699

Li, F., Lin, L. and Cui, X. (2010). Covariate-adjusted partially linear regression models. *Communications in Statistics: Theory and Methods* **39**, 1054–1074. MR2745361 https://doi.org/10.1080/03610920902846539

Li, F., Lin, L., Lu, Y. and Feng, S. (2018). An adaptive estimation for covariate-adjusted nonparametric regression model. *Statistical Papers*. Available at https://doi.org/10.1007/s00362-019-01084-0. https://doi.org/10.1007/s00362-019-01084-0

Li, F. and Lu, Y. (2018). Lasso-type estimation for covariate-adjusted linear model. *Journal of Applied Statistics* **45**, 26–42. MR3736856 https://doi.org/10.1080/02664763.2016.1267121

Li, G., Zhang, J. and Feng, S. (2016). *Modern Measurement Error Models*. Beijing: Science Press.

Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27**, 1519–1535. MR1742498 https://doi.org/10.1214/aos/1017939140

Liang, H. and Ren, H. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics* **14**, 237–250. MR2137900 https://doi.org/10.1198/106186005X37481

Liu, H. and Xia, X. (2018). Estimation and empirical likelihood for single-index multiplicative models. *Journal of Statistical Planning and Inference* **193**, 70–88. MR3713464 https://doi.org/10.1016/j.jspi.2017.08.003

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics* **19**, 1725–1747. MR1135146 https://doi.org/10.1214/aos/1176348368

Partlett, C. and Patil, P. (2017). Measuring asymmetry and testing symmetry. *Annals of the Institute of Statistical Mathematics* **69**, 429–460. MR3611527 https://doi.org/10.1007/s10463-015-0547-4

Patil, P. N., Patil, P. P. and Bagkavos, D. (2012). A measure of asymmetry. *Statistical Papers* **53**, 971–985. MR2992928 https://doi.org/10.1007/s00362-011-0401-6

Rosman, J. B., ter Wee, P. M., Meijer, S., Piers-Becht, T. P. M., Sluiter, W. J. and Donker, A. J. M. (1984). Prospective randomized trial of early protein restriction in chronic renal failure. *The Lancet* **2**, 1291–1296.

Şentürk, D. and Müller, H.-G. (2005). Covariate adjusted correlation analysis via varying coefficient models. *Scandinavian Journal of Statistics: Theory and Applications* **32**, 365–383. MR2204625 https://doi.org/10.1111/j.1467-9469.2005.00450.x

Şentürk, D. and Müller, H.-G. (2006). Inference for covariate adjusted regression via varying coefficient models. *The Annals of Statistics* **34**, 654–679. MR2281880 https://doi.org/10.1214/009053606000000083

Zhang, J., Gai, J., Cui, X. and Li, G. (2019). Supplement to "Measuring symmetry and asymmetry of multiplicative distortion measurement errors data." https://doi.org/10.1214/19-BJPS432SUPP.

Zhao, J. and Xie, C. (2018). A nonparametric test for covariate-adjusted models. *Statistics & Probability Letters* **133**, 65–70. MR3732354 https://doi.org/10.1016/j.spl.2017.10.004

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 394–413
https://doi.org/10.1214/18-BJPS420

Stein characterizations for linear combinations of gamma random variables

Benjamin Arras[a], Ehsan Azmoodeh[b], Guillaume Poly[c] and Yvik Swan[d]

[a]*Université Pierre et Marie Curie*
[b]*Ruhr University Bochum*
[c]*Université de Rennes 1*
[d]*Université de Liège*

Abstract. In this paper we propose a new, simple and explicit mechanism allowing to derive Stein operators for random variables whose characteristic function satisfies a simple ODE. We apply this to study random variables which can be represented as linear combinations of (not necessarily independent) gamma distributed random variables. The connection with Malliavin calculus for random variables in the second Wiener chaos is detailed. An application to McKay Type I random variables is also outlined.

References

Arras, B., Azmoodeh, E., Poly, G. and Swan, Y. (2019). A bound on the Wasserstein-2 distance between linear combinations of independent random variables. *Stochastic Processes and Their Applications* **129**(7), 2341–2375.

Azmoodeh, A., Campese, S. and Poly, G. (2014). Fourth moment theorems for Markov diffusion generators. *Journal of Functional Analysis* **266**, 2341–2359. MR3150163 https://doi.org/10.1016/j.jfa.2013.10.014

Azmoodeh, E., Malicet, D., Mijoule, G. and Poly, G. (2016). Generalization of the Nualart–Peccati criterion. *Annals of Probability* **44**, 924–954. MR3474463 https://doi.org/10.1214/14-AOP992

Azmoodeh, E., Peccati, G. and Poly, G. (2014). Convergence towards linear combinations of chi-squared random variables: A Malliavin-based approach. In *Séminaire de Probabilités XLVII (Special Volume in Memory of Marc Yor)*, 339–367. MR3444306 https://doi.org/10.1007/978-3-319-18585-9_16

Bapat, R. B. (1989). Infinite divisibility of multivariate gamma distributions and M-matrices. *Sankhya Series A* **51**, 73–78. MR1065560

Barbour, A. D. (1990). Stein's method for diffusion approximations. *Probability Theory and Related Fields* **84**, 297–322. MR1035659 https://doi.org/10.1007/BF01197887

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford: Clarendon Press. MR1163825

Chen, L. H. Y., Goldstein, L. and Shao, Q. M. (2010). *Normal Approximation by Stein's Method*. Springer. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

Döbler, C. (2015). Stein's method of exchangeable pairs for the Beta distribution and generalizations. *Electronic Journal of Probability* **20**, 1–34. MR3418541 https://doi.org/10.1214/EJP.v20-3933

Döbler, C. and Peccati, G. (2018). The gamma Stein equation and noncentral de Jong theorems. *Bernoulli* **24**, 3384–3421. MR3788176 https://doi.org/10.3150/17-BEJ963

Eden, R. and Viquez, J. (2015). Nourdin–Peccati analysis on Wiener and Wiener–Poisson space for general distributions. *Stochastic Processes and Their Applications* **125**, 182–216. MR3274696 https://doi.org/10.1016/j.spa.2014.09.001

Eichelsbacher, P. and Thäle, C. (2014). Malliavin–Stein method for variance-Gamma approximation on Wiener space. *Electronic Journal of Probability* **20**, 1–28. MR3425543 https://doi.org/10.1214/EJP.v20-4136

Eisenbaum, N. and Kaspi, H. (2006). A characterization of the infinitely divisible squared Gaussian processes. *Annals of Probability* **34**, 728–742. MR2223956 https://doi.org/10.1214/009117905000000684

Eisenbaum, N. and Kaspi, H. (2009). On permanental processes. *Stochastic Processes and Their Applications* **119**, 1401–1415. MR2513113 https://doi.org/10.1016/j.spa.2008.07.003

Gaunt, R. E. (2013). Rates of convergence of variance-Gamma approximations via Stein's method. MR3194737 https://doi.org/10.1214/EJP.v19-3020

Key words and phrases. Stein's method, second Wiener chaos, multivariate gamma distribution, McKay distribution.

Gaunt, R. E. (2014). Variance-Gamma approximation via Stein's method. *Electronic Journal of Probability* **19**, 38, 1–33. MR3194737 https://doi.org/10.1214/EJP.v19-3020

Gaunt, R. E. (2017). On Stein's method for products of normal random variables and zero bias couplings. *Bernoulli* **23**, 3311–3345. MR3654808 https://doi.org/10.3150/16-BEJ848

Gaunt, R. E. (2018). Products of normal, beta and gamma random variables: Stein characterisations and distributional theory. *Brazilian Journal of Probability and Statistics* **32**, 437–466. MR3787761 https://doi.org/10.1214/16-BJPS349

Gaunt, R. E., Mijoule, G. and Swan, Y. (2019). An algebra of Stein operators. *Journal of Mathematical Analysis and Applications* **469**(1), 260–279.

Griffiths, R. C. (1984). Characterization of infinitely divisible multivariate gamma distributions. *Journal of Multivariate Analysis* **15**, 13–20. MR0755813 https://doi.org/10.1016/0047-259X(84)90064-2

Holm, H. and Alouini, M. S. (2004). Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution. *IEEE Transactions on Communications* **52**, 1367–1376.

Krishnamoorthy, A. S. and Parthasarathy, M. (1951). A multivariate gamma-type distribution. *The Annals of Mathematical Statistics* **22**, 549–557. MR0044790

Kusuoka, S. and Tudor, C. A. (2012). Stein's method for invariant measures of diffusions via Malliavin calculus. *Stochastic Processes and Their Applications* **122**, 1627–1651. MR2914766 https://doi.org/10.1016/j.spa.2012.02.005

Ley, C., Reinert, G. and Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probability Surveys* **14**, 1–52. MR3595350 https://doi.org/10.1214/16-PS278

Ley, C. and Swan, Y. (2013). Local Pinsker inequalities via Stein's discrete density approach. *IEEE Transactions on Information Theory* **59**, 5584–5591. MR3096942 https://doi.org/10.1109/TIT.2013.2265392

Loeve, M. (1977). *Probability Theory, I*, 4th ed. Springer. MR0651017

Luk, H. M. (1994). Stein's method for the gamma distribution and related statistical applications. ProQuest. MR2693204

Nourdin, I. and Peccati, G. (2009a). Noncentral convergence of multiple integrals. *Annals of Probability* **37**, 1412–1426. MR2546749 https://doi.org/10.1214/08-AOP435

Nourdin, I. and Peccati, G. (2009b). Stein's method on Wiener chaos. *Probability Theory and Related Fields* **145**, 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

Nourdin, I. and Peccati, G. (2010). Cumulants on the Wiener space. *Journal of Functional Analysis* **258**, 3775–3791. MR2606872 https://doi.org/10.1016/j.jfa.2009.10.024

Nourdin, I. and Peccati, G. (2012). *Normal Approximations Using Malliavin Calculus: From Stein's Method to Universality*. *Cambridge Tracts in Mathematics*. Cambridge University. MR2962301 https://doi.org/10.1017/CBO9781139084659

Nourdin, I. and Poly, G. (2012). Convergence in law in the second Wiener/Wigner chaos. *Electronic Communications in Probability* **17**, 1–12. MR2970700 https://doi.org/10.1214/ecp.v17-2023

Peköz, E., Röllin, A. and Ross, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *The Annals of Applied Probability* **23**, 1188–1218. MR3076682 https://doi.org/10.1214/12-AAP868

Pickett, A. (2004). Rates of convergence of Chi-square approximations via Stein's method. (Ph.D.)-Thesis, University of Oxford.

Royen, T. (2016). A note on the existence of the multivariate gamma distribution. arXiv preprint. Available at arXiv:1606.04747.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, II*, 583–602. MR0402873

Stein, C. (1986). *Approximate Computation of Expectations*. *Lecture Notes-Monograph Series* **7**. IMS. MR0882007

Teschl, G. (2012). *Ordinary Differential Equations and Dynamical Systems*. *Graduate Studies in Mathematics* **140**. Providence: Amer. Math. Soc. MR2961944 https://doi.org/10.1090/gsm/140

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand Journal of Mathematics* **26**, 125–149. MR1450811

Zolotarev, V. M. (1983). Probability metrics. *Teoriâ Veroâtnostej I Ee Primeneniâ* **28**, 264–287. MR0700210

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 414–425
https://doi.org/10.1214/18-BJPS422

Oriented first passage percolation in the mean field limit

Nicola Kistler, Adrien Schertzer and Marius A. Schmidt

Goethe-Universität

Abstract. The *Poisson clumping heuristic* has lead Aldous to conjecture the value of the oriented first passage percolation on the hypercube in the limit of large dimensions. Aldous' conjecture has been rigorously confirmed by Fill and Pemantle (*Ann. Appl. Probab.* **3** (1993) 593–629) by means of a variance reduction trick. We present here a streamlined and, we believe, more natural proof based on ideas emerged in the study of Derrida's random energy models.

References

Aldous, D. (2013). *Probability Approximations Via the Poisson Clumping Heuristic*. *Applied Mathematical Sciences* **77**. New York: Springer. MR0969362 https://doi.org/10.1007/978-1-4757-6283-9

Arguin, L.-P. (2016). Extrema of log-correlated random variables: Principles and examples. In *Advances in Disordered Systems, Random Processes and Some Applications* (P. Contucci and C. Giardiná, eds.). Cambridge: Cambridge Univiersity Press. MR3644278

Bolthausen, E. and Kistler, N. (2006). On a nonhierarchical version of the generalized random energy model. *The Annals of Applied Probability* **16**, 1–14. MR2209333 https://doi.org/10.1214/105051605000000665

Bovier, A. and Kurkova, I. (2009). A short course on mean field spin glasses. In *Spin Glasses: Statics and Dynamics. Summer School Paris, 2007* (A. Boutet de Monvel and A. Bovier, eds.). Basel–Boston–Berlin: Birkhäuser. MR2761979 https://doi.org/10.1007/978-3-7643-9891-0_1

Derrida, B. (1981). Random-energy model: An exactly solvable model of disordered systems. *Physical Review B* **24**, 2613–2626. MR0627810 https://doi.org/10.1103/physrevb.24.2613

Derrida, B. (1985). A generalization of the random energy model which includes correlations between energies. *Journal de Physique Lettres* **46**, 401–407.

Fill, J. A. and Pemantle, R. (1993). Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *The Annals of Applied Probability* **3**, 593–629. MR1221168

Hegarty, P. and Martinsson, A. (2014). On the existence of accessible paths in various models of fitness landscapes. *The Annals of Applied Probability* **24**, 1375–1395. MR3210999 https://doi.org/10.1214/13-AAP949

Kistler, N. (2015). Derrida's random energy models. From spin glasses to the extremes of correlated radom fields. In *Correlated Random Systems: Five Different Methods* (V. Gayrard and N. Kistler, eds.), *Lecture Notes in Mathematics* **2143**. Cham: Springer. MR3380419 https://doi.org/10.1007/978-3-319-17674-1_3

Martinsson, A. (2016). Unoriented first-passage percolation on the n-cube. *The Annals of Applied Probability* **26**, 2597–2625. MR3563188 https://doi.org/10.1214/15-AAP1155

Key words and phrases. First passage percolation, mean field approximation, Derrida's REMs.

Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 426–438
https://doi.org/10.1214/19-BJPS430

Branching random walks with uncountably many extinction probability vectors

Daniela Bertacchi[a] **and Fabio Zucca**[b]

[a]*Università di Milano–Bicocca*
[b]*Politecnico di Milano*

Abstract. Given a branching random walk on a set X, we study its extinction probability vectors $\mathbf{q}(\cdot, A)$. Their components are the probability that the process goes extinct in a fixed $A \subseteq X$, when starting from a vertex $x \in X$. The set of extinction probability vectors (obtained letting A vary among all subsets of X) is a subset of the set of the fixed points of the generating function of the branching random walk. In particular here we are interested in the cardinality of the set of extinction probability vectors. We prove results which allow to understand whether the probability of extinction in a set A is different from the one of extinction in another set B. In many cases there are only two possible extinction probability vectors and so far, in more complicated examples, only a finite number of distinct extinction probability vectors had been explicitly found. Whether a branching random walk could have an infinite number of distinct extinction probability vectors was not known. We apply our results to construct examples of branching random walks with uncountably many distinct extinction probability vectors.

References

Bertacchi, D., Coletti, C. F. and Zucca, F. (2017). Global survival of branching random walks and tree-like branching random walks. *ALEA—Latin American Journal of Probability and Mathematical Statistics* **14**, 382–402. MR3647298

Bertacchi, D. and Zucca, F. (2008). Critical behaviours and critical values of branching random walks on multigraphs. *Journal of Applied Probability* **45**, 481–497. MR2426846 https://doi.org/10.1239/jap/1214950362

Bertacchi, D. and Zucca, F. (2009). Characterization of the critical values of branching random walks on weighted graphs through infinite-type branching processes. *Journal of Statistical Physics* **134**, 53–65. MR2489494 https://doi.org/10.1007/s10955-008-9653-5

Bertacchi, D. and Zucca, F. (2012). Recent results on branching random walks. In *Statistical Mechanics and Random Walks: Principles, Processes and Applications*, 289–340. New York: Nova Science Publishers.

Bertacchi, D. and Zucca, F. (2014). Strong local survival of branching random walks is not monotone. *Advances in Applied Probability* **46**, 400–421. MR3215539 https://doi.org/10.1239/aap/1401369700

Bertacchi, D. and Zucca, F. (2017). A generating function approach to branching random walks. *Brazilian Journal of Probability and Statistics* **31**, 229–253. MR3635904 https://doi.org/10.1214/16-BJPS311

Braunsteins, P. and Hautphenne, S. (2017). Extinction in lower Hessenberg branching processes with countably many types. Preprint. Available at arXiv:1706.02919.

Braunsteins, P. and Hautphenne, S. (2018). The probabilities of extinction in a branching random walk on a strip. Preprint. Available at arXiv:1805.07634.

Machado, F. P., Menshikov, M. V. and Popov, S. Y. (2001). Recurrence and transience of multitype branching random walks. *Stochastic Processes and Their Applications* **91**, 21–37. MR1807360 https://doi.org/10.1016/S0304-4149(00)00055-7

Moyal, J. E. (1962). Multiplicative population chains. *Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences* **266**, 518–526. MR0138132 https://doi.org/10.1098/rspa.1962.0075

Pemantle, R. and Stacey, A. M. (2001). The branching random walk and contact process on Galton–Watson and nonhomogeneous trees. *Annals of Probability* **29**, 1563–1590. MR1880232 https://doi.org/10.1214/aop/1015345762

Key words and phrases. Branching random walk, generating function, fixed point, extinction probability vector, tree, comb.

Spataru, A. (1989). Properties of branching processes with denumerable many types. *Revue Roumaine de Mathématiques Pures et Appliquées* **34**, 747–759. MR1031234

Woess, W. (2000). *Random Walks on Infinite Graphs and Groups*. *Cambridge Tracts in Mathematics* **138**. Cambridge: Cambridge Univ. Press. MR1743100 https://doi.org/10.1017/CBO9780511470967

Zucca, F. (2011). Survival, extinction and approximation of discrete-time branching random walks. *Journal of Statistical Physics* **142**, 726–753. MR2773785 https://doi.org/10.1007/s10955-011-0134-x

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: http://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 http://www.e-publications.org/ims/submission.

Preparation of Manuscripts: http://www.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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998. 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:
http://www.imstat.org/individual-membership/

Individual and General Subscriptions:
http://www.imstat.org/orders/

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

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