# Contents



BJPS is an IMS supported journal ISSN 0103-0752 Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 34 • Number 1 • February 2020

# Brazilian Journal of Probability and Statistics

Volume 34 • Number 1 • February 2020

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 34, Number 1, February 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. 1, 1
https://doi.org/10.1214/19-BJPS341ED


# A message from the editorial board

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 2–17
https://doi.org/10.1214/18-BJPS409


# Simple step-stress models with a cure fraction

**Nandini Kannan[a] and Debasis Kundu[b]**

[a]*National Science Foundation*
[b]*Indian Institute of Technology Kanpur*

**Abstract.** In this article, we consider models for time-to-event data obtained from experiments in which stress levels are altered at intermediate stages during the observation period. These experiments, known as step-stress tests, belong to the larger class of accelerated tests used extensively in the reliability literature. The analysis of data from step-stress tests largely relies on the popular cumulative exposure model. However, despite its simple form, the utility of the model is limited, as it is assumed that the hazard function of the underlying distribution is discontinuous at the points at which the stress levels are changed, which may not be very reasonable. Due to this deficiency, Kannan et al. (*Journal of Applied Statistics* **37** (2010b) 1625–1636) introduced the cumulative risk model, where the hazard function is continuous. In this paper, we propose a class of parametric models based on the cumulative risk model assuming the underlying population contains long-term survivors or 'cured' fraction. An EM algorithm to compute the maximum likelihood estimators of the unknown parameters is proposed. This research is motivated by a study on altitude decompression sickness. The performance of different parametric models will be evaluated using data from this study.

## References

Bai, D. S., Kim, M. S. and Lee, S. H. (1989). Optimum simple step-stress accelerated life test with censoring. *IEEE Transactions on Reliability* **38**, 528–532.

Balakrishnan, N. (2009). A synthesis of exact inferential results for exponential step-stress models and associated optimal accelerated life-tests. *Metrika* **69**, 351–396. MR2481928 https://doi.org/10.1007/s00184-008-0221-4

Balakrishnan, N., Kundu, D., Ng, H. K. T. and Kannan, N. (2007). Point and interval estimation for a simple step-stress model with type-II censoring. *Journal of Quality Technology* **9**, 35–47.

Beltrami, J. (2011). Competing risks in the step-stress model with lagged effects. Ph.D. thesis, The University of Texas at San Antonio. MR2942153

Bhattacharyya, G. K. and Soejoeti, Z. (1989). A tampered failure rate model for step-stress accelerated life test. *Communications in Statistics Theory and Methods* **18**, 1627–1643. MR1010126 https://doi.org/10.1080/03610928908829990

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society, Series B* **11**, 15–53.

Cancho, V. G. and Bolafarine, H. (2001). Modeling the presence of immunes by using the exponentiated-Weibull model. *Journal of Applied Statistics* **28**, 659–671. MR1858829 https://doi.org/10.1080/02664760120059200

Chen, M.-H., Ibrahim, J. and Sinha, D. (1999). A new Bayesian model for survival data with surviving fraction. *Journal of the American Statistical Association* **94**, 909–918. MR1723307 https://doi.org/10.2307/2670006

Gamel, J. W., Mclean, I. W. and Rosenberg, S. H. (1999). Proportion cured and mean long-survival time as function of tumor size. *Statistics in Medicine* **9**, 999–1006.

Ghitany, M. E. and Maller, R. A. (1992). Asymptotic results for exponential mixture models with long-term survivors. *Statistics* **23**, 321–336. MR1238254 https://doi.org/10.1080/02331889208802379

Gouno, E. and Balakrishnan, N. (2001). *Step-Stress Accelerated Life Test*. *Hand Book of Statistics*. Amsterdam, The Netherlands: North-Holland. MR1861939 https://doi.org/10.1016/S0169-7161(01)20025-X

Greven, S., Bailer, J., Kupper, L. L., Muller, K. E. and Craft, J. L. (2004). A parametric model for studying organism fitness using step-stress experiments. *Biometrics* **60**, 793–799. MR2089456 https://doi.org/10.1111/j.0006-341X.2004.00230.x

*Key words and phrases.* Cumulative exposure model, step-stress tests, cured fraction, maximum likelihood estimation, EM algorithm.

Gupta, R. D. and Kundu, D. (1999). Generalized exponential distribution. *Australian & New Zealand Journal of Statistics* **41**, 173–188. MR1705342 https://doi.org/10.1111/1467-842X.00072

Ismail, A. A. (2016). Statistical inference for a step-stress partially-accelerated life test model with an adaptive type-I progressively hybrid censored data from Weibull distribution. *Statistical Papers* **57**, 271–301. MR3483189 https://doi.org/10.1007/s00362-014-0639-x

Kannan, N., Kundu, D. and Balakrishnan, N. (2010a). Survival models for step-stress experiments with lagged effects. In *Advances in Degradation Modelling* (M. Nikulin, N. Limnios and N. Balakrishnan, eds.) 355–369. New York: Birkhauser. MR2642594 https://doi.org/10.1007/978-0-8176-4924-1_23

Kannan, N., Kundu, D., Nair, R. C. and Tripathi, R. C. (2010b). The generalized exponential cure rate model with covariates. *Journal of Applied Statistics* **37**, 1625–1636. MR2758672 https://doi.org/10.1080/02664760903117739

Kannan, N., Raychaudhury, A. and Pilmanis, A. A. (1998). A logistic model for altitude decomposition sickness. *Aviation, Space, and Environmental Medicine* **69**, 965–970.

Kateri, M. and Balakrishnan, N. (2008). Inference for a simple step-stress model with type-II censoring and Weibull distributed lifetimes. *IEEE Transactions on Reliability* **57**, 616–626.

Kateri, M. and Kamps, U. (2015). Inference in step-stress models based on failure rates. *Statistical Papers* **56**, 639–660. MR3369423 https://doi.org/10.1007/s00362-014-0601-y

Khamis, I. H. and Higgins, J. J. (1998). A new model for step-stress testing. *IEEE Transactions on Reliability* **47**, 131–134.

Maller, R. A. and Zhou, S. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley. MR1453117

Nelson, W. B. (1980). Accelerated life testing: Step-stress models and data analysis. *IEEE Transactions on Reliability* **29**, 103–108.

Nelson, W. B. (1990). *Accelerated Life Testing, Statistical Models, Test Plans and Data Analysis*. New York: John Wiley & Sons.

Sedyakin, N. M. (1966). On one physical principle in reliability theory. *Technical Cybernetics* **3**, 80–87 (in Russian).

Self, S. G. and Liang, K.-L. (1987). Asymptotic properties of the maximum likelihood estimators and likelihood ratio test under non-standard conditions. *Journal of the American Statistical Association* **82**, 605–610. MR0898365

Sha, N. and Pan, R. (2014). Bayesian analysis for step-stress accelerated life testing using Weibull proportional hazard model. *Statistical Papers* **55**, 715–726. MR3227548 https://doi.org/10.1007/s00362-013-0521-2

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54**, 1508–1516.

Xiong, C. (1998). Inferences on a simple step-stress model with type-II censored exponential data. *IEEE Transactions on Reliability* **47**, 142–146.

Xiong, C. and Milliken, G. A. (1999). Step-stress life-testing with random change times for exponential data. *IEEE Transactions on Reliability* **48**, 141–148.

Yu, B., Tiwari, R. C., Cronin, K. A. and Feuer, E. C. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine* **23**, 1733–1747.

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 18–34
https://doi.org/10.1214/18-BJPS412


# Bootstrap-based testing inference in beta regressions

**Fábio P. Lima and Francisco Cribari-Neto**

*Universidade Federal de Pernambuco*

**Abstract.** We address the issue of performing testing inference in small samples in the class of beta regression models. We consider the likelihood ratio test and its standard bootstrap version. We also consider two alternative resampling-based tests. One of them uses the bootstrap test statistic replicates to numerically estimate a Bartlett correction factor that can be applied to the likelihood ratio test statistic. By doing so, we avoid estimation of quantities located in the tail of the likelihood ratio test statistic null distribution. The second alternative resampling-based test uses a fast double bootstrap scheme in which a single second level bootstrapping resample is performed for each first level bootstrap replication. It delivers accurate testing inferences at a computational cost that is considerably smaller than that of a standard double bootstrapping scheme. The Monte Carlo results we provide show that the standard likelihood ratio test tends to be quite liberal in small samples. They also show that the bootstrap tests deliver accurate testing inferences even when the sample size is quite small. An empirical application is also presented and discussed.

## References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, 267–281. MR0483125

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* **30**, 9–14. MR0507075 https://doi.org/10.1007/BF02480194

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences* **160**, 268–282.

Bayer, F. M. and Cribari-Neto, F. (2013). Bartlett corrections in beta regression models. *Journal of Statistical Planning and Inference* **143**, 531–547. MR2995113 https://doi.org/10.1016/j.jspi.2012.08.018

Bayer, F. M. and Cribari-Neto, F. (2015). Bootstrap-based model selection criteria for beta regressions. *Test* **24**, 776–795. MR3414517 https://doi.org/10.1007/s11749-015-0434-6

Bayer, F. M. and Cribari-Neto, F. (2017). Model selection criteria in beta regression with varying dispersion. *Communications in Statistics Simulation and Computation* **46**, 729–746. MR3563524 https://doi.org/10.1080/03610918.2014.977918

Cordeiro, G. M. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*, 1st ed. New York: Springer. MR3289997 https://doi.org/10.1007/978-3-642-55255-7

Cribari-Neto, F. and Cordeiro, G. M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339–367. MR1423902 https://doi.org/10.1080/07474939608800361

Cribari-Neto, F. and Lucena, S. E. F. (2015). Nonnested hypothesis testing in the class of varying dispersion beta regressions. *Journal of Applied Statistics* **42**, 967–985. MR3315740 https://doi.org/10.1080/02664763.2014.993368

Cribari-Neto, F. and Queiroz, M. P. F. (2014). On testing inference in beta regressions. *Journal of Statistical Computation and Simulation* **84**, 186–203. MR3169320 https://doi.org/10.1080/00949655.2012.700456

Davidson, J. (2006). Alternative bootstrap procedures for testing cointegration in fractionally integrated processes. *Journal of Econometrics* **133**, 741–777. MR2252915 https://doi.org/10.1016/j.jeconom.2005.06.012

Davidson, R. and MacKinnon, J. G. (2000). Improving the reliability of bootstrap tests. Working Papers No. 995, Dept. Economics, Queen's Univ.

Davidson, R. and MacKinnon, J. G. (2002). Fast double bootstrap tests of nonnested linear regression models. *Econometric Reviews* **21**, 419–429. MR1951651 https://doi.org/10.1081/ETC-120015384

*Key words and phrases.* Bartlett correction, beta regression, bootstrap, double bootstrap, fast double bootstrap, likelihood ratio test.

Davidson, R. and Trokic, M. (2011). The iterated bootstrap. Working Paper, McGill University.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7**, 1–26. MR0515681

Espinheira, P. L., 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

Espinheira, P. L., Ferrari, S. L. P. and Cribari-Neto, F. (2014). Bootstrap prediction intervals in beta regressions. *Computational Statistics* **29**, 1263–1277. MR3266058 https://doi.org/10.1007/s00180-014-0490-5

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

Ferrari, S. L. P., Espinheira, P. L. and Cribari-Neto, F. (2011). Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica* **65**, 337–351. MR2857878 https://doi.org/10.1111/j.1467-9574.2011.00488.x

Ferrari, S. L. P. and Pinheiro, E. C. (2011). Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation* **81**, 431–443. MR2782138 https://doi.org/10.1080/00949650903389993

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

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237 https://doi.org/10.1093/biomet/43.3-4.295

MacKinnon, J. G. (2006). Applications of the fast double bootstrap. Working Papers No. 1023, Dept. Economics, Queen's Univ.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692. MR1130937 https://doi.org/10.1093/biomet/78.3.691

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*, 2nd ed. New York: Springer. MR2244940

Omtzigt, P. H. and Fachin, S. (2002). Bootstrapping and Bartlett corrections in the cointegrated VAR model. Econometrics Discussion Paper 15.

Ospina, R., Cribari-Neto, F. and Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis* **51**, 960–981. Errata: 55, 2011, 2445. MR2297500 https://doi.org/10.1016/j.csda.2005.10.002

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

Ouysse, R. (2011). Computationally efficient approximation for the double bootstrap mean bias correction. *Economics Bulletin* **31**, 2388–2403.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. New York: Cambridge University Press. MR1414682

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society B* **31**, 350–371. MR0290502

Rocke, D. M. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association* **84**, 598–601. MR1010351

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

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons. MR0595165

Simas, A. B., Barreto-Souza, W. and Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* **54**, 348–366. MR2756431 https://doi.org/10.1016/j.csda.2009.08.017

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics* **28**, 3–32. MR1844348 https://doi.org/10.1111/1467-9469.00223

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

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 35–50
https://doi.org/10.1214/18-BJPS416


# A joint mean-correlation modeling approach for longitudinal zero-inflated count data

**Weiping Zhang, Jiangli Wang Fang Qian and Yu Chen**

*University of Science and Technology of China*

**Abstract.** Longitudinal zero-inflated count data are widely encountered in many fields, while modeling the correlation between measurements for the same subject is more challenge due to the lack of suitable multivariate joint distributions. This paper studies a novel mean-correlation modeling approach for longitudinal zero-inflated regression model, solving both problems of specifying joint distribution and parsimoniously modeling correlations with no constraint. The joint distribution of zero-inflated discrete longitudinal responses is modeled by a copula model whose correlation parameters are innovatively represented in hyper-spherical coordinates. To overcome the computational intractability in maximizing the full likelihood function of the model, we further propose a computationally efficient pairwise likelihood approach. We then propose separated mean and correlation regression models to model these key quantities, such modeling approach can also handle irregularly and possibly subject-specific times points. The resulting estimators are shown to be consistent and asymptotically normal. Data example and simulations support the effectiveness of the proposed approach.

## References

Alfo, M. and Maruotti, A. (2010). Two-part regression models for longitudinal zero-inflated count data. *Canadian Journal of Statistics* **38**, 197–216. MR2682758 https://doi.org/10.1002/cjs.10056

Atkins, D. C. and Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology* **21**, 726–735. https://doi.org/10.1037/0893-3200.21.4.726

Bergsma, W., Croon, M. and Hagenaars, J. A. (2009). *Marginal Models for Dependent, Clustered, and Longitudinal Categorical Data*. Berlin: Springer. MR1057178

Berk, K. N. and Lachenbruch, P. A. (2002). Repeated measures with zeros. *Statistical Methods in Medical Research* **11**, 303–316. https://doi.org/10.1191/0962280202sm293ra

Bulsara, M. K., Holman, C. D. J., Davis, E. A. and Jones, T. W. (2004). Evaluating risk factors associated with severe hypoglycaemia in epidemiology studies—what method should we use? *Diabetic Medicine* **21**, 914–919. https://doi.org/10.1111/j.1464-5491.2004.01250.x

Buu, A., Li, R., Tan, X. and Zucker, R. A. (2012). Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine* **31**, 4074–4086. MR3041794 https://doi.org/10.1002/sim.5510

Creal, D., Koopman, S. J. and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* **29**, 552–563. MR2879242 https://doi.org/10.1198/jbes.2011.10070

Deb, P., Trivedi, P. and Zimmer, D. M. (2014). Cost-offsets of prescription drug expenditures: Data analysis via a copula-based bivariate dynamic hurdle model. *Health Economics* **23**, 1242–1259. https://doi.org/10.1002/hec.2982

Fan, J., Liu, H., Ning, Y. and Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society, Series B* **79**, 405–421. MR3611752 https://doi.org/10.1111/rssb.12168

Fang, H., Fang, K. and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* **82**, 1–16. MR1918612 https://doi.org/10.1006/jmva.2001.2017

*Key words and phrases.* Copula, hyperspherical coordinates, mean-correlation regression, pairwise likelihood, zero inflated negative binomial.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **62**, 424–431. MR2227490 https://doi.org/10.1111/j.1541-0420.2006.00507.x

Ghosh, S. K., Mukhopadhyay, P. and Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference* **136**, 1360–1375. MR2253768 https://doi.org/10.1016/j.jspi.2004.10.008

Ground, M. and Koch, S. F. (2008). Hurdle models of alcohol and tobacco expenditure in South African households. *The South African Journal of Economics* **76**, 132–143. https://doi.org/10.1111/j.1813-6982.2008.00156.x

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Boca Raton: CRC Press. MR1462613 https://doi.org/10.1201/b13150

Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics* **30**, 63–77. MR1957361 https://doi.org/10.1080/0266476022000018510

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. New York: Marcel Dekker. MR1169465

Lee, A. H., Wang, K., Scott, J. A., et al (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medicine Research* **15**, 47–61. MR2225145 https://doi.org/10.1191/0962280206sm429oa

Leng, C., Zhang, W. and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association* **105**, 181–193. MR2656048 https://doi.org/10.1198/jasa.2009.tm08485

Lewsey, J. D. and Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology* **32**, 183–189.

Liu, H., Lafferty, J. D. and Wasserman, L. A. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research* **10**, 2295–2328. MR2563983

Liu, M., Zhang, W. and Chen, Y. (2018). Bayesian joint semiparametric mean-covariance modeling for longitudinal data. *Communications in Mathematics and Statistics* **7**, 1–15. MR3611276

Liu, X. and Zhang, W. (2013). A moving average Cholesky factor model in joint mean-covariance modeling for longitudinal data. *Science China. Mathematics* **56**, 2367–2379. MR3123576 https://doi.org/10.1007/s11425-013-4608-y

Madsen, L. and Fang, Y. (2011). Joint regression analysis for discrete longitudinal data. *Biometrics* **67**, 1171–1175. MR2829253 https://doi.org/10.1111/j.1541-0420.2010.01494.x

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modeling* **5**, 1–19. MR2133525 https://doi.org/10.1191/1471082X05st084oa

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Berlin: Springer. MR2171048

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**, 341–365. MR0867980 https://doi.org/10.1016/0304-4076(86)90002-3

Neighbors, C., Lewis, M. A., Atkins, D. C., Jensen, M. M., Walter, T., Fossos, N., et al (2010). Efficacy of web-based personalized normative feedback: A two-year randomized controlled trial. *Journal of Consulting and Clinical Psychology* **78**, 898–911.

Pan, J. and Mackenzie, G. (2003). On modelling mean–covariance structures in longitudinal studies. *Biometrika* **90**, 239–244. MR1966564 https://doi.org/10.1093/biomet/90.1.239

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677–690. MR1723786 https://doi.org/10.1093/biomet/86.3.677

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–435. MR1782488 https://doi.org/10.1093/biomet/87.2.425

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika* **94**, 1006–1013. MR2376812 https://doi.org/10.1093/biomet/asm073

Renard, D., Molenberghs, G. and Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis* **44**, 649–667. MR2026438 https://doi.org/10.1016/S0167-9473(02)00263-3

Rose, C. E., Martin, S. W., Wannemuehler, K. A. and Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics* **16**, 463–481. MR2242134 https://doi.org/10.1080/10543400600719384

Shi, P. and Zhang, W. (2015). Private information in healthcare utilization: Specification of a copula-based hurdle model. *Journal of the Royal Statistical Society Series A Statistics in Society* **178**, 337–361. MR3300007 https://doi.org/10.1111/rssa.12065

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de L'Institut de Statistiques de l'Universite de Paris* **8**, 229–231. MR0125600

Smith, M. S. and Khaled, M. A. (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association* **107**, 290–303. MR2949360 https://doi.org/10.1080/01621459.2011.644501

Song, P. X.-K., Li, M. and Yuan, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* **65**, 60–68. MR2665846 https://doi.org/10.1111/j.1541-0420.2008.01058.x

Song, P. X. K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics* **27**, 305–320. MR1777506 https://doi.org/10.1111/1467-9469.00191

Tang, C. Y., Zhang, W. and Leng, C. (2018). Discrete longitudinal data modeling with a mean-corrrelation regression approach. *Statistica Sinica*. https://doi.org/10.5705/ss.202016.0435. Preprint. https://doi.org/10.5705/ss.202016.0435

Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Berlin: Springer. MR1029032 https://doi.org/10.1007/978-1-4613-9655-0

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42. MR2796852

White, H. R. and Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol* **50**, 30–37.

Ye, H. and Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika* **93**, 927–941. MR2285080 https://doi.org/10.1093/biomet/93.4.927

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.

Zhang, W. and Leng, C. (2012). A moving average cholesky factor model in covariance modeling for longitudinal data. *Biometrika* **99**, 141–150. MR2899669 https://doi.org/10.1093/biomet/asr068

Zhang, W., Leng, C. and Tang, C. Y. (2015). A joint modeling approach for longitudinal studies. *Journal of the Royal Statistical Society, Series B* **77**, 219–238. MR3299406 https://doi.org/10.1111/rssb.12065

Zimmer, D. (2018). Using copulas to estimate the coefficient of a binary endogenous regressor in a Poisson regression: Application to the effect of insurance on doctor visits. *Health Economics* **27**, 545–556.

Zimmer, D. and Trivedi, P. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business & Economic Statistics* **24**, 63–76. MR2234712 https://doi.org/10.1198/073500105000000153

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 51–70
https://doi.org/10.1214/18-BJPS417


# Robust Bayesian model selection for heavy-tailed linear regression using finite mixtures

**Flávio B. Gonçalves**[a], **Marcos O. Prates**[a] **and Victor Hugo Lachos**[b]

[a]*Universidade Federal de Minas Gerais*
[b]*University of Connecticut*

**Abstract.** In this paper, we present a novel methodology to perform Bayesian model selection in linear models with heavy-tailed distributions. We consider a finite mixture of distributions to model a latent variable where each component of the mixture corresponds to one possible model within the symmetrical class of normal independent distributions. Naturally, the Gaussian model is one of the possibilities. This allows for a simultaneous analysis based on the posterior probability of each model. Inference is performed via Markov chain Monte Carlo—a Gibbs sampler with Metropolis–Hastings steps for a class of parameters. Simulated examples highlight the advantages of this approach compared to a segregated analysis based on arbitrarily chosen model selection criteria. Examples with real data are presented and an extension to censored linear regression is introduced and discussed.

## References

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

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

Cabral, C. R. B., Lachos, V. H. and Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis* **56**, 126–142. MR2833042 https://doi.org/10.1016/j.csda.2011.06.026

Carlin, B. P. (2006). Comments to discussion of deviance information criteria for missing data. *Bayesian Analysis* **1**, 675–676. MR2282198 https://doi.org/10.1214/06-BA122A

Chen, M.-H. (2006). Comments to discussion of deviance information criteria for missing data. *Bayesian Analysis* **1**, 677–680.

Choy, S. T. B. and Chan, J. S. K. (2008). Scale mixtures distributions in statistical modelling. *Australian & New Zeland Journal of Statistics* **50**, 135–146. MR2516871 https://doi.org/10.1111/j.1467-842X.2008.00504.x

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York: Wiley. MR1285353 https://doi.org/10.1002/9780470316863

Dey, D. K., Chen, M. H. and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53**, 1239–1252.

Fernandez, C. and Steel, M. F. J. (1999). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika* **86**, 153–167. MR1688079 https://doi.org/10.1093/biomet/86.1.153

Fonseca, T. C. O., Ferreira, M. A. R. and Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**, 325–333. MR2521587 https://doi.org/10.1093/biomet/asn001

Galea, M., Paula, G. A. and Cysneiros, F. J. A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters* **73**, 459–467. MR2187861 https://doi.org/10.1016/j.spl.2005.04.033

Galea, M., Paula, G. A. and Uribe-Opazo, M. (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers* **44**, 23–45. MR1963364 https://doi.org/10.1007/s00362-002-0132-9

Garay, A. M., Bolfarine, H., Lachos, V. H. and Cabral, C. R. B. (2015). Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. *Journal of Applied Statistics* **42**, 2694–2714. MR3428840 https://doi.org/10.1080/02664763.2015.1048671

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection (Corr: V75 p765). *Journal of the American Statistical Association* **74**, 153–160. MR0529531

*Key words and phrases.* Scale mixtures of normal, Student-t, slash, penalised complexity priors, MCMC.

Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24**, 997–1016. MR3253850 https://doi.org/10.1007/s11222-013-9416-2

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **85**, 398–409.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Disc: P189-193). In *Bayesian Statistics, Vol. 4. Proceedings of the Fourth Valencia International Meeting* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 169–188. New York: Clarendon Press [Oxford University Press]. MR1380276

Gonçalves, F. B., Gamerman, D. and Soares, T. M. (2013). Simultaneous multifactor DIF analysis and detection in item response theory. *Computational Statistics & Data Analysis* **59**, 144–160. MR3000048 https://doi.org/10.1016/j.csda.2012.10.011

Gonçalves, F. B., Prates, M. O. and Lachos, V. H. (2020). Supplement to "Robust Bayesian model selection for heavy-tailed linear regression using finite mixtures." https://doi.org/10.1214/18-BJPS417SUPP.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

Lachos, V. H., Angolini, T. and Abanto-Valle, C. (2011). On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers* **52**, 567–590. MR2821058 https://doi.org/10.1007/s00362-009-0270-4

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **20**, 303–322. MR2640696

Lange, K. L., Little, R. and Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association* **84**, 881–896. MR1134486

Lange, K. L. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**, 175–198. MR1272391 https://doi.org/10.2307/1390698

Lin, J.-G. and Cao, C.-Z. (2013). On estimation of measurement error models with replication under heavy-tailed distributions. *Computational Statistics* **28**, 809–829. MR3064480 https://doi.org/10.1007/s00180-012-0330-4

Martins, T. G. and Rue, H. (2013). Prior for flexibility parameters: The Student's t case. Preprint 08/2013, Norwegian University of Science and Technology.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* **55**, 765–799.

Osorio, F., Paula, G. A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis* **51**, 4354–4368. MR2364450 https://doi.org/10.1016/j.csda.2006.06.004

Pinheiro, J. C., Liu, C. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* **10**, 249–276. MR1939700 https://doi.org/10.1198/10618600152628059

Raftery, A., Madigan, D. and Volinsky, C. T. (1995). Accounting for model uncertainty in survival analysis improves predictive performance. In *Bayesian Statistics, Vol. 5*, 323–349. New York: Oxford University Press. MR1425413

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

Rosa, G., Padovani, C. and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian mcmc implementation. *Biometrical Journal* **45**, 573–590. MR1998137 https://doi.org/10.1002/bimj.200390034

Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* **32**, 1–28. MR3634300 https://doi.org/10.1214/16-STS576

Spiegelhalter, D., Best, N., Carlin, B. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639. MR1979380 https://doi.org/10.1111/1467-9868.00353

Steel, M. F. J. and Fernandez, C. (1999). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika* **86**, 153–168. MR1688079 https://doi.org/10.1093/biomet/86.1.153

Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *The Annals of Applied Probability* **8**, 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594. MR2756194

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 71–89
https://doi.org/10.1214/18-BJPS413


# Effects of gene–environment and gene–gene interactions in case-control studies: A novel Bayesian semiparametric approach

**Durba Bhattacharya and Sourabh Bhattacharya**

*Indian Statistical Institute*

**Abstract.** Present day bio-medical research is pointing towards the fact that cognizance of gene–environment interactions along with genetic interactions may help prevent or detain the onset of many complex diseases like cardiovascular disease, cancer, type2 diabetes, autism or asthma by adjustments to lifestyle.

In this regard, we propose a Bayesian semiparametric model to detect not only the roles of genes and their interactions, but also the possible influence of environmental variables on the genes in case-control studies. Our model also accounts for the unknown number of genetic sub-populations via finite mixtures composed of Dirichlet processes. An effective parallel computing methodology, developed by us harnesses the power of parallel processing technology to increase the efficiencies of our conditionally independent Gibbs sampling and Transformation based MCMC (TMCMC) methods.

Applications of our model and methods to simulation studies with biologically realistic genotype datasets and a real, case-control based genotype dataset on early onset of myocardial infarction (MI) have yielded quite interesting results beside providing some insights into the differential effect of gender on MI.

## References

Ahn, J., Mukherjee, B., Ghosh, M. and Gruber, S. B. (2013). Bayesian semiparametric analysis of two-phase studies of gene–environment interaction. *Annals of Applied Statistics* **7**, 543–569. MR3086430 https://doi.org/10.1214/12-AOAS599

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer. MR0804611 https://doi.org/10.1007/978-1-4757-4286-2

Bhattacharjee, S., Wang, Z., Ciampa, J., Kraft, P., Chanock, S., Yu, K. and Chatterjee, N. (2010). Using principal components of genetic variation for robust and powerful detection of gene–gene interactions in case-control and case-only studies. *American Journal of Human Genetics* **86**, 331–342.

Bhattacharya, D. and Bhattacharya, S. (2016). A Bayesian semiparametric approach to learning about gene–gene interactions in case-control studies. Preprint. Available at http://arxiv.org/abs/1411.7571. MR3860648 https://doi.org/10.1080/02664763.2018.1444741

Bhattacharya, D. and Bhattacharya, S. (2020). Supplement to "Effects of gene–environment and gene–gene interactions in case-control studies: A novel Bayesian semiparametric approach." https://doi.org/10.1214/18-BJPS413SUPP.

Brockwell, A. (2006). Parallel Markov chain Monte Carlo simulation by pre-fetching. *Journal of Computational and Graphical Statistics* **15**, 246–261. MR2269370 https://doi.org/10.1198/106186006X100579

Calderhead, B. (2014). A general construction for parallelizing Metropolis–Hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17408–17413.

Chen, Y., Freitas, N. D., Eskelin, M., Fang, J. and Welling, M. (2016). Herded Gibbs sampling. *Journal of Machine Learning Research* **17**, 263–291. MR3491104

De Iorio, M., Favaro, S. and Teh, Y. W. (2015). Bayesian inference on population structure: From parametric to nonparametric modeling. In *Nonparametric Bayesian Inference in Biostatistics*, 135–151. Cham: Springer. MR3411018

Dutta, S. and Bhattacharya, S. (2014). Markov chain Monte Carlo based on deterministic transformations. *Statistical Methodology* **16**, 100–116. Also available at http://arxiv.org/abs/1106.5850. Supplement available at http://arxiv.org/abs/1306.6684.

*Key words and phrases.* Case-control study, Dirichlet process, gene–gene and gene–environment interaction, matrix normal, parallel processing, transformation based MCMC.

Erdmann, J., Linsel-Nitschke, P. and Schunkert, H. (2010). Genetic causes of myocardial infarction. *Deutsches Ärzteblatt International* **107**, 694–699.

Hunter, D. J. (2005). Gene environment interactions in human diseases. *Nature Reviews Genetics* **6**, 287–298.

Jacob, P., Robert, C. P. and Smith, M. H. (2011). Using parallel computation to improve independent Metropolis–Hastings based estimation. *Journal of Computational and Graphical Statistics* **3**, 616–635. MR2878993 https://doi.org/10.1198/jcgs.2011.10167

Khouri, M. J. (2005). Do we need genomic research for the prevention of common diseases with environmental causes? *American Journal of Epidemiology* **161**, 799–805.

Ko, Y.-A., Saha Chaudhuri, P., Vokonas, P. S., Park, S. K. and Mukherjee, B. (2013). Likelihood ratio tests for detecting gene environment interaction in longitudinal studies. *Genetic Epidemiology* **37**, 581–591.

Lucas, G., Lluis-Ganella, C., Subirana, I., Masameh, M. D. and Gonzalez, J. R. (2012). Hypothesis-based analysis of gene–gene interaction and risk of myocardial infraction. *PLoS ONE* **7**, e41730.

Majumdar, A., Bhattacharya, S., Basu, A. and Ghosh, S. (2013). A novel Bayesian semiparametric algorithm for inferring population structure and adjusting for case-control association tests. *Biometrics* **69**, 164–173. MR3058063 https://doi.org/10.1111/biom.12004

Mapp, C. (2003). The role of genetic factors in occupational asthma. *European Respiratory Journal* **21**, 173–178.

Martino, L., Elvira, V. and Camps-Valls, G. (2018). The recycling Gibbs sampler for efficient learning. *Digital Signal Processing* **74**, 1–13. MR3754555 https://doi.org/10.1016/j.dsp.2017.11.012

Martino, L., Elvira, V., Luengo, D., Corander, J. and Louzada, F. (2016). Orthogonal parallel MCMC methods for sampling and optimization. *Digital Signal Processing* **58**, 64–84.

Mather, K. and Caligary, P. (1976). Genotype x environmental interactions. *Heredity* **36**, 41–48.

Mukherjee, B., Ahn, J., Gruber, S. B. and Chatterjee, N. (2012). Testing gene environment interaction in large-scale association studies. *American Journal of Epidemiology* **175**, 177–190.

Mukherjee, B., Ahn, J., Gruber, S. B., Ghosh, M. and Chatterjee, N. (2010). Bayesian sample size determination for case-control studies of gene–environment interaction. *Biometrics* **66**, 934–948. MR2758230 https://doi.org/10.1111/j.1541-0420.2009.01357.x

Mukherjee, B., Ahn, J., Gruber, S. B., Moreno, V. and Chatterjee, N. (2008). Testing gene–environment interaction from case-control data: A novel study of type-I error, power and designs. *Genetic Epidemiology* **32**, 615–626.

Mukherjee, B. and Chatterjee, N. (2008). Exploiting gene–environment independence for analysis of case-control studies: An empirical-Bayes type shrinkage estimator to trade off between bias and efficiency. *Biometrics* **64**, 685–694. MR2526617 https://doi.org/10.1111/j.1541-0420.2007.00953.x

Mukhopadhyay, S., Bhattacharya, S. and Dihidar, K. (2011). On Bayesian "central clustering": Application to landscape classification of Western Ghats. *Annals of Applied Statistics* **5**, 1948–1977. MR2884928 https://doi.org/10.1214/11-AOAS454

Ottman, R. (2010). Gene environment interactions: Definitions and study designs. *Pubmed* **6**, 764–770.

Pinelli, M., Scala, G., Amato, R., Cocozza, S. and Miele, G. (2012). Simulating gene–gene and gene–environment interactions in complex diseases: Gene–environment iNteraction simulator 2. *BMC Bioinformatics* **13**, 132.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research* **5**, 554–571.

Qi, L., Ma, J., Qi, Q., Hartiala, J., Allayee, H. and Campos, H. (2011). Genetic risk score and risk of myocardial infarction in hispanics. *Circulation* **123**, 374–380.

Sanchez, B., Kang, S. and Mukherjee, B. (2012). A latent variable approach to study of gene–environment interactions in the presence of multiple correlated exposures. *Biometrics* **68**, 466–476. MR2959613 https://doi.org/10.1111/j.1541-0420.2011.01677.x

Scott, G. and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics* **38**, 2587–2619. MR2722450 https://doi.org/10.1214/10-AOS792

Scott, S. A. (2011). Personalizing medicine with clinical pharmacogenetics. *Genetics in Medicine* **13**, 987–995.

Wang, Q., Rao, S., Shen, G.-Q., Li, L., Moliterno, D. J., Newby, L. K., Rogers, W. J., Cannata, R., Zirzow, E., Elston, R. C. and Topol, E. J. (2004). Premature myocardial infarction novel susceptibility locus on chromosome 1P34-36 identified by genomewide linkage analysis. *Circulation* **74**, 262–271.

Wang, X., Elston, R. C. and Zhu, X. (2010). The meaning of interaction. *Human Heredity* **70**, 269–277.

Wright, A. F., Carothers, A. D. and Campbell, H. (2002). Gene–environment interactions—The BioBank UK study. *Pharmacogenomics Journal* **2**, 75–82.

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 90–111
https://doi.org/10.1214/18-BJPS414


# On the Nielsen distribution

**Fredy Castellares**[a], **Artur J. Lemonte**[b] **and Marcos A. C. Santos**[a]

[a]*Departamento de Estatística, ICEX, Universidade Federal de Minas Gerais*
[b]*Departamento de Estatística, CCET, Universidade Federal do Rio Grande no Norte*

**Abstract.** We introduce a two-parameter discrete distribution that may have a zero vertex and can be useful for modeling overdispersion. The discrete Nielsen distribution generalizes the Fisher logarithmic (i.e., logarithmic series) and Stirling type I distributions in the sense that both can be considered displacements of the Nielsen distribution. We provide a comprehensive account of the structural properties of the new discrete distribution. We also show that the Nielsen distribution is infinitely divisible. We discuss maximum likelihood estimation of the model parameters and provide a simple method to find them numerically. The usefulness of the proposed distribution is illustrated by means of three real data sets to prove its versatility in practical applications.

## References

Allison, P. D. (2012). *Logistic Regression Using SAS: Theory and Application*, 2nd ed. Cary, North Carolina: SAS Institute Inc.

Barbiero, A. (2014). An alternative discrete skew Laplace distribution. *Statistical Methodology* **16**, 47–67. MR3110887 https://doi.org/10.1016/j.stamet.2013.07.002

Catcheside, D. G., Lea, D. E. and Thoday, J. M. (1946a). Types of chromosome structural change induced by the irradiation of Tradescantia microspores. *Journal of Genetics* **47**, 113–136.

Catcheside, D. G., Lea, D. E. and Thoday, J. M. (1946b). The production of chromosome structural changes in Tradescantia microspores in relation to dosage, intensity and temperature. *Journal of Genetics* **47**, 137–149.

Englehardt, J. D. and Li, R. C. (2011). The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis* **31**, 370–381.

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

Fisher, R. A., Corbet, A. S. and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**, 42–58.

Flajonet, P. and Sedgewick, R. (2009). *Analytic Combinatorics*. New York: Cambridge University Press. MR2483235 https://doi.org/10.1017/CBO9780511801655

Gómez-Déniz, E., Sarabia, J. M. and Calderin-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance Mathematics & Economics* **48**, 406–412. MR2820054 https://doi.org/10.1016/j.insmatheco.2011.01.007

Gossiaux, A. and Lemaire, J. (1981). Methodes d'ajustement de distributions de sinistres. *Bulletin of the Association of Swiss Actuaries* **81**, 87–95.

Graham, R., Knuth, D. and Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. New York: Addison-Wesley. MR1397498

Inusah, S. and Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference* **136**, 1090–1102. MR2181990 https://doi.org/10.1016/j.jspi.2004.08.014

Jazi, M. A., Lai, C. D. and Alamatsaz, M. H. (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology* **7**, 121–132. MR2591715 https://doi.org/10.1016/j.stamet.2009.11.001

Johnson, N. L. and Kotz, S. (1982). Developments in discrete distributions, 1969–1980. *International Statistical Review* **50**, 71–101. MR0668611 https://doi.org/10.2307/1402460

Jones, M. C. (2015). On families of distributions with shape parameters. *International Statistical Review* **83**, 175–192. MR3377071 https://doi.org/10.1111/insr.12055

*Key words and phrases.* Discrete distribution, Fisher logarithmic distribution, Nielsen expansion, Stirling distributions.

Kozubowski, T. J. and Inusah, S. (2006). A skew Laplace distribution on integers. *Annals of the Institute of Statistical Mathematics* **58**, 555–571. MR2327893 https://doi.org/10.1007/s10463-005-0029-1

Krishna, H. and Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology* **6**, 177–188. MR2649616 https://doi.org/10.1016/j.stamet.2008.07.001

Nekoukhou, V., Alamatsaz, M. H. and Bidram, H. (2013). Discrete generalized exponential distribution of a second type. *Statistics* **47**, 876–887. MR3175721 https://doi.org/10.1080/02331888.2011.633707

Nielsen, N. (1906). *Handbuch der Theorie der Gammafunction*. Leipzig: Teubner.

Nooghabi, M. S., Roknabadi, A. H. R. and Borzadaran, G. M. (2011). Discrete modified Weibull distribution. *Metron* **LXIX**, 207–222. MR3041266 https://doi.org/10.1007/BF03263557

Patil, G. P. (1963). Minimum variance unbiased estimation and certain problems of additive number theory. *The Annals of Mathematical Statistics* **34**, 1050–1056. MR0155388 https://doi.org/10.1214/aoms/1177704029

Patil, G. P. and Wani, J. K. (1965). On certain structural properties of the logarithmic series distribution and the first type Stirling distribution. *Sankhya Series A* **27**, 271–280. MR0202233

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ross, S. M. (2013). *Simulation*, 5th ed. London: Academic Press.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability* **53**, 255–260.

Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data. *Psychometrika* **16**, 107–127. MR0042100 https://doi.org/10.1007/BF02313431

Ward, M. (1934). The representation of Stirling's numbers and Stirling's polynomials as sums of factorial. *American Journal of Mathematics* **56**, 87–95. MR1507004 https://doi.org/10.2307/2370916

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 112–126
https://doi.org/10.1214/18-BJPS418


# Nonparametric discrimination of areal functional data

**Ahmad Younso**

*Damascus University*

**Abstract.** We consider a new nonparametric rule of classification, inspired from the classical moving window rule, that allows for the classification of spatially dependent functional data containing some completely missing curves. We investigate the consistency of this classifier under mild conditions. The practical use of the classifier will be illustrated through simulation studies.

## References

Abraham, C., Biau, G. and Cadre, B. (2006). On the kernel rule for function classification. *Annals of the Institute of Statistical Mathematics* **58**, 619–633. MR2327897 https://doi.org/10.1007/s10463-006-0032-1

Berlinet, A., Biau, G. and Rouvière, L. (2008). Functional supervised classification with wavelets. *Annales de l'ISUP* **52**, 61–80. MR2435041

Biau, G., Bunea, F. and Wegkamp, M. H. (2005). Functional classification in Hilbert spaces. *IEEE Transactions on Information Theory* **51**, 2163–2172. MR2235289 https://doi.org/10.1109/TIT.2005.847705

Biau, G. and Cadre, B. (2004). Nonparametric spatial prediction. *Statistical Inference for Stochastic Processes* **3**, 327–349. MR2111294 https://doi.org/10.1023/B:SISP.0000049116.23705.88

Biau, G., Cérou, F. and Guyader, A. (2010). Rates of convergence of the functional k-nearest neighbor estimate. *IEEE Transactions on Information Theory* **56**, 2034–2040. MR2654492 https://doi.org/10.1109/TIT.2010.2040857

Carbon, M., Francq, C. and Tran, L. T. (2007). Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference* **137**, 778–798. MR2301715 https://doi.org/10.1016/j.jspi.2006.06.008

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis* **92**, 24–41. MR2102242 https://doi.org/10.1016/j.jmva.2003.08.008

Carlo, G., Paolo, G. and Roberto, P. (2017). Spatial clustering of curves with an application of satellite data. *Spatial Statistics* **20**, 110–124. MR3654006 https://doi.org/10.1016/j.spasta.2017.01.006

Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probabilités Et Statistique* **10**, 340–355. MR2247925 https://doi.org/10.1051/ps:2006014

Chang, C., Chen, Y. and Ogden, R. T. (2014). Functional data classification: A wavelet approach. *Computational Statistics* **29**, 1497–1513. MR3279004 https://doi.org/10.1007/s00180-014-0503-4

Dabo-Niang, S. and Yao, A. F. (2007). Kernel regression estimation for continuous spatial processes. *Mathematical Methods of Statistics* **16**, 298–317. MR2378278 https://doi.org/10.3103/S1066530707040023

Dabo-Niang, S. and Yao, A. F. (2013). Kernel spatial density estimation in infinite dimension space. *Metrika* **76**, 19–52. MR3018356 https://doi.org/10.1007/s00184-011-0374-4

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

Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilitic Theory of Pattern Recognition*. New York: Spriner-Verlag. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

Devroye, L. and Krzyżak, A. (2013). An equivalence theorem for L1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference* **23**, 71–82. MR1029241 https://doi.org/10.1016/0378-3758(89)90040-2

Ferraty, F., Van Keilegom, I. and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis* **109**, 10–28. MR2922850 https://doi.org/10.1016/j.jmva.2012.02.008

Górecki, T., Krzysko, M. and Wolynski (2015). Classification problems based on regression models for multidimensional functional data. *Statistics in Transition New Series* **16**, 97–110.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Teoriâ Veroâtnostej I Ee Primeneniâ* **7**, 361–392. MR0148125

*Key words and phrases.* Bayes rule, spectral reflectance curve, functional data, moving window rule, random field, consistency.

Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis* **71**, 92–106. MR3131956 https://doi.org/10.1016/j.csda.2012.12.004

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

Kolmogorov, A. N. and Tihomirov, V. M. (1961). $\epsilon$-entropy and $\epsilon$-capacity of sets in functional spaces. *Translations - American Mathematical Society* **17**, 277–364. MR0124720

Kulkarni, S. R. and Posner, S. E. (1995). Rate of convergence of nearest neighbor estimation under arbitrary sampling. *IEEE Transactions on Information Theory* **41**, 1028–1039. MR1366756 https://doi.org/10.1109/18.391248

Lin, Z. and Yan, L. (2016). A support vector machine classifier based on a new kernel function model for hyperspectral data. *GIScience and Remote Sensing* **53**, 85–101.

Moughal, T. A. (2013). Hyperspectral image classification using support vector machine. *Journal of Physics Conference Series* **439**.

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

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

Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. *Mathématiques et Applications*. Berlin: Spriner. MR2117923

Romano, E., Balzanella, A. and Verde, R. (2010). Clustering spatio-functional data: A model based approach. In *Classification as a Tool for Research*, 167–175. Berlin: Springer. MR2722134 https://doi.org/10.1007/978-3-642-10745-0_17

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America* **42**, 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

Ruiz-Medina, M. D., Espejo, R. and Romano, E. (2014). Spatial functional normal mixed effect approach for curve classification. *Advances in Data Analysis and Classification* **8**, 257–285. MR3253860 https://doi.org/10.1007/s11634-014-0174-6

Saltyte-Benth, J. and Ducinskas, K. (2005). Linear discriminant analysis of multivariate spatial–temporal regressions. *Scandinavian Journal of Statistics* **32**, 281–294. MR2188674 https://doi.org/10.1111/j.1467-9469.2005.00421.x

Ternynck, C. (2014). Spatial regression estimation for functional data with spatial dependency. *Journal de la Société Française de Statistique* **155**, 673–684. MR3211759

Younso, A. (2017a). On the consistency of a new kernel rule for spatially dependent data. *Statistics & Probability Letters* **131**, 64–71. MR3706697 https://doi.org/10.1016/j.spl.2017.08.008

Younso, A. (2017b). On nonparametric classification for weakly dependent functional processes. *ESAIM Probabilités Et Statistique* **21**, 452–466. MR3743922 https://doi.org/10.1051/ps/2017002

Younso, A. (2018). On the consistency of kernel classification rule for functional random field. *Journal de la Société Française de Statistique* **159**, 68–87. MR3803124

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 127–135
https://doi.org/10.1214/18-BJPS415


# A primer on the characterization of the exchangeable Marshall–Olkin copula via monotone sequences

**Natalia Shenkman**

*Technische Universität München*

**Abstract.** While derivations of the characterization of the $d$-variate exchangeable Marshall–Olkin copula via $d$-monotone sequences relying on basic knowledge in probability theory exist in the literature, they contain a myriad of unnecessary relatively complicated computations. We revisit this issue and provide proofs where all undesired artefacts are removed, thereby exposing the simplicity of the characterization. In particular, we give an insightful analytical derivation of the monotonicity conditions based on the monotonicity properties of the survival probabilities.

## References

Cuadras, C. M. (2009). Constructing copula functions with weighted geometric means. *Journal of Statistical Planning and Inference* **139**, 3766–3772. MR2553761 https://doi.org/10.1016/j.jspi.2009.05.016

Esary, J. D. and Marshall, A. W. (1973). Multivariate geometric distributions generated by a cumulative damage process. In *Naval Postgraduate School*. *Monterey, California*. Available at http://calhoun.nps.edu/handle/10945/30003.

Genest, C., Quessy, J.-F. and Rémillard, B. (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *The Annals of Statistics* **35**, 166–191. MR2332273 https://doi.org/10.1214/009053606000000984

Georges, P., Lamy, A. G., Nicolas, E., Quibel, G. and Roncalli, T. (2001). Multivariate survival modelling: A unified approach with copulas. https://doi.org/10.2139/ssrn.1032559

Giesecke, K. (2003). A simple exponential model for dependent defaults. *The Journal of Fixed Income* **13**, 74–83.

Gnedin, A. V. and Pitman, J. (2008). Moments of convex distribution functions and completely alternating sequences. *Probability and Statistics: Essays in Honor of David A. Freedman* **2**, 30–41. MR2459948 https://doi.org/10.1214/193940307000000374

Hausdorff, F. (1921). Summationsmethoden und Momentfolgen. I. *Mathematische Zeitschrift* **9**, 74–109. MR1544453 https://doi.org/10.1007/BF01378337

Mai, J.-F. (2010). Extendibility of Marshall–Olkin distributions via Lévy subordinators and an application to portfolio risk. Dissertation, Technische Universität München. Available at https://mediatum.ub.tum.de/doc/969547/file.pdf.

Mai, J.-F. and Scherer, M. (2009). Lévy-frailty copulas. *Journal of Multivariate Analysis* **100**, 1567–1585. MR2514148 https://doi.org/10.1016/j.jmva.2009.01.010

Mai, J.-F. and Scherer, M. (2011). Reparameterizing Marshall–Olkin copulas with applications to sampling. *Journal of Statistical Computation and Simulation* **81**, 59–78. MR2747378 https://doi.org/10.1080/00949650903185961

Mai, J.-F. and Scherer, M. (2012). *Simulating Copulas*. *Series in Quantitative Finance* **4**. London: Imperial College Press. MR2906392 https://doi.org/10.1142/p842

Mai, J.-F., Scherer, M. and Shenkman, N. (2013). Multivariate geometric distributions, (logarithmically) monotone sequences, and infinitely divisible laws. *Journal of Multivariate Analysis* **115**, 457–480. MR3004570 https://doi.org/10.1016/j.jmva.2012.11.012

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association* **62**(317), 30–44. MR0215400 https://doi.org/10.2307/2282907

Nelsen, R. B. (2006). *An Introduction to Copulas*. *Springer Series in Statistics*. New York: Springer. MR2197664

Ressel, P. (2011). Monotonicity properties of multivariate distribution and survival functions with an application to Lévy-frailty copulas. *Journal of Multivariate Analysis* **102**, 393–404. MR2755005 https://doi.org/10.1016/j.jmva.2010.10.001

*Key words and phrases.* Marshall–Olkin distribution, lack of memory, exchangeability, $d$-monotone sequence.

Ressel, P. (2013). Finite exchangeability, Lévy-frailty copulas and higher-order monotonic sequences. *Journal of Theoretical Probability* **26**, 666–675. MR3090545 https://doi.org/10.1007/s10959-011-0389-9

Schoenberg, I. J. (1932). On finite and infinite completely monotonic sequences. *Bulletin of the American Mathematical Society* **38**, 72–76. MR1562329 https://doi.org/10.1090/S0002-9904-1932-05330-7

Shenkman, N. (2017). A natural parametrization of multivariate distributions with limited memory. *Journal of Multivariate Analysis* **155**, 234–251. MR3607893 https://doi.org/10.1016/j.jmva.2017.01.004

Spizzichino, F. (2009). A concept of duality for multivariate exchangeable survival models. *Fuzzy Sets and Systems* **160**, 325–333. MR2473106 https://doi.org/10.1016/j.fss.2007.10.009

Weiss, S. (2009). The inclusion exclusion principle and its more general version. Available at http://www.compsci.hunter.cuny.edu/~sweiss/resources/inclusion_exclusion.pdf.

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 136–149
https://doi.org/10.1214/18-BJPS411


# Multivariate normal approximation of the maximum likelihood estimator via the delta method

**Andreas Anastasiou**[a] **and Robert E. Gaunt**[b]

[a]*The London School of Economics and Political Science*
[b]*School of Mathematics, The University of Manchester*

**Abstract.** We use the delta method and Stein's method to derive, under regularity conditions, explicit upper bounds for the distributional distance between the distribution of the maximum likelihood estimator (MLE) of a $d$-dimensional parameter and its asymptotic multivariate normal distribution. Our bounds apply in situations in which the MLE can be written as a function of a sum of i.i.d. $t$-dimensional random vectors. We apply our general bound to establish a bound for the multivariate normal approximation of the MLE of the normal distribution with unknown mean and variance.

## References

Anastasiou, A. (2018). Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data. To appear *in the Electronic Journal of Statistics.*

Anastasiou, A. and Ley, C. (2017). Bounds for the asymptotic normality of the maximum likelihood estimator using the Delta method. *ALEA: Latin American Journal of Probability and Mathematical Statistics* **14**, 153–171. MR3622464

Anastasiou, A. and Reinert, G. (2017). Bounds for the normal approximation of the maximum likelihood estimator. *Bernoulli* **23**, 191–218. MR3556771 https://doi.org/10.3150/15-BEJ741

Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics* **32**, 12–40. MR0123420 https://doi.org/10.1214/aoms/1177705136

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **30**, 248–275. MR0237052

Davison, A. C. (2008). *Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press. MR1998913 https://doi.org/10.1017/CBO9780511815850

Gaunt, R. E. (2016). Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the Stein equation. *Journal of Theoretical Probability* **29**, 231–247. MR3463084 https://doi.org/10.1007/s10959-014-0562-z

Gaunt, R. E. and Reinert, G. (2016). The rate of convergence of some asymptotically chi-square distributed statistics by Stein's method. https://arxiv.org/pdf/1603.01889.pdf.

Mäkeläinen, T., Schmidt, K. and Styan, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *The Annals of Statistics* **9**, 758–767. MR0619279

Winkelbauer, A. (2012). Moments and absolute moments of the normal distribution. https://arxiv.org/pdf/1209.4340.pdf.

*Key words and phrases.* Multi-parameter maximum likelihood estimation, multivariate normal distribution, Stein's method.

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 150–166
https://doi.org/10.1214/18-BJPS410


# Application of weighted and unordered majorization orders in comparisons of parallel systems with exponentiated generalized gamma components

**Abedin Haidari**[a], **Amir T. Payandeh Najafabadi**[a] **and Narayanaswamy Balakrishnan**[b]

[a]*Department of Mathematical Sciences, Shahid Beheshti University*
[b]*Department of Mathematics and Statistics, McMaster University*

**Abstract.** Consider two parallel systems, say $A$ and $B$, with respective lifetimes $T_1$ and $T_2$ wherein independent component lifetimes of each system follow exponentiated generalized gamma distribution with possibly different exponential shape and scale parameters. We show here that $T_2$ is smaller than $T_1$ with respect to the usual stochastic order (reversed hazard rate order) if the vector of logarithm (the main vector) of scale parameters of System $B$ is weakly weighted majorized by that of System $A$, and if the vector of exponential shape parameters of System $A$ is unordered mojorized by that of System $B$. By means of some examples, we show that the above results can not be extended to the hazard rate and likelihood ratio orders. However, when the scale parameters of each system divide into two homogeneous groups, we verify that the usual stochastic and reversed hazard rate orders can be extended, respectively, to the hazard rate and likelihood ratio orders. The established results complete and strengthen some of the known results in the literature.

## References

Balakrishnan, N., Haidari, A. and Masoumifard, K. (2015). Stochastic comparisons of series and parallel systems with generalized exponential components. *IEEE Transactions on Reliability* **64**, 333–348.

Balakrishnan, N. and Zhao, P. (2013a). Ordering properties of order statistics from heterogeneous populations: A review with an emphasis on some recent developments. *Probability in the Engineering and Informational Sciences* **27**, 403–443. MR3150103 https://doi.org/10.1017/S0269964813000156

Balakrishnan, N. and Zhao, P. (2013b). Hazard rate comparison of parallel systems with heterogeneous gamma components. *Journal of Multivariate Analysis* **113**, 153–160. MR2984362 https://doi.org/10.1016/j.jmva.2011.05.001

Belzunce, F., Martínez-Riquelme, C. and Mulero, J. (2016). *An Introduction to Stochastic Orders*. London: Academic Press. MR3430823 https://doi.org/10.1016/B978-0-12-803768-3.00001-6

Cheng, K. W. (1977). Majorization: Its extensions and preservation theorems. Tech. Rep. No. 121, Department of Statistics, Stanford University, Stanford, CA.

Cordeiro, G. M., Ortega, E. M. M. and Silva, G. O. (2009). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation* **81**, 827–842. MR2806928 https://doi.org/10.1080/00949650903517874

Gupta, R. C., Gupta, P. L. and Gupta, R. D. (1998). Modeling failure time data by Lehmann alternatives. *Communications in Statistics Theory and Methods* **27**, 887–904. MR1613497 https://doi.org/10.1080/03610929808832134

Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zeland Journal of Statistics* **41**, 173–188. MR1705342 https://doi.org/10.1111/1467-842X.00072

Khaledi, B., Farsinezhad, S. and Kochar, S. C. (2011). Stochastic comparisons of order statistics in the scale model. *Journal of Statistical Planning and Inference* **141**, 276–286. MR2719493 https://doi.org/10.1016/j.jspi.2010.06.006

Kochar, S. C. and Torrado, N. (2015). On stochastic comparisons of largest order statistics in the scale model. *Communications in Statistics Theory and Methods* **44**, 4132–4143. MR3406335 https://doi.org/10.1080/03610926.2014.985839

---

*Key words and phrases.* Exponentiated generalized gamma distribution, parallel system, stochastic orders, unordered majorization, weighted majorization.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*. New York: Springer. MR2759813 https://doi.org/10.1007/978-0-387-68276-1

Misra, N. and Misra, A. K. (2013). On comparison of reversed hazard rates of two parallel systems comprising of independent gamma components. *Statistics & Probability Letters* **83**, 1567–1570. MR3048324 https://doi.org/10.1016/j.spl.2013.03.002

Mitrinović, D. S., Pečarić, J. E. and Fink, A. M. (1993). *New Inequalities in Analysis*. Dordrecht: Kluwer Academic Publishers.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure rate data. *IEEE Transactions on Reliability* **42**, 299–302.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. New York: Wiley. MR1889865

Parker, D. S. and Ram, P. (1997). Greed and majorization. Tech. Report, Department of Computer Science, University of California, Los Angeles.

Pečarić, J. A., Proschan, F. and Tong, Y. L. (1992). *Convex Functions, Partial Orderings, and Statistical Applications*. San Diego: Academic Press. MR1162312

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

Zhao, P. and Balakrishnan, N. (2015). Comparisons of largest order statistics from multiple-outlier gamma models. *Methodology and Computing in Applied Probability* **17**, 617–645. MR3377852 https://doi.org/10.1007/s11009-013-9377-0

Zhao, P., Hu, Y. and Zhang, Y. (2015). Some new results on the largest order statistics from multiple-outlier gamma models. *Probability in the Engineering and Informational Sciences* **29**, 597–621. MR3412160 https://doi.org/10.1017/S0269964815000212

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 167–182
https://doi.org/10.1214/18-BJPS407


# On estimating the location parameter of the selected exponential population under the LINEX loss function

**Mohd Arshad and Omer Abdalghani**

*Aligarh Muslim University*

**Abstract.** Suppose that $\pi_1, \pi_2, \dots, \pi_k$ be $k (\geq 2)$ independent exponential populations having unknown location parameters $\mu_1, \mu_2, \dots, \mu_k$ and known scale parameters $\sigma_1, \dots, \sigma_k$. Let $\mu_{[k]} = \max\{\mu_1, \dots, \mu_k\}$. For selecting the population associated with $\mu_{[k]}$, a class of selection rules (proposed by Arshad and Misra [*Statistical Papers* **57** (2016) 605–621]) is considered. We consider the problem of estimating the location parameter $\mu_S$ of the selected population under the criterion of the LINEX loss function. We consider three natural estimators $\delta_{N,1}, \delta_{N,2}$ and $\delta_{N,3}$ of $\mu_S$, based on the maximum likelihood estimators, uniformly minimum variance unbiased estimator (UMVUE) and minimum risk equivariant estimator (MREE) of $\mu_i$'s, respectively. The uniformly minimum risk unbiased estimator (UMRUE) and the generalized Bayes estimator of $\mu_S$ are derived. Under the LINEX loss function, a general result for improving a location-equivariant estimator of $\mu_S$ is derived. Using this result, estimator better than the natural estimator $\delta_{N,1}$ is obtained. We also shown that the estimator $\delta_{N,1}$ is dominated by the natural estimator $\delta_{N,3}$. Finally, we perform a simulation study to evaluate and compare risk functions among various competing estimators of $\mu_S$.

## References

Abughalous, M. M. and Bansal, N. K. (1994). On the problem of selecting the best population in life testing models. *Communications in Statistics Theory and Methods* **23**, 1471–1481. MR1281222 https://doi.org/10.1080/03610929408831334

Abughalous, M. M. and Miescke, K. J. (1989). On selecting the largest success probability with unequal sample sizes. *Journal of Statistical Planning and Inference* **21**, 53–68. MR0995591 https://doi.org/10.1016/0378-3758(89)90018-9

Arshad, M. and Misra, N. (2015a). Selecting the exponential population having the larger guarantee time with unequal sample sizes. *Communications in Statistics Theory and Methods* **44**, 4144–4171. MR3406336 https://doi.org/10.1080/03610926.2014.973526

Arshad, M. and Misra, N. (2015b). Estimation after selection from uniform populations with unequal sample sizes. *American Journal of Mathematical and Management Sciences* **34**, 367–391.

Arshad, M. and Misra, N. (2016). Estimation after selection from exponential populations with unequal scale parameters. *Statistical Papers* **57**, 605–621. MR3557363 https://doi.org/10.1007/s00362-015-0670-6

Arshad, M. and Misra, N. (2017). On estimating the scale parameter of the selected uniform population under the entropy loss function. *Brazilian Journal of Probability and Statistics* **31**, 303–319. MR3635907 https://doi.org/10.1214/16-BJPS314

Arshad, M., Misra, N. and Vellaisamy, P. (2015). Estimation after selection from gamma populations with unequal known shape parameters. *Journal of Statistical Theory and Practice* **9**, 395–418. MR3311110 https://doi.org/10.1080/15598608.2014.912601

Brewster, J. F. and Zidek, Z. V. (1974). Improving on equivarient estimators. *The Annals of Statistics* **2**, 21–38. MR0381098

Gangopadhyay, A. K. and Kumar, S. (2005). Estimating average worth of the selected subset from two-parameter exponential populations. *Communications in Statistics Theory and Methods* **34**, 2257–2267. MR2209714 https://doi.org/10.1080/03610920500257220

Kumar, S., Mahapatra, A. K. and Vellaisamy, P. (2009). Reliability estimation of the selected exponential populations. *Statistical Probability Letters* **79**, 1372–1377. MR2537512 https://doi.org/10.1016/j.spl.2009.02.012

*Key words and phrases.* Exponential populations, estimation after selection, equivariant estimators, LINEX loss function, inadmissible estimators, UMRU estimator.

Meena, K. R., Arshad, M. and Gangopadhyay, A. K. (2018). Estimating the parameter of selected uniform population under the squared log error loss function. *Communications in Statistics Theory and Methods* **47**, 1679–1692. MR3766676 https://doi.org/10.1080/03610926.2017.1324986

Misra, N., Anand, R. and Singh, H. (1998). Estimation after subset selection from exponential populations: Location parameter case. *American Journal Mathematical Managemment Sciences* **18**, 291–326. MR1678490 https://doi.org/10.1080/01966324.1998.10737468

Misra, N. and Arshad, M. (2014). Selecting the best of two gamma populations having unequal shape parameters. *Statistical Methodology* **18**, 41–63. MR3151863 https://doi.org/10.1016/j.stamet.2013.08.008

Misra, N. and Dhariyal, I. D. (1994). Non-minimaxity of natural decision rules under heteroscedasticity. *Statistics & Decisions* **12**, 79–89. MR1292657

Misra, N. and Singh, G. N. (1993). On the UMVUE for estimating the parameter of the selected exponential population. *Journal of Indian Statistical Association* **31**, 61–69. MR1240855

Misra, N. and van der Meulen, E. C. (2001). On estimation following selection from nonregular distributions Vol. 30, 2543–2561. MR1877355 https://doi.org/10.1081/STA-100108447

Nematollahi, N. and Jozani, M. J. (2016). On risk unbiased estimation after selection. *Brazilian Journal of Probability and Statistics* **30**, 91–106. MR3453516 https://doi.org/10.1214/14-BJPS259

Nematollahi, N. and Motamed-Shariati, F. (2012). Estimation of the parameter of the selected uniform population under the entropy loss function. *Journal of Statistical Planning and Inference* **142**, 2190–2202. MR2903422 https://doi.org/10.1016/j.jspi.2012.01.016

Nematollahi, N. and Pagheh, A. (2017). Estimation of the location parameter and the average worth of the selected subset of two parameter exponential populations under LINEX loss function. *Communications in Statistics Theory and Methods* **46**, 3901–3914. MR3590846 https://doi.org/10.1080/03610926.2015.1076472

Parsian, A. and Farsipour, N. S. (1999). Estimation of the mean of the selected population under asymmetric loss function. *Metrika* **50**, 89–107. MR1745751 https://doi.org/10.1007/s001840050037

Risko, K. J. (1985). Selecting the better binomial population with unequal sample sizes. *Communications in Statistics Theory and Methods* **14**, 123–158. MR0788790 https://doi.org/10.1080/03610928508828900

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In *Studies in Bayesian Econometric and Statistics in Honor of Leonard J. Savage*, 195–208.

Vellaisamy, P. (1996). A note on the estimation of the selected scale parameters. *Journal of Statistical Planning and Inference* **55**, 39–46. MR1423957 https://doi.org/10.1016/0378-3758(95)00178-6

Vellaisamy, P. (2009). A note on unbiased estimation following selection. *Statistical Methodology* **6**, 389–396. MR2751081 https://doi.org/10.1016/j.stamet.2008.12.001

Vellaisamy, P., Kumar, S. and Sharma, D. (1988). Estimating the mean of the selected uniform population. *Communications in Statistics Theory and Methods* **17**, 3447–3475. MR0964346 https://doi.org/10.1080/03610928808829814

Vellaisamy, P. and Punnen, A. P. (2002). Improved estimators for the selected location parameters. *Statistical Papers* **43**, 291–299. MR1903550 https://doi.org/10.1007/s00362-002-0102-2

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 183–187
https://doi.org/10.1214/18-BJPS419


# A note on the "L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications"

**Saralees Nadarajah and Yuancheng Si**

*School of Mathematics, University of Manchester*

**Abstract.** Da Paz, Balakrishnan and Bazan [Braz. J. Probab. Stat. **33** (2019), 455–479] introduced the L-logistic distribution, studied its properties including estimation issues and illustrated a data application. This note derives a closed form expression for moment properties of the distribution. Some computational issues are discussed.

## References

Alhennawi, H. R., El Ayadi, M. M. H., Ismail, M. H. and Mourad, H. A. M. (2016). Closed-form exact and asymptotic expressions for the symbol error rate and capacity of the $H$-function fading channel. *IEEE Transactions on Vehicular Technology* **65**, 1957–1974.

Da Paz, R. F., Balakrishnan, N. and Bazan, J. L. (2019). L-Logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. *Brazilian Journal of Probability and Statistics* **33**, 455–479.

El Ayadi, M. M. H. and Ismail, M. H. (2014). Novel closed-form exact expressions and asymptotic analysis for the symbol error rate of single- and multiple-branch MRC and EGC receivers over $\alpha - \mu$ fading. *IEEE Transactions on Vehicular Technology* **63**, 4277–4291.

Ji, Z., Dong, C., Wang, Y. and Lu, J. (2014). On the analysis of effective capacity over generalized fading channels. In *Proceedings of the 2014 IEEE International Conference on Communications*. https://doi.org/10.1109/ICC.2014.6883613. https://doi.org/10.1109/ICC.2014.6883613

Kilbas, A. A., Srivastava, H. M. and Trujillo, J. J. (2006). *Theory and Applications of Fractional Differential Equations*. Amsterdam: Elsevier. MR2218073

Mathai, A. M. and Saxena, R. K. (1978). *The $H$-Function with Applications in Statistics and Other Disciplines*. New York: Wiley. MR0513025

Mathai, A. M., Saxena, R. K. and Haubold, H. J. (2010). *The $H$ Function: Theory and Applications*. New York: Springer. MR2562766 https://doi.org/10.1007/978-1-4419-0916-9

Saxena, R. K. (1977). On the $H$ function of $n$ variables. *Kyungpook Mathematical Journal* **17**, 221–226. MR0463525

Srivastava, H. M., Gupta, K. C. and Goyal, S. P. (1982). *The $H$-Functions of One and Two Variables with Applications*. New Delhi: South Asian Publishers. MR0691138

Wright, E. M. (1935). The asymptotic expansion of the generalized hypergeometric function. *Journal of the London Mathematical Society* **10**, 286–293. MR0003876 https://doi.org/10.1112/plms/s2-46.1.389

You, M., Sun, H., Jiang, J. and Zhang, J. (2017). Unified framework for the effective rate analysis of wireless communication systems over MISO fading channels. *IEEE Transactions on Communications* **65**, 1775–1785.

*Key words and phrases.* L-logistic distribution, moments, Wright generalized hypergeometric function.

*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 188–201
https://doi.org/10.1214/18-BJPS408


# $W^{1,p}$-Solutions of the transport equation by stochastic perturbation

**David A. C. Mollinedo**

*Universidade Tecnológica Federal do Paraná*

**Abstract.** We consider the stochastic transport equation with a possibly unbounded Hölder continuous vector field. Well-posedness is proved, namely, we show existence, uniqueness and strong stability of $W^{1,p}$-weak solutions.

## References

Alberti, G., Bianchini, S. and Crippa, G. (2010). Divergence-free vector fields in $\mathbb{R}^2$. *Journal of Mathematical Sciences* **170**, 283–293. MR2752636

Ambrosio, L. (2004). Transport equation and Cauchy problem for *BV* vector fields. *Inventiones Mathematicae* **158**, 227–260. MR2096794

Ambrosio, L. and Crippa, G. (2014). Continuity equations and ODE fows with non-smooth velocity, Lecture notes of a course given at HeriottWatt University, Edinburgh. *Proceedings of the Royal Society of Edinburgh Section A Mathematics* **144**, 1191–1244. MR3283066

Catuogno, P. and Olivera, C. (2013). $L^p$-Solutions of the stochastic transport equation. *Random Operators and Stochastic Equations* **21**, 125–134. MR3068412

Colombini, F., Luo, T. and Rauch, J. (2004). Nearly Lipschitzean divergence-free transport propagates neither continuity nor BV regularity. *Communications in Mathematical Sciences* **2**, 207–212. MR2119938

Dafermos, C. M. (2010). *Hyperbolic Conservation Laws in Continuum Physics*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **325**. Berlin: Springer. MR2574377

De Lellis, C. (2007). *Ordinary Differential Equations with Rough Coefficients and the Renormalization Theorem of Ambrosio*. *Bourbaki Seminar* 1–26. MR2487734

DiPerna, R. and Lions, P. L. (1989). Ordinary differential equations, transport theory and Sobolev spaces. *Inventiones Mathematicae* **98**, 511–547. MR1022305

Fedrizzi, E. and Flandoli, F. (2013). Noise prevents singularities in linear transport equations. *Journal of Functional Analysis* **264**, 1329–1354. MR3017266

Fedrizzi, E., Neves, W. and Olivera, C. (2018). On a class of stochastic transport equations for $L^2_{\text{loc}}$ vector fields. *Annali Della Scuola Normale Superiore Di Pisa Classe Di Scienze* **18**, 397–419. MR3801283 https://doi.org/10.2422/2036-2145.201512\protect\T1\textunderscore008

Flandoli, F., Gubinelli, M. and Priola, E. (2010a). Well-posedness of the transport equation by stochastic perturbation. *Inventiones Mathematicae* **180**, 1–53. MR2593276

Flandoli, F., Gubinelli, M. and Priola, E. (2010b). Flow of diffeomorphisms for SDEs with unbounded Hölder continuous drift. *Bulletin Des Sciences Mathématiques* **134**, 405–422. MR2651899

Flandoli, F., Gubinelli, M. and Priola, E. (2012). Remarks on the stochastic transport equation with Hölder drift. *Rendiconti del Seminario Matematico. Universita e Politecnico Torino* **70**, 53–73. MR3305566

Hauray, M. (2003). On two-dimensional Hamiltonian transport equations with $L^p_{\text{loc}}$ coefficients. *Annales de L'Institut Henri Poincaré Analyse Non Linéaire* **20**, 625–644. MR1981402

Kunita, H. (1984a). Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'Été de Probabilités de Saint-Flour XII-1982*, 143–303. Berlin: Springer. MR0876080

Kunita, H. (1984b). First order stochastic partial differential equations. In *Stochastic Analysis (Katata/Kyoto, 1982). North-Holland Math. Library* **32**, 249–269. MR0780761

Lions, P. L. (1996). *Mathematical Topics in Fluid Mechanics, Vol. I: Incompressible Models*. *Oxford Lecture Series in Mathematics and Its Applications* **3**. New York: Oxford University Press. MR1422251

Lions, P. L. (1998). *Mathematical Topics in Fluid Mechanics, Vol. II: Compressible Models*. *Oxford Lecture Series in Mathematics and Its Applications* **10**. New York: Oxford University Press. MR1637634

Mollinedo, D. A. C. and Olivera, C. (2017a). Stochastic continuity equation with non-smooth velocity. *Annali di Matematica Pura ed Applicata. Series IV* **196**, 1669–1684. MR3694739

*Key words and phrases.* Stochastic transport equation, stochastic characteristic method, regularization.

Mollinedo, D. A. C. and Olivera, C. (2017b). Well-posedness of the stochastic transport equation with unbounded drift. *Bulletin of the Brazilian Mathematical Society, New Series* **48**, 663–677. MR3735753

Neves, W. and Olivera, C. (2015). Wellposedness for stochastic continuity equations with Ladyzhenskaya–Prodi–Serrin condition. *NoDEA Nonlinear Differential Equations and Applications* **22**, 1247–1258. MR3399177

Neves, W. and Olivera, C. (2016). Stochastic continuity equations—A general uniqueness result. *Bulletin of the Brazilian Mathematical Society, New Series* **47**, 631–639. MR3514426

## 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/