Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 36 • Number 3 • September 2022

Brazilian Journal of Probability and Statistics

Volume 36 • Number 3 • September 2022

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 36, Number 3, September 2022. Published quarterly by the Brazilian Statistical Association.

POSTMASTER:

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

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

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 421–440
https://doi.org/10.1214/22-BJPS533

Bucket plot: A visual tool for skewness and kurtosis comparisons

Fernanda De Bastiani[1,a], Dimitrios M. Stasinopoulos[2,c], Robert A. Rigby[2,d], Gillian Z. Heller[3,4,e] and Lucas A. Silva[1,b]

[1]*Statistics Department, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235—Cidade Universitária, Recife—Prince Edward Island, Brazil, 50670-901,* [a]*fernanda.bastiani@ufpe.br,* [b]*lucas.asilva5@ufpe.br*

[2]*STORM, London Metropolitan University, 166-220 Holloway Rd, London N7 8DB, United Kingdom,* [c]*d.stasinopoulos@londonmet.ac.uk,* [d]*r.rigby@londonmet.ac.uk*

[3]*NHMRC Clinical Trials Centre, University of Sydney, Locked bag 77, Camperdown New South Wales 1450, Austrália,* [e]*gillian.heller@sydney.edu.au*

[4]*School of Mathematical and Physical Sciences, Macquarie University, Sydney New South Wales 2109, Australia*

Abstract. This study introduces the bucket plot, a visual tool to detect skewness and kurtosis in a continuously distributed random variable. The plot can be applied to both moment and centile skewness and kurtosis. The bucket plot is used to detect skewness and kurtosis either in a response variable, or in the residuals from a fitted model as a diagnostic tool by which to assess the adequacy of a fitted distribution to the response variable regarding skewness and kurtosis. We demonstrate the bucket plot in nine simulated skewness and kurtosis scenarios, and the usefulness of the plot is shown in a real-data situation.

References

Augustin, N. H., Sauleaub, E.-A. and Wood, S. N. (2012). On quantile quantile plots for generalized linear models. *Computational Statistics & Data Analysis* **56**, 2404–2409. MR2910057 https://doi.org/10.1016/j.csda.2012.01.026

Balanda, K. P. and MacGillivray, H. L. (1990). Kurtosis and spread. *Canadian Journal of Statistics* **18**, 17–30. MR1058911 https://doi.org/10.2307/3315414

Brys, G., Hubert, M. and Struyf, A. (2003). A comparison of some new measures of skewness. Developments in Robust Statistics. 98–113. MR1977469

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* **11**, 1305–1319.

Cullen, A. C. and Frey, H. C. (1999). *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Berlin: Springer.

De Bastiani, F., Rigby, R. A., Stasinopoulos, D. M. and Heller, G. Z. (2019). A skewness and kurtosis comparison for continuous distributions. In *Proceedings of the 34th International Workshop on Statistical Modelling Volume I: International Conference, IWSM 2019*. *Guimarães, Portugal*.

De Bastiani, F., Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z. and Silva, L. A. (2022). Supplement to "Bucket plot: A visual tool for skewness and kurtosis comparisons." https://doi.org/10.1214/22-BJPS533SUPP

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 236–244.

Fernandez, C., Osiewalski, J. and Steel, M. J. F. (1995). Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association* **90**, 1331–1340. MR1379475

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* **93**, 359–371. MR1614601 https://doi.org/10.2307/2669632

Horsewell, R. and Looney, S. (1993). Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality. *Communications in Statistics Simulation and Computation* **22**, 437–459.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis* **52**, 5186–5201. MR2526585 https://doi.org/10.1016/j.csda.2007.11.008

Key words and phrases. Centiles and moments, GAMLSS, leptokurtic, platykurtic, residual diagnostic plot, skew exponential power type 3.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* **6**, 255–259. MR0615323 https://doi.org/10.1016/0165-1765(80)90024-5

Joanes, D. N. and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society Series D The Statistician* **47**, 183–189.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–176. MR0033994 https://doi.org/10.1093/biomet/36.1-2.149

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika* **96**, 761–780. MR2564489 https://doi.org/10.1093/biomet/asp053

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. *Lecture Notes in Statistics* **9**. New York: Springer. MR0648107

Kim, T. H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* **1**, 56–73.

Lopatatzidis, A. and Green, P. J. (2000). Private Communication.

MacGillivray, H. (1986). Skewness and asymmetry: Measures and orderings. *The Annals of Statistics* **14**, 994–1011. MR0856802 https://doi.org/10.1214/aos/1176350046

MacGillivray, H. and Balanda, K. (1988). The relationships between skewness and kurtosis. *Australian Journal of Statistics* **30**, 319–337. MR0991198 https://doi.org/10.1111/j.1467-842x.1988.tb00626.x

Mavrodiev, H. (2019). London bike sharing dataset. https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset.

McDonald, J. B. (1984). Some generalized functions for the size distributions of income. *Econometrica* **52**, 647–663.

McDonald, J. B. and Xu, Y. J. (1995). A generalisation of the beta distribution with applications. *Journal of Econometrics* **66**, 133–152.

Nolan, J. P. (2012). *Stable Distributions—Models for Heavy Tailed Data*. Boston: Birkhauser. In progress, Chapter 1 online at academic2.american.edu/~jpnolan. MR4230105 https://doi.org/10.1007/978-3-030-52915-4

Rigby, R. A. and Stasinopoulos, D. M. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box Cox Power Exponential distribution. *Statistics in Medicine* **23**, 3053–3076. MR2252360 https://doi.org/10.1191/1471082X06st122oa

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics* **54**, 507–554. MR2137253 https://doi.org/10.1111/j.1467-9876.2005.00510.x

Rigby, R. A. and Stasinopoulos, D. M. (2006). Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. *Statistical Modelling* **6**, 209. MR2252360 https://doi.org/10.1191/1471082X06st122oa

Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z. and De Bastiani, F. (2019). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Boca Raton: CRC Press. MR3799717 https://doi.org/10.1177/1471082X18759144

Royston, P. and Wright, E. (2000). Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine* **19**, 2943–2962.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V. and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Boca Raton: CRC Press. MR3799717 https://doi.org/10.1177/1471082X18759144

van Buuren, S. and Fredriks, M. (2001). Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine* **20**, 1259–1277.

van Zwet, W. R. (1964a). Convex transformations: A new approach to skewness and kurtosis. *Statistica Neerlandica* **18**, 433–441. MR0175217 https://doi.org/10.1111/j.1467-9574.1964.tb00530.x

van Zwet, W. R. (1964b). *Convex Transformations of Random Variables*. *Mathematica Centre Tracts* **7**. Amsterdam: Mathematisch Centrum. MR0176511

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**, 180–185.

Zhang, Y. and Mi, Z. (2018). Environmental benefits of bike sharing: A big data-based analysis. *Applied Energy* **220**, 296–301.

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 441–462
https://doi.org/10.1214/22-BJPS534

On outliers detection and prior distribution sensitivity in standard skew-probit regression models

Fabiano R. Coelho[1,2,a], **Cibele M. Russo**[1,b] **and Jorge L. Bazán**[1,c]

[1]*Department of Applied Mathematics and Statistics, University of São Paulo, São Carlos, Brazil,*
[a]*fabiano_icmc@usp.br,* [b]*cibele@icmc.usp.br,* [c]*jlbazan@icmc.usp.br*
[2]*Department of Statistics, University of São Carlos, São Carlos, Brazil*

Abstract. Regression models with probit and logit link functions are the most frequently used for binary response variables. However, traditional approaches may not be adequate when data are unbalanced. This paper deals with standard skew-probit regression models. Parameters were estimated through a new Bayesian approach which consists of the use of Hamiltonian Monte Carlo (HMC) and the original likelihood function. Simulation studies assessed the efficiency of the estimation method and the sensitivity of prior distributions for parameters related to asymmetry calculating the RMSE (root mean square error). The proposed estimation method was compared when used for detecting outliers. The results show that the proposed method is more efficient than INLA and is successful in the recovery of true parameter values. The sensitivity study enabled the proposal of a new prior distribution configuration for the asymmetry parameter, and the randomized quantile residual proved to be more suitable for detecting outliers. The methodology was applied to a diabetes dataset towards illustrating the results.

References

Agresti, A. (2003). *Categorical Data Analysis, Vol. 482*. New York: Wiley. MR1914507 https://doi.org/10.1002/0471249688

Albert, J. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. In *Biometrika, Vol. 82*, 747–769. London: Oxford University Press. MR1380812 https://doi.org/10.1093/biomet/82.4.747

Atkinson, A. C. and Plots, T. (1985). *Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. *Oxford Statistical Science Series*. Oxford: Oxford University Press.

Basu, S. and Mukhopadhyay, S. (2000). Bayesian analysis of binary regression using symmetric and asymmetric links. *Sankhyā: The Indian Journal of Statistics, Series B. JSTOR*, 372–387. MR1834162

Bayes, C. L. and Branco, M. D. E. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics. JSTOR*, 141–163. MR2397043

Bazán, J. L., et al (2006). A skew item response model. Bayesian analysis. *International Society for Bayesian Analysis* **1**, 861–892. MR2282209 https://doi.org/10.1214/06-BA128

Bazán, J. L., et al (2014). Extensions of the skew-normal ogive item response model. *Brazilian Journal of Probability and Statistics* **28**, 1–23. MR3165426 https://doi.org/10.1214/12-BJPS191

Bazán, J. L., Bolfarine, H. and Branco, M. D. (2010). A framework for skew-probit links in binary regression. *Communications in Statistics Theory and Methods* **39**, 678–697. MR2745312 https://doi.org/10.1080/03610920902783849

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC. MR3203203

Chen, M. H. (2004). Skewed link models for categorical response data. In *kew-Elliptical Distributions and Their Applications*, 151–172. London: Chapman and Hall/CRC. MR2155327

Chen, M. H., et al (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* **94**, 1172–1186. MR1731481 https://doi.org/10.2307/2669933

Chen, M. H., Dey, D. K. and Shao, Q. M. (2001). Bayesian analysis of binary data using skewed logit models. In *Calcutta Statistical Association Bulletin, Vol. 51*, 11–30. New Delhi, India: SAGE Publications Sage India. MR1892782 https://doi.org/10.1177/0008068320010102

Key words and phrases. Skew-probit regression models, outliers detection, HMC estimation method, unbalanced binary data.

Dey, D. K. and Birmiwal, L. R. (1994). Robust Bayesian analysis using divergence measures. *Statistics & Probability Letters* **20**, 287–294. MR1294607 https://doi.org/10.1016/0167-7152(94)90016-7

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 319–392.

Farias, R. B. A., et al (2012). Latent residual analysis in binary regression with skewed link. *Brazilian Journal of Probability and Statistics* **26**, 344–357. MR2949083 https://doi.org/10.1214/11-BJPS143

Farias, R. B. A. and Branco, M. D. (2011). Efficient algorithms for Bayesian binary regression model with skew-probit link. In *Recent Advances in Biostatistics: False Discovery Rates, Survival Analysis, and Related Topics*, 143–168. Singapore: World Scientific.

Friedrich, L. and Evgenia, D. (2010). mlbench: Machine Learning Benchmark Problems. R package version 2.1-1.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall/CRC. MR1385925

Gelman, A., Lee, D. and Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics* **40**, 530–543.

Gelman, A., Rubin, D. B., et al (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472. MR3264555 https://doi.org/10.1214/14-STS489

Hailpern, S. M. and Visintainer, P. F. (2003). Odds ratios and logistic regression: Further examples of their use and interpretation. *Stata Journal* **3**, 213–225.

Henderson, N. C., Louis, T. A., Wang, C. and Varadhan, R. (2016). Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. *Health Services and Outcomes Research Methodology* **4**, 213–233.

Hoffman, 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

King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis* **9**, 137–163.

Lee, D. and Sinha, S. (2019). Identifiability and bias reduction in the skew-probit model for a binary response. *Journal of Statistical Computation and Simulation* **89**, 1621–1648. MR3934749 https://doi.org/10.1080/00949655.2019.1590579

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian Biostatistics*. New York: Wiley. MR3236827 https://doi.org/10.1002/9781119942412

Meltzer, E. B., et al (2011). Bayesian probit regression model for the diagnosis of pulmonary fibrosis: Proof-of-principle. BMC medical genomics. *BioMed Central* **4**, 1–13.

Naranjo, L., Pérez, C. J. and Martín, J. (2019). Skewed link-based regression models for misclassified binary data. *Revista de la Real Academia de Ciencias Exactas, Físicas Y Naturales Serie A, Matemáticas* **113**, 1585–1599. MR3942416 https://doi.org/10.1007/s13398-018-0571-3

Neal, R. M., et al (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* **2**, 2. MR2858447

Newman, D. J., et al. (1998). UCI Repository of machine learning databases. Available from. http://www.ics.uci.edu/~mlearn/MLRepository.htm.

Peng, F. and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics* **23**, 199–213.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics. JSTOR.*, 761–768.

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, Statistical Methodology* **71**, 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

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, Statistical Methodology* **64**, 583–639. MR1979380 https://doi.org/10.1111/1467-9868.00353

Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2. Available from. http://mc-stan.org/.

Van der Paal, B. (2014). A comparison of different methods for modelling rare events data. PhD thesis, Ghent University.

Van Niekerk, J. and Rue, H. (2021). Skewed probit regression–identifiability, contraction and reformulation. *REVSTAT Statistical Journal* **19**, 1–22.

Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**, 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

Wang, X., et al (2010). Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *Annals of Applied Statistics* **4**, 2000–2023. MR2829944 https://doi.org/10.1214/10-AOAS354

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

Yan, G. and Sedransk, J. (2010). A note on Bayesian residuals as a hierarchical model diagnostic technique. *Statistical Papers* **51**, 1. MR2556583 https://doi.org/10.1007/s00362-007-0111-2

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 463–481
https://doi.org/10.1214/22-BJPS535

Additive ratio type exponential estimator of finite population mean of sensitive variable using non-sensitive auxiliary information based on optional randomized response model

Lovleen Kumar Grover[a] and Amanpreet Kaur[b]

Department of Mathematics, Guru Nanak Dev University, Amritsar-143005, Punjab, India,
[a]*lovleen_2@yahoo.co.in,* [b]*amanpreetk321@gmail.com*

Abstract. The appropriate use of auxiliary information in sample surveys increases the efficiency of estimator for parameter of interest. In this paper, we have proposed an exponential type estimator for the population mean of a sensitive study variable based on an optional randomized response model by using the known information on a non-sensitive auxiliary variable. Expressions for the bias and the mean square error (MSE) of the proposed estimator are derived, up to first order of approximation. For this proposed estimator, efficiency comparisons with the existing estimators have been carried out both theoretically and numerically. It has been shown that our proposed estimator perform better than the existing estimators based on the same optional randomized response model even for the small correlation between auxiliary variable and study variable. To support the results obtained, we have also studied the performance of the proposed exponential estimator using simulation technique.

References

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

Diana, G. and Perri, P.F. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers* **52**, 633–650. MR2821061 https://doi.org/10.1007/s00362-009-0273-1

Eichhorn, B. H. and Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and inference* **7**, 307–316.

Greenberg, B. G., Abul-Ela Abdel-Latif, A., Simmons, W. R. and Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association* **64**, 520–539. MR0247719

Grover, L.K. and Kaur, A. (2019). An Efficient Scrambled Estimator of Population Mean of Quantitative Sensitive Variable Using General Linear Transformation of Non-sensitive Auxiliary Variable. *Communications in Mathematics and Statistics* **7**, 401–415. MR4024612 https://doi.org/10.1007/s40304-018-0146-9

Grover, L.K. and Kaur, A. (2020). Corrected estimator of sensitive population proportion using unknown repeated trials in the unrelated question randomized response model. *Biometrical Journal* **62**, 1337–1342. MR4149484 https://doi.org/10.1080/03610926.2019.1628992

Gupta, S., Gupta, B. and Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and inference* **100**, 239–247. MR1877192 https://doi.org/10.1016/S0378-3758(01)00137-9

Gupta, S., Kalucha, G. and Shabbir, J. (2017). A regression estimator for finite population mean of a sensitive variable using an optional randomized response model. *Communications in Statistics-Simulation and Computation* **46**, 2393–2405. MR3625289 https://doi.org/10.1080/03610918.2015.1044614

Gupta, S. and Shabbir, J. (2004). Sensitivity estimation for personal interview survey questions. *Statistica* **64**, 643–653. MR2324023

Gupta, S., Shabbir, J. and Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference* **140**, 2870–2874. MR2651974 https://doi.org/10.1016/j.jspi.2010.03.010

Key words and phrases. Auxiliary information, bias, mean square error, randomized response technique, ratio estimator, exponential estimator, optional randomized response model.

Gupta, S., Mehta, S., Shabbir, J. and Dass, B. K. (2013). Generalized scrambling in quantitative optional randomized response models. *Communications in Statistics-Theory and Methods* **42**, 4034–4042. MR3170981 https://doi.org/10.1080/03610926.2011.638427

Gupta, S., Shabbir, J. Sousa, R. and Corte-Real, P. (2016). Improved exponential type estimators of the mean of a sensitive variable in the presence of nonsensitive auxiliary information. *Communications in Statistics-Simulation and Computation* **45**, 3317–3328. MR3532410 https://doi.org/10.1080/03610918.2014.941487

Huang, K.-C. (2010). Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. *Metrika* **71**, 341–352. MR2610002 https://doi.org/10.1007/s00184-009-0234-7

Hussain, Z. and Shabbir, J. (2009). Improved estimation procedures for the mean of sensitive variable using randomized response model. *Pakistan Journal of Statistics* **25**, 205–220. MR2668157

Kalucha, G., Gupta, S. and Dass, B. K. (2015). Ratio estimation of finite population mean using optional randomized response models. *Journal of Statistical Theory and Practice* **9**, 633–645. MR3339774 https://doi.org/10.1080/15598608.2014.977982

Kim, J.-M. and Warde, W. D. (2005). A mixed randomized response model. *Journal of Statistical Planning and Inference* **133**, 211–221. MR2162576 https://doi.org/10.1016/j.jspi.2004.03.011

Koyuncu, N., Gupta, S. and Sousa, R. (2014). Exponential-type estimators of the mean of a sensitive variable in the presence of nonsensitive auxiliary information. *Communications in statistics-Simulation and Computation* **43**, 1583–1594. MR3215794 https://doi.org/10.1080/03610918.2012.737492

Mangat, N. S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika* **77**, 439–442. MR1064823 https://doi.org/10.1093/biomet/77.2.439

Noor-Ul-Amin, M., Mushtaq, N. and Hanif, M. (2018). Estimation of mean using generalized optional scrambled responses in the presence of non-sensitive auxiliary variable. *Journal of Statistics and Management Systems* **21**, 287–304.

Özgül, N. and Çingi, H. (2017). A new estimator based on auxiliary information through quantitative randomized response techniques. *Journal of Modern Applied Statistical Methods* **16**, 364–387.

Peng, R. D. (2019). *R programming for data science*. Victoria, BC, Canada: Leanpub.

Ryu, J.-B., Kim, J.-M., Heo, T.-Y. and Park, C.G (2006). On stratified randomized response sampling. *Model Assisted Statistics and Applications* **1**, 31–36. MR2339080

Saha, A. (2008). A randomized response technique for quantitative data under unequal probability sampling. *Journal of Statistical Theory and Practice* **2**, 589–596. MR2655433 https://doi.org/10.1080/15598608.2008.10411897

Saleem, I., Sanaullah, A. and Hanif, M. (2019). Double-sampling regression-cum-exponential estimator of the mean of a sensitive variable. *Mathematical Population Studies: An international Journal of Mathematical Demography* **26**, 163–182. MR3993650 https://doi.org/10.1080/08898480.2019.1565273

Sanaullah, A., Saleem, I., Gupta, S. and Hanif, M. (2020). Mean estimation with generalized scrambling using two-phase sampling. *Communications in Statistics-Simulation and Computation*, 1–15. https://doi.org/10.1080/03610918.2020.1778032

Shah, S.F., Hussain, Z. and Cheema, S.A. (2020). Combining answers to direct and indirect questions: An implementation of Kuks randomized response model. *Communications in Statistics-Theory and Methods* **49**, 3933–3949. MR4121931 https://doi.org/10.1080/03610926.2019.1593458

Shahzad, U., Perri, P.F. and Hanif, M. (2019). A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information. *Communications in Statistics-Simulation and Computation* **48**, 2566–2585. MR4001221 https://doi.org/10.1080/03610918.2018.1458133

Singh, S. (2003). *Advanced Sampling Theory With Applications*. Dordrecht: Springer. MR2032219 https://doi.org/10.1007/978-94-007-0789-4

Sousa, R., Shabbir, J., Corte-Real, P. and Gupta, S. (2010). Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistical Theory and Practice* **4**, 495–507. MR2758690 https://doi.org/10.1080/15598608.2010.10411999

Vishwakarma, G. K. and Singh, N. (2021). Computing the effect of measurement errors under additive scramble response of the sensitive variable. *Journal of Computational and Applied Mathematics* **395**, 113593. MR4250647 https://doi.org/10.1016/j.cam.2021.113593

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60**, 63–69.

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 482–491
https://doi.org/10.1214/22-BJPS536

Estimation of trace-variogram using Legendre–Gauss quadrature

Gilberto Sassi[1,a,b] **and Chang Chiann**[2,c]

[1]*Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, Bahia, Brazil,* [a]*sassi.pereira.gilberto@gmail.com,* [b]*https://gilberto-sassi.github.io*
[2]*Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, São Paulo, São Paulo, Brazil,* [c]*chang@ime.usp.br*

Abstract. Functional Data Analysis is known for its application in several fields of science. In some cases, functional datasets are constituted by spatially indexed curves. The primary goal of this paper is to supply a straightforward and precise approach to interpolate these curves, that is, the aim is to estimate a curve at an unmonitored location. It is proven that the best linear unbiased estimator for this unsampled curve is the solution of a linear system, where the coefficients and the constant terms of the system are formed using a function called trace-variogram. In this paper, we propose using Legendre-Gauss quadrature to estimate the trace-variogram. This estimator's suitable numerical properties are shown in simulation studies for normal and non-normal datasets. Simulation results indicated that the proposed methodology outperforms the established estimation procedure. An `R` package was built and is available at the CRAN repository. The novel estimation methodology is illustrated with a real dataset on temperature curves from 35 weather stations in Canada.

References

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85**, 549–559. MR1665873 https://doi.org/10.1093/biomet/85.3.549

Beyaztas, U. and Shang, H. L. (2019). Forecasting functional time series using weighted likelihood methodology. *Journal of Statistical Computation and Simulation* **89**, 3046–3060. MR4000283 https://doi.org/10.1080/00949655.2019.1650935

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.

Chen, Y., Goldsmith, J. and Ogden, R. T. (2019). Functional data analysis of dynamic PET data. *Journal of the American Statistical Association* **114**, 595–609. MR3963165 https://doi.org/10.1080/01621459.2018.1497495

Fang, K., Zhang, X., Ma, S. and Zhang, Q. (2020). Smooth and locally sparse estimation for multiple-output functional linear regression. *Journal of Statistical Computation and Simulation* **90**, 341–354. MR4038811 https://doi.org/10.1080/00949655.2019.1680676

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Berlin: Springer. MR2229687

Gabrys, R. and Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association* **102**, 1338–1348. MR2412554 https://doi.org/10.1198/016214507000001111

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., Mateu, J. and Delicado, P. (2012). `geofd`: An `R` package for function-valued geostatistical prediction. *Revista Colombiana de Estadística* **35**, 385–407. MR3075152

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.

Khuri, A. I. (2003). *Advanced Calculus with Applications in Statistics, Vol. 486*. New York: Wiley. MR1934984

Key words and phrases. Trace-variogram, kriging, functional data analysis, geostatistics, fourier series, spatial statistics.

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

Lee, W., Miranda, M. F., Rausch, P., Baladandayuthapani, V., Fazio, M., Downs, J. C. and Morris, J. S. (2019). Bayesian semiparametric functional mixed models for serially correlated functional data, with application to glaucoma data. *Journal of the American Statistical Association* **114**, 495–513. MR3963158 https://doi.org/10.1080/01621459.2018.1476242

Menafoglio, A., Guadagnini, A. and Secchi, P. (2014). A Kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. *Stochastic Environmental Research and Risk Assessment* **28**, 1835–1851.

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 Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* **53**, 539–572. MR1125714

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2018) fda: Functional Data Analysis. R package version 2.4.8.

Reyes, A., Giraldo, R. and Mateu, J. (2012). Residual Kriging for functional data. Application to the spatial prediction of salinity curves. *Revista Colombiana de Estadística* **35**, 385–407.

Salazar, E., Giraldo, R. and Porcu, E. (2015). Spatial prediction for infinite-dimensional compositional data. *Stochastic Environmental Research and Risk Assessment* **29**, 1737–1749.

Scott, D. W. (2009). Sturges' rule. *Wiley Interdisciplinary Reviews: Computational Statistics* **1**, 303–306.

Vale, C. D. and Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika* **48**, 465–471. ISSN 1860-0980.

Zamani, A., Hashemi, M. and Haghbin, H. (2019). Improved functional portmanteau tests. *Journal of Statistical Computation and Simulation* **89**, 1423–1436. MR3929230 https://doi.org/10.1080/00949655.2019.1584199

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 492–503
https://doi.org/10.1214/22-BJPS537

Joint mixture quantile regressions and time-to-event analysis

Getachew A. Dagne[a]

College of Public Health, University of South Florida, USA, [a]*gdagne@usf.edu*

Abstract. Growth curve mixture models for longitudinal data are often developed on the conditional mean of a response, focusing only on the central section of the distribution. There is, however, an increasing desire to provide holistic information on different parts of the distribution of the response such as lower and higher quantiles. This article presents quantile regression analysis within the framework of growth curve models by jointly analyzing time to an event and longitudinal data with multiphasic features. The multiphasic patterns are accounted for at different quantiles by modeling heterogeneous growth trajectories which show gradual changes from a declining trend to an increasing trend over time within latent classes. Thus, we assess these important features of longitudinal data using bent-cable models along with a joint modeling of time to event process and response process. The proposed methods are illustrated using a real data set from an AIDS clinical study.

References

Acosta, E. P., Wu, H., Walawander, A., Eron, J., Pettinelli, C., Yu, S., Neath, D., Ferguson, E., Saah, A. J., Kuritzkes, D. R. and Gerber, J. G. (2004). Comparison of two indinavir/ritonavir regimens in treatment-experienced HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes* **37**, 1358–1366.

Bartolucci, F. and Farcomeni, A. (2015). A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics* **71**, 80–89. MR3335352 https://doi.org/10.1111/biom.12224

Bollen, K. A. and Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Approach*. *Wiley Series on Probability and Mathematical Statistics*. New Jersey: John Wiley & Sons. MR2184502

Chiua, G., Lockharta, R. and Routledgea, R. (2006). Bent-cable regression theory and applications. *Journal of the American Statistical Association* **101**, 542–553. MR2256174 https://doi.org/10.1198/016214505000001177

Dagne, G. A. (2018). Joint bent-cable Tobit models for longitudinal and time-to-event data. *Journal of Biopharmaceutical Statistics* **28**, 385–401. MR3713660 https://doi.org/10.1002/sim.7429

DeGruttola, V. and Tu, X. M. (1994). Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

Ferrer, E. and McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science* **19**, 149–154.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Boca Raton: Chapman and Hall/CRC. MR2027492

Henderson, R., Diggle, P. J. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

Ho, D. D., Neumann, A. U., Perelson, A. S., et al (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126.

Khan, S. A., Chiu, G. and Dubin, J. A. (2009). Atmospheric concentration of chloroflurocarbons: Addressing the global concern with the longitudinal bent-cable model. *Chance* **22**, 8–17.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation* **81**, 1565–1578. MR2851270 https://doi.org/10.1080/00949655.2010.496117

Lambert, P., Collett, D., Kimber, A. and Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine* **23**, 3177–3192.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, an extensibility. *Statistics and Computing* **10**, 325–337.

Margolick, J. B. J., Gange, S. S. J., Detels, R., O'Gorman, M. R. G., Rinaldo, C. R., et al (2006). Impact of inversion of the CD4/CD8 ratio on the natural history of HIV-1 infection. *Journal of Acquired Immune Deficiency Syndromes* **42**, 620–626.

Key words and phrases. Change-point model, Bayesian inference, latent variable, survival analysis.

Muthen, B. O. and Curra, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods* **2**, 371–402.

Paterson, D. L., Swindells, S., Mohr, J., et al (2000). Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Annals of Internal Medicine* **133**, 21–30.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Sage Publications: Thousand Oaks, CA.

Reed, C. and Yu, K. (2009). A partially collapsed Gibbs sampler for Bayesian quantile regression. Technical report, Department of Mathematical Sciences, Brunel University London. Available at https://bura.brunel.ac.uk/bitstream/2438/3593/1/fulltext.pdf.

Reich, B. J., Bondell, H. D. and Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* **11**, 337–352.

Rizopoulos, D., Verbeke, G., Lesaffre, E. and Vanrenterghem, Y. (2008). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* **64**, 611–619. MR2432435 https://doi.org/10.1111/j.1541-0420.2007.00894.x

Serrano-Villar, S., Gutiérrez, C., Vallejo, A., Hernández-Novoa, B., Díaz, L., et al (2013). The CD4/CD8 ratio in HIV-infected subjects is independently associated with T-cell activation despite long-term viral suppression. *Journal of Infection* **66**, 57–66.

Taylor, J. M. G., Fahey, J. L., Detels, R., et al (1989). CD4 percentage, CD4 number and CD4:CD8 ratio in HIV infection: Which to choose and how to use. *Journal of Acquired Immune Deficiency Syndromes* **2**, 114–124.

Tishler, A., Zang, I. (1981). A new maximum likelihood algorithm for piece-wise regression. *Journal of the American Statistical Association* **76**, 980–987.

Tsiatis, A. A. and Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* **14**, 793–818. MR2087974

Viviani, S., Alfò, M. and Rizopoulos, D. (2014). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing* **24**, 417–427. MR3192265 https://doi.org/10.1007/s11222-013-9378-4

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association* **97**, 955–964. MR1951254 https://doi.org/10.1198/016214502388618744

Wu, L., Liu, W. and Hu, X. J. (2010). Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics* **66**, 327–335. MR2758812 https://doi.org/10.1111/j.1541-0420.2009.01308.x

Ye, W., Lin, X. and Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data– a two-stage regression calibration approach. *Biometrics* **64**, 1238–1246. MR2522273 https://doi.org/10.1111/j.1541-0420.2007.00983.x

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* **54**, 437–447. MR1861390 https://doi.org/10.1016/S0167-7152(01)00124-9

Yuan, Y. and Yin, G. (2011). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66**, 105–114. MR2756696 https://doi.org/10.1111/j.1541-0420.2009.01269.x

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 504–539
https://doi.org/10.1214/22-BJPS538

Unadjusted Langevin algorithm for sampling a mixture of weakly smooth potentials

Dao Nguyen[a]

Department of Mathematics, University of Mississippi, University, Mississippi 38677, USA,
[a]*dxnguyen@olemiss.edu*

Abstract. Discretization of continuous-time diffusion processes is a widely recognized method for sampling. However, it seems to be a considerable restriction when the potentials are often required to be smooth (gradient Lipschitz). This paper studies the problem of sampling through Euler discretization, where the potential function is assumed to be a mixture of weakly smooth distributions and satisfies weakly dissipative. We establish the convergence in Kullback–Leibler (KL) divergence with the number of iterations to reach ϵ-neighborhood of a target distribution in only polynomial dependence on the dimension. We relax the degenerated convex at infinity conditions of (Erdogdu and Hosseinzadeh (2020)) and prove convergence guarantees under Poincaré inequality or non-strongly convex outside the ball. In addition, we also provide convergence in L_β-Wasserstein metric for the smoothing potential.

References

Atchadé, Y. F. (2015). A Moreau–Yosida approximation scheme for a class of high-dimensional posterior distributions. Preprint. Available at arXiv:1505.07072.

Barkhagen, M., Chau, N. H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021). On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case. *Bernoulli* **27**, 1–33. MR4177359 https://doi.org/10.3150/19-BEJ1187

Bobkov, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *The Annals of Probability* **27**, 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

Bolley, F. and Villani, C. (2005). Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse: Mathématiques* **14**, 331–352. MR2172583

Brosse, N., Durmus, A. and Moulines, E. (2018). The promises and pitfalls of stochastic gradient Langevin dynamics. In *Advances in Neural Information Processing Systems* **31**.

Brosse, N., Durmus, A., Moulines, É. and Sabanis, S. (2019). The tamed unadjusted Langevin algorithm. *Stochastic Processes and Their Applications* **129**, 3638–3663. MR3997657 https://doi.org/10.1016/j.spa.2018.10.002

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge: Cambridge University Press. MR2409394 https://doi.org/10.1017/CBO9780511546921

Chatterji, N. S., Diakonikolas, J., Jordan, M. I. and Bartlett, P. L. (2019). Langevin Monte Carlo without smoothness. Preprint. Available at arXiv:1905.13285.

Chau, N. H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021). On stochastic gradient Langevin dynamics with dependent data streams: The fully nonconvex case. *SIAM Journal on Mathematics of Data Science* **3**, 959–986. MR4313846 https://doi.org/10.1137/20M1355392

Chen, Y., Dwivedi, R., Wainwright, M. J. and Yu, B. (2018). Fast MCMC sampling algorithms on polytopes. *The Journal of Machine Learning Research* **19**, 2146–2231. MR3874163

Cheng, X. and Bartlett, P. L. (2018). Convergence of Langevin MCMC in KL-divergence. *The Proceedings of Machine Learning Research* **83**, 186–211. MR3857306

Cheng, X., Chatterji, N. S., Abbasi-Yadkori, Y., Bartlett, P. L. and Jordan, M. I. (2018). Sharp convergence rates for Langevin dynamics in the nonconvex setting. Preprint. Available at arXiv:1805.01648.

Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **79**, 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

Key words and phrases. Langevin Monte Carlo, non-convex sampling, Kullback–Leibler divergence, mixture weakly smooth.

Dalalyan, A. S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and Their Applications* **129**, 5278–5311. MR4025705 https://doi.org/10.1016/j.spa.2019.02.016

Durmus, A. and Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability* **27**, 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

Durmus, A., Moulines, E., et al (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli* **25**, 2854–2882. MR4003567 https://doi.org/10.3150/18-BEJ1073

Durmus, A., Moulines, E. and Pereyra, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau. *SIAM Journal on Imaging Sciences* **11**, 473–506. MR3763089 https://doi.org/10.1137/16M1108340

Dyer, M., Frieze, A. and Kannan, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)* **38**, 1–17. MR1095916 https://doi.org/10.1145/102782.102783

Eberle, A. (2016). Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields* **166**, 851–886. MR3568041 https://doi.org/10.1007/s00440-015-0673-1

Eberle, A., Guillin, A. and Zimmer, R. (2019). Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Transactions of the American Mathematical Society* **371**, 7135–7173. MR3939573 https://doi.org/10.1090/tran/7576

Erdogdu, M. A. and Hosseinzadeh, R. (2020). On the convergence of Langevin Monte Carlo: The interplay between tail growth and smoothness. Preprint. Available at arXiv:2005.13097.

Erdogdu, M. A., Mackey, L. and Shamir, O. (2018). Global non-convex optimization with discretized diffusions. Preprint. Available at arXiv:1810.12361.

Holley, R. and Stroock, D. W. (1986). Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics* **46**, 1159–1194. MR0893137 https://doi.org/10.1007/BF01011161

Hsieh, Y.-P., Kavis, A., Rolland, P. and Cevher, V. (2018). Mirrored Langevin dynamics. In *Advances in Neural Information Processing Systems*, 2878–2887.

Hutzenthaler, M., Jentzen, A. and Kloeden, P. E. (2012). Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability* **22**, 1611–1641. MR2985171 https://doi.org/10.1214/11-AAP803

Kaipio, J. and Somersalo, E. (2006). *Statistical and Computational Inverse Problems* **160**. New York: Springer. MR2102218

Kečkić, J. D. and Vasić, P. M. (1971). Some inequalities for the gamma function. *Publications de l'Institut Mathématique* **11**, 107–114. MR0308446

Ledoux, M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités XXXV*, 167–194. Berlin: Springer. MR1837286 https://doi.org/10.1007/978-3-540-44671-2_13

Lim, D.-Y., Neufeld, A., Sabanis, S. and Zhang, Y. (2021). Non-asymptotic estimates for TUSLA algorithm for non-convex learning with applications to neural networks with ReLU activation function. Preprint. Available at arXiv:2107.08649.

Liu, Y. (2020). The Poincaré inequality and quadratic transportation-variance inequalities. *Electronic Journal of Probability* **25**, 1–16. MR4053901 https://doi.org/10.1214/19-ejp403

Lovas, A., Lytras, I., Rásonyi, M. and Sabanis, S. (2020). Taming neural networks with TUSLA: Non-convex learning via adaptive stochastic gradient Langevin algorithms. Preprint. Available at arXiv:2006.14514.

Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms* **30**, 307–358. MR2309621 https://doi.org/10.1002/rsa.20135

Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N. and Jordan, M. I. (2019). Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences* **116**, 20881–20885. MR4025861 https://doi.org/10.1073/pnas.1820003116

Majka, M. B., Mijatović, A. and Szpruch, Ł. (2020). Nonasymptotic bounds for sampling algorithms without log-concavity. *The Annals of Applied Probability* **30**, 1534–1581. MR4132634 https://doi.org/10.1214/19-AAP1535

Marie-Caroline, C., Denis, K., Emilie, C., Jean-Yves, T. and Jean-Christophe, P. (2019). Preconditioned P-ULA for joint deconvolution-segmentation of ultrasound images. Preprint. Available at arXiv:1903.08111.

McNabb, A. (1986). Comparison theorems for differential equations. *Journal of Mathematical Analysis and Applications* **119**, 417–428. MR0856678 https://doi.org/10.1016/0022-247X(86)90163-0

Neal, R. M. (1993). Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems*, 475–482.

Nesterov, Y. and Spokoiny, V. (2017). Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics* **17**, 527–566. MR3627456 https://doi.org/10.1007/s10208-015-9296-2

Nguyen, D. (2022). Supplement to "Unadjusted Langevin algorithm for sampling a mixture of weakly smooth potentials." https://doi.org/10.1214/22-BJPS538SUPP

Nguyen, D., Dang, X. and Chen, Y. (2021). Unadjusted Langevin algorithm for non-convex weakly smooth potentials. Preprint. Available at arXiv:2101.06369.

Parisi, G. (1981). Correlation functions and computer simulations. *Nuclear Physics B* **180**, 378–384. MR0615176 https://doi.org/10.1016/0550-3213(81)90056-0

Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations* **60**. New York: Springer. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

Polyanskiy, Y. and Wu, Y. (2016). Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory* **62**, 3992–4002. MR3515734 https://doi.org/10.1109/TIT.2016.2562630

Rademacher, L. and Vempala, S. (2008). Dispersion of mass and the complexity of randomized geometric algorithms. *Advances in Mathematics* **219**, 1037–1069. MR2442061 https://doi.org/10.1016/j.aim.2008.06.004

Raginsky, M., Rakhlin, A. and Telgarsky, M. (2017). Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. Preprint. Available at arXiv:1702.03849.

Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*. New York: Springer. MR1707311 https://doi.org/10.1007/978-1-4757-3071-5

Sabanis, S. (2013). A note on tamed Euler approximations. *Electronic Communications in Probability* **18**, 1–10. MR3070913 https://doi.org/10.1214/ECP.v18-2824

Sabanis, S. (2016). Euler approximations with varying coefficients: The case of superlinearly growing diffusion coefficients. *The Annals of Applied Probability* **26**, 2083–2105. MR3543890 https://doi.org/10.1214/15-AAP1140

Sabanis, S. and Zhang, Y. (2019). Higher order Langevin Monte Carlo algorithm. *Electronic Journal of Statistics* **13**, 3805–3850. MR4015336 https://doi.org/10.1214/19-EJS1615

Vempala, S. and Wibisono, A. (2019). Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems*, 8094–8106.

Villani, C. (2008). *Optimal Transport: Old and New* **338**. Berlin: Springer. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

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

Xu, P., Chen, J., Zou, D. and Gu, Q. (2018). Global convergence of Langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems*, 3122–3133.

Yan, M. (2012). Extension of convex function. Preprint. Available at arXiv:1207.0944. MR3331205

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 540–559
https://doi.org/10.1214/22-BJPS539

Cumulative past Fisher information measure and its extensions

Narayanaswamy Balakrishnan[1,a] **and Omid Kharazmi**[2,b]

[1]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada,*
[a]*bala@mcmaster.ca*
[2]*Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Iran,*
[b]*omid.kharazmi@vru.ac.ir*

Abstract. In this work, we define the cumulative past Fisher (CPF) information and the relative cumulative past Fisher (RCRF) information measures for parameter as well as for the distribution function of the underlying random variables. We show that these cumulative past Fisher information measures can be expressed in terms of the reversed hazard rate function. We also define three extensions of the CPF information measure. Further, we study these cumulative information measures and their Bayes versions for some well-known models used in reliability, economics and survival analysis. The associated results reveal some interesting connections between the proposed Fisher type information measures with some well-known information divergences and reliability measures.

References

Asadi, M. (2017). A new measure of association between random variables. *Metrika* **80**, 649–661. MR3716609 https://doi.org/10.1007/s00184-017-0620-5

Asadi, M., Ebrahimi, N., Soofi, E. S. and Zohrevand, Y. (2016). Jensen–Shannon information of the coherent system lifetime. *Reliability Engineering & Systems Safety* **156**, 244–255.

Asadi, M. and Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference* **137**, 1931–1941. MR2323874 https://doi.org/10.1016/j.jspi.2006.06.035

Balakrishnan, N. and Stepanov, A. (2006). On the Fisher information in record data. *Statistics & Probability Letters* **76**, 537–545. MR2266609 https://doi.org/10.1016/j.spl.2005.08.021

Bercher, J. F. (2013). Some properties of generalized Fisher information in the context of nonextensive thermostatistics. *Physica A: Statistical Mechanics and its Applications* **392**, 3140–3154. MR3061519 https://doi.org/10.1016/j.physa.2013.03.062

Bobkov, S. G. (2019). Moments of the scores. *IEEE Transactions on Information Theory* **65**, 5294–5301. MR4009234 https://doi.org/10.1109/TIT.2019.2903248

Cover, T. M. and Thomas, J. A. (1991). Information theory and statistics. *Elements of Information Theory* **1**, 279–335. MR1122806 https://doi.org/10.1002/0471200611

Cvetkovski, Z. (2012). *Inequalities: Theorems, Techniques and Selected Problems*. New York: Springer. MR3015124 https://doi.org/10.1007/978-3-642-23792-8

Di Crescenzo, A. and Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference* **139**, 4072–4087. MR2558351 https://doi.org/10.1016/j.jspi.2009.05.038

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London, Series A* **125**, 54–59.

Gupta, R. C. and Wu, H. (2001). Analyzing survival data by proportional reversed hazard model. *International Journal of Reliability and Applications* **2**, 1–26.

Kayal, S. (2016). On generalized cumulative entropies. *Probability in the Engineering and Informational Sciences* **30**, 640–662. MR3569139 https://doi.org/10.1017/S0269964816000218

Kharazmi, O. and Asadi, M. (2018). On the time-dependent Fisher information of a density function. *Brazilian Journal of Probability and Statistics* **32**, 795–814. MR3845030 https://doi.org/10.1214/17-BJPS366

Kharazmi, O. and Balakrishnan, N. (2021). Cumulative residual and relative cumulative residual Fisher information and their properties. *IEEE Transactions on Information Theory* **67**, 6306–6312. MR4345063 https://doi.org/10.1109/TIT.2021.3073789

Key words and phrases. Shannon entropy, Fisher information, cumulative past Fisher information, relative cumulative past Fisher information, Jensen inequality, Hazard function, reversed hazard rate function, Bayes–Fisher information.

Kirmani, S. N. U. A. and Gupta, R. C. (2001). On the proportional odds model in survival analysis. *Annals of the Institute of Statistical Mathematics* **53**, 203–216. MR1841132 https://doi.org/10.1023/A:1012458303498

Kundu, C. and Ghosh, A. (2017). Inequalities involving expectations of selected functions in reliability theory to characterize distributions. *Communications in Statistics Theory and Methods* **46**, 8468–8478. MR3680770 https://doi.org/10.1080/03610926.2016.1183784

Marshall, A. W. and Olkin, I. (2007). *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*. New York: Springer. MR2344835

Mehrali, Y., Asadi, M. and Kharazmi, O. (2018). A Jensen–Gini measure of divergence with application in parameter estimation. *Metron* **76**, 115–131. MR3774524 https://doi.org/10.1007/s40300-017-0119-x

Navarro, J., Ruiz, J. M. and Del Aguila, Y. (2008). Characterizations and ordering properties based on log-odds functions. *Statistics* **46**, 313–328. MR2449611 https://doi.org/10.1080/02331880701835762

Park, S., Noughabi, H. A. and Kim, I. (2018). General cumulative Kullback–Leibler information. *Communications in Statistics Theory and Methods* **47**, 1551–1560. MR3766667 https://doi.org/10.1080/03610926.2017.1321767

Park, S., Rao, M. and Shin, D. W. (2012). On cumulative residual Kullback–Leibler information. *Statistics & Probability Letters* **82**, 2025–2032. MR2970308 https://doi.org/10.1016/j.spl.2012.06.015

Rao, M., Chen, Y., Vemuri, B. C. and Wang, F. (2004). Cumulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory* **50**, 1220–1228. MR2094878 https://doi.org/10.1109/TIT.2004.828057

Sánchez-Moreno, P., Zarzo, A. and Dehesa, J. S. (2012). Jensen divergence based on Fisher's information. *Journal of Physics A: Mathematical and Theoretical* **45**, 125305. MR2904391 https://doi.org/10.1088/1751-8113/45/12/125305

Schlattmann, P. (2009). *Medical Applications of Finite Mixture Models*. Berlin: Springer.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Veres-Ferrer, E. J. and Pavía, J. M. (2014). On the relationship between the reversed hazard rate and elasticity. *Statistical Papers* **55**, 275–284. MR3188403 https://doi.org/10.1007/s00362-012-0470-1

Zegers, P. (2015). Fisher information properties. *Entropy* **17**, 4918–4939.

Zimmer, W. J., Wang, Y. and Pathak, P. K. (1998). Log-odds rate and monotone log-odds rate distributions. *Journal of Quality Technology* **30**, 376–385.

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 560–578
https://doi.org/10.1214/22-BJPS540

Log-symmetric models with cure fraction with application to leprosy reactions data

Joyce B. Rocha[a], **Francisco M. C. Medeiros**[b] **and Dione M. Valença**[c]

Department of Statistics, Federal University of Rio Grande do Norte, UFRN, Natal, Brazil,
[a]*joycerochaa@live.com,* [b]*sesiommedeiros@gmail.com,* [c]*dioneval@gmail.com*

Abstract. In this paper, we propose a log-symmetric survival model with cure fraction, considering that the distributions of lifetimes for susceptible individuals belong to the log-symmetric class of distributions. This class has continuous, strictly positive, and asymmetric distributions, including the log-normal, log-Student-t, Birnbaum–Saunders, log-logistic I, log-logistic II, contaminated log-normal, log-power-exponential, and log-slash distributions. The log-symmetric class is quite flexible and allows for including bimodal distributions and outliers. It has two parameters interpreted directly as location and scale, where the location is the median, which is a robust measure in the presence of outliers and quite informative in survival analysis. The proposed model includes explanatory variables through the parameter associated with the cure fraction. We evaluate the performance of such model through extensive simulation studies and consider a real data application to evaluate the effect of factors on the immunity to leprosy reactions in patients with Hansen's disease.

References

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47**, 501–515.

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

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* **33**, 261–304. MR2086350 https://doi.org/10.1177/0049124104268644

Cancho, V. G., Louzada, F. and Barriga, G. D. (2012). The geometric Birnbaum–Saunders regression model with cure rate. *Journal of Statistical Planning and Inference* **142**, 993–1000. MR2863887 https://doi.org/10.1016/j.jspi.2011.11.007

Cancho, V. G., Louzada, F. and Ortega, E. M. (2013). The power series cure rate model: An application to a cutaneous melanoma data. *Communications in Statistics Simulation and Computation* **42**, 586–602. MR3020088 https://doi.org/10.1080/03610918.2011.639971

Cancho, V. G., Macera, M. A. C., Suzuki, A. K., Louzada, F. and Zavaleta, K. E. C. (2020). A new long-term survival model with dispersion induced by discrete frailty. *Lifetime Data Analysis* **26**, 221–244. MR4079665 https://doi.org/10.1007/s10985-019-09472-2

Cancho, V. G., Rodrigues, J. and de Castro, M. (2011). A flexible model for survival data with a cure rate: A Bayesian approach. *Journal of Applied Statistics* **38**, 57–70. MR2747683 https://doi.org/10.1080/02664760903254052

Carneiro, H. and Valença, D. M. (2016). Gradient and likelihood ratio tests in cure rate models. *International Journal of Statistics and Probability* **5**, 9–21.

Carrasco, J. M., Ortega, E. M. and Cordeiro, G. M. (2008). A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics & Data Analysis* **53**, 450–462. MR2649099 https://doi.org/10.1016/j.csda.2008.08.023

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* **94**, 909–919. MR1723307 https://doi.org/10.2307/2670006

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276. MR0400509 https://doi.org/10.1093/biomet/62.2.269

Key words and phrases. Cure rate, log-symmetric models, maximum likelihood, survival analysis.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837
Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall. MR0751780
D'Andrea, A., Rocha, R., Tomazella, V. and Louzada, F. (2018). Negative binomial Kumaraswamy-G cure rate regression model. *Journal of Risk and Financial Management* **11**, 6.
Flores, J. D., Borges, P., Cancho, V. G. and Louzada, F. (2013). The complementary exponential power series distribution. *Brazilian Journal of Probability and Statistics* **27**, 565–584. MR3105044 https://doi.org/10.1214/11-bjps182
Fonseca, R. S., Valença, D. M. and Bolfarine, H. (2013). Cure rate survival models with missing covariates: A simulation study. *Journal of Statistical Computation and Simulation* **83**, 97–113. MR3005337 https://doi.org/10.1080/00949655.2011.613396
Gallardo, D. I., de Castro, M. and Gomez, H. W. (2021). An alternative promotion time cure model with overdispersed number of competing causes: An application to melanoma data. *Mathematics* **9**, 1–14.
Gallardo, D. I., Gomez, H. W. and Bolfarine, H. (2017). A new cure rate model based on the Yule–Simon distribution with application to a melanoma data set. *Journal of Applied Statistics* **44**, 1153–1164. MR3638437 https://doi.org/10.1080/02664763.2016.1194385
Gallardo, D. I., Gomez, Y. M. and de Castro, M. (2018). A flexible cure rate model based on the polylogarithm distribution. *Journal of Statistical Computation and Simulation* **88**, 2137–2149. MR3804193 https://doi.org/10.1080/00949655.2018.1451850
Gallardo, D. I., Gomez, Y. M., Gomez, H. W. and de Castro, M. (2020). On the use of the modified power series family of distributions in a cure rate model context. *Statistical Methods in Medical Research* **29**, 1831–1845. MR4114816 https://doi.org/10.1177/0962280219876962
Ghitany, M. E. and Maller, R. A. (1992). Asymptotic results for exponential mixture models with long-term survivors. *Statistics: A Journal of Theoretical and Applied Statistics* **23**, 321–336. MR1238254 https://doi.org/10.1080/02331889208802379
Ghitany, M. E., Maller, R. A. and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of Multivariates Analysis* **49**, 218–241. MR1276436 https://doi.org/10.1006/jmva.1994.1023
Hashimoto, E. M., Ortega, E. M., Cordeiro, G. M. and Cancho, V. G. (2014). The Poisson Birnbaum–Saunders model with long-term survivors. *Statistics* **48**, 1394–1413. MR3269743 https://doi.org/10.1080/02331888.2013.822503
Jorgensen, B. (2012). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Berlin: Springer. MR0648107
Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley. MR1924807 https://doi.org/10.1002/9781118032985
Lai, C. D. (2013). Constructions and applications of lifetime distributions. *Applied Stochastic Models in Business and Industry* **29**, 127–140. MR3047739 https://doi.org/10.1002/asmb.948
Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*. New Jersey: John Wiley & Sons. MR0640866
Leão, J., Bourguignon, M., Gallardo, D. I., Rocha, R. and Tomazella, V. (2020). A new cure rate model with flexible competing causes with applications to melanoma and transplantation data. *Statistics in Medicine* **24**, 3272–3284. MR4157839 https://doi.org/10.1002/sim.8664
Leão, J., Leiva, V., Saulo, H. and Tomazella, V. (2018). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Statistics in Medicine* **29**, 4421–4440. MR3879437 https://doi.org/10.1002/sim.7929
Loose, L. H., Valença, D. M. and Bayer, F. M. (2018). On bootstrap testing inference in cure rate models. *Journal of Statistical Computation and Simulation* **17**, 3437–3454. MR3857104 https://doi.org/10.1080/00949655.2018.1522316
Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics Theory and Methods* **26**, 1165–1182. MR1450228 https://doi.org/10.1080/03610929708831974
Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long–Term Survivors*. New York: Wiley. MR1453117
Medeiros, F. M. C. and Ferrari, S. L. P. (2017). Small-sample testing inference in symmetric and log-symmetric linear regression models. *Statistica Neerlandica* **71**, 200–224. MR3680260 https://doi.org/10.1111/stan.12107
Mizoi, M. F., Bolfarine, H. and Pedroso-De-Lima, A. C. (2007). Cure rate model with measurement errors. *Communications in Statistics Simulation and Computation* **36**, 185–196. MR2370933 https://doi.org/10.1080/03610910601095991
Ortega, E. M., Cancho, V. G. and Paula, G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis* **15**, 79. MR2471626 https://doi.org/10.1007/s10985-008-9096-y
Pal, A., Mondal, S. and Kundu, D. (2021). A cure rate model for exponentially distributed lifetimes with competing risks. *Journal of Statistical Theory and Practice* **15**, 21. MR4198530 https://doi.org/10.1007/s42519-020-00145-0

Peng, Y., Dear, K. B. and Denham, J. W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine* **17**, 813–830.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rodrigues, A. S., Calsavara, V. F., Bertolli, E., Peres, S. V. and Tomazella, V. (2021). Bayesian long-term survival model including a frailty term: Application to melanoma data. *Chilean Journal of Statistics* **12**, 53–70. MR4299220

Rodrigues, J., Cancho, V. G., de Castro, M. and Louzada-Neto, F. (2009). On the unification of long-term survival models. *Statistics & Probability Letters* **79**, 753–759. MR2662300 https://doi.org/10.1016/j.spl.2008.10.029

Scolas, S., Legrand, C., Oulhaj, A. and El Ghouch, A. (2018). Diagnostic checks in mixture cure models with interval-censoring. *Statistical Methods in Medical Research* **27**, 2114–2131. MR3807925 https://doi.org/10.1177/0962280216676502

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics* **33**, 1187–1192. MR0143277 https://doi.org/10.1214/aoms/1177704481

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

Vanegas, L. H. and Paula, G. A. (2016). Log-symmetric distributions: Statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics* **30**, 196–220. MR3481101 https://doi.org/10.1214/14-BJPS272

Vanegas, L. H. and Paula, G. A. (2017). Log-symmetric regression models under the presence of non-informative left-or right-censored observations. *Test* **26**, 405–428. MR3650533 https://doi.org/10.1007/s11749-016-0517-z

Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. *Biometrie et Analyse de Donnees Spatio-Temporelles* **12**, 66–82.

Yiqi, B., Cancho, V. G., Dey, D. K., Balakrishnan, N. and Suzuki, A. K. (2020). Power series cure rate model for spatially correlated interval-censored data based on generalized extreme value distribution. *Journal of Computational and Applied Methematics* **364**, 112362. MR3986501 https://doi.org/10.1016/j.cam.2019.112362

Yu, B., Tiwari, R. C., Cronin, K. A. and Feuer, E. J. (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
2022, Vol. 36, No. 3, 579–592
https://doi.org/10.1214/22-BJPS541

Trajectory fitting estimation for a class of SDEs with small Lévy noises

Xuekang Zhang[1,a] **and Huisheng Shu**[2,b]
[1]*School of Mathematics-Physics and Finance, and Energy Internet Engineering Research Center of Anhui Provincial Department of Education, Anhui Polytechnic University, Wuhu, 241000, China,* [a]*xkzhang@ahpu.edu.cn*
[2]*College of Science, Donghua University, Shanghai, 201620, China,* [b]*hsshu@dhu.edu.cn*

Abstract. In this paper, we consider the problem of trajectory fitting estimation for a class of stochastic differential equations with small Lévy noises based on continuous-time observations. The consistency, the rate of convergence, and asymptotic distribution of the trajectory fitting estimator are established as a small dispersion coefficient tends to zero.

References

Dietz, H. (2001). Asymptotic behavior of trajectory fitting estimators for certain non-ergodic SDE. *Statistical Inference for Stochastic Processes* **4**, 249–258. MR1868721 https://doi.org/10.1023/A:1012254332474

Dietz, H. and Kutoyants, Y. (1997). A class of minimum-distance estimators for diffusion processes with ergodic properties. *Statistics & Decisions* **15**, 211–227. MR1484377

Dietz, H. and Kutoyants, Y. (2003). Parameter estimation for some non-recurrent solutions of SDE. *Statistics & Decisions* **21**, 29–45. MR1985650 https://doi.org/10.1524/stnd.21.1.29.20321

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21**, 99–116. MR1056065 https://doi.org/10.1080/02331889008802231

Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and Their Applications* **119**, 679–699. MR2500255 https://doi.org/10.1016/j.spa.2008.04.004

Hu, Y. and Long, H. (2007). Parameter estimation for Ornstein–Uhlenbeck processes driven by α-stable Lévy motions. *Communications on Stochastic Analysis* **1**, 175–192. MR2397392 https://doi.org/10.31390/cosa.1.2.01

Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*. New York: Marcel Dekker. MR1306279

Kallenberg, O. (1992). Some time change representations of stable integrals, via predictable transformations of local martingales. *Stochastic Processes and Their Applications* **40**, 199–223. MR1158024 https://doi.org/10.1016/0304-4149(92)90012-F

Kunitomo, N. and Takahashi, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims. *Mathematical Finance* **11**, 117–151. MR1807851 https://doi.org/10.1111/1467-9965.00110

Kutoyants, Y. (1991). Minimum distance parameter estimation for diffusion type observations. *Comptes Rendus de l'Académie des Sciences, Series I* **312**, 637–642. MR1101048

Kutoyants, Y. (1994). *Identification of Dynamical Systems with Small Noise*. Dordrecht: Kluwer. MR1332492 https://doi.org/10.1007/978-94-011-1020-4

Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Berlin, Heidelberg: Springer. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

Laredo, C. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *The Annals of Statistics* **18**, 1158–1171. MR1062703 https://doi.org/10.1214/aos/1176347744

Li, J. and Wu, J. (2016). On drift parameter estimation for mean-reversion type stochastic differential equations with discrete observations. *Advances in Difference Equations* **90**, 1–23. MR3480640 https://doi.org/10.1186/s13662-016-0819-1

Long, H. (2009). Least squares estimator for discretely observed Ornstein–Uhlenbeck processes with small Lévy noises. *Statistics & Probability Letters* **79**, 2076–2085. MR2571771 https://doi.org/10.1016/j.spl.2009.06.018

Key words and phrases. Trajectory fitting estimation, stochastic differential equations, small Lévy noises, consistency, asymptotic distribution.

Long, H. (2010). Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations. *Acta Mathematica Scientia* **30B**, 645–663. MR2675775 https://doi.org/10.1016/S0252-9602(10)60067-7

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

Long, H., Shimizu, Y. and Sun, W. (2013). Least squares estimators for discretely observed stochastic processes driven by small Lévy noises. *Journal of Multivariate Analysis* **116**, 422–439. MR3049914 https://doi.org/10.1016/j.jmva.2013.01.012

Ma, C. (2010). A note on "Least squares estimator for discretely observed Ornstein–Uhlenbeck processes with small Lévy noise". *Statistics & Probability Letters* **80**, 1528–1531. MR2669756 https://doi.org/10.1016/j.spl.2010.06.006

Mattingly, J., Stuart, A. and Higham, D. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and Their Applications* **101**, 185–232. MR1931266 https://doi.org/10.1016/S0304-4149(02)00150-3

Ren, P. and Wu, J. (2019). Least squares estimator for path-dependent McKean–Vlasov SDEs via discrete-time observations. *Acta Mathematica Scientia* **39B**, 691–716. MR4066500 https://doi.org/10.1007/s10473-019-0305-4

Rosinski, J. and Woyczynski, W. A. (1986). On Ito stochastic integration with respect to p-stable motion: Inner clock, integrability of sample paths, double and multiple integrals. *The Annals of Probability* **14**, 271–286. MR0815970

Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051–1069. MR2046817 https://doi.org/10.3150/bj/1072215200

Takahashi, A. and Yoshida, N. (2004). An asymptotic expansion scheme for optimal investment problems. *Statistical Inference for Stochastic Processes* **7**, 153–188. MR2061183 https://doi.org/10.1023/B:SISP.0000026045.26381.1d

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scandinavian Journal of Statistics* **31**, 553–566. MR2101539 https://doi.org/10.1111/j.1467-9469.2004.00406.x

Uchida, M. (2008). Approximate martingale estimating functions for stochastic differential equations with small noises. *Stochastic Processes and Their Applications* **118**, 1706–1721. MR2442376 https://doi.org/10.1016/j.spa.2007.10.008

Uchida, M. and Yoshida, N. (2004). Asymptotic expansion for small diffusions applied to option pricing. *Statistical Inference for Stochastic Processes* **7**, 189–223. MR2111290 https://doi.org/10.1023/B:SISP.0000049093.20850.11

Yang, X. (2017). Maximum likelihood type estimation for discretely observed CIR model with small α-stable noises. *Statistics & Probability Letters* **120**, 18–27. MR3567917 https://doi.org/10.1016/j.spl.2016.09.014

Zhang, X., Yi, H. and Shu, H. (2020). Parameter estimation for non-stationary reflected Ornstein–Uhlenbeck processes driven by α-stableb noises. *Statistics & Probability Letters* **156**, 108617. MR4008041 https://doi.org/10.1016/j.spl.2019.108617

Zhang, X., Yi, H. and Shu, H. (2021). Parameter estimation for certain nonstationary processes driven by α-stable motions. *Communications in Statistics Theory and Methods* **50**, 95–104. MR4195010 https://doi.org/10.1080/03610926.2019.1630436

Zhao, H. and Zhang, C. (2019). Minimum distance parameter estimation for SDEs with small α-stable noises. *Statistics & Probability Letters* **145**, 301–311. MR3873919 https://doi.org/10.1016/j.spl.2018.10.009

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 593–612
https://doi.org/10.1214/22-BJPS542

Finite mixtures of multivariate skew Student's *t* distributions with independent logistic skewing functions

Hok Shing Kwong and Saralees Nadarajah[a]

Department of Mathematics, University of Manchester, Manchester M13 9PL, UK,
[a]*saralees.nadarajah@manchester.ac.uk*

Abstract. This paper extends the multivariate skew *t* distributions with independent logistic skewing functions (MSTIL) introduced in Kwong and Nadarajah (*Methodology and Computing in Applied Probability* **24** (2022) 1669–1691) to finite mixture models (FM-MSTIL). A stochastic EM-type algorithm is proposed for fitting the FM-MSTIL, and a divisive hierarchical algorithm is proposed for initialisations and model selections. We show that the model can outperform other finite mixture models in the literature for some simulated data sets. The performance of the FM-MSTIL in cluster analysis is also investigated. We show that the FM-MSTIL-R, a nested version of the FM-MSTIL, performs well for automatic gating tasks on some flow cytometry data sets in the FlowCap-I challenge. The FM-MSTIL-R achieved a better overall score than all other competing algorithms in the original challenge. An efficient implementation of the FM-MSTIL is available as an R package in GitHub.

References

Aghaeepour, N., Finak, G., Hoos, H., Mosmann, T. R., Brinkman, R., Gottardo, R. and Scheuermann, R. H. (2013). Critical assessment of automated flow cytometry data analysis techniques. *Nature Methods* **10**, 228–238.

Aghaeepour, N., Nikolic, R., Hoos, H. H. and Brinkman, R. R. (2011). Rapid cell population identification in flow cytometry data. *Cytometry Part A* **79**, 6–13.

Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis* **96**, 93–116. MR2202402 https://doi.org/10.1016/j.jmva.2004.10.002

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. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **65**, 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391

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

Biernacki, C., Celeux, G. and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 719–725.

Bilmes, J., Asanovic, K., Chin, C.-W. and Demmel, J. (1997). Using PHiPAC to speed error back-propagation learning. In *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5*, 4153–4156.

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

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

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

Key words and phrases. Cluster analyses, EM algorithm, flow cytometry, multivariate distributions, robust distributions.

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics* **11**, 317–336.

Hartigan, J. A. (1975). *Clustering Algorithms*. New York: John Wiley and Sons. MR0405726

Ho, H. J., Pyne, S. and Lin, T. I. (2012). Maximum likelihood inference for mixtures of skew Student-t-normal distributions through practical EM-type algorithms. *Statistics and Computing* **22**, 287–299. MR2865071 https://doi.org/10.1007/s11222-010-9225-9

Kwong, H. S. and Nadarajah, S. (2022). A new robust class of skew elliptical distributions. *Methodology and Computing in Applied Probability* **24**, 1669–1691. MR4457560 https://doi.org/10.1007/s11009-021-09883-5

Lee, S. X. and McLachlan, G. J. (2011). On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. ArXiv preprint. Available at arXiv:1109.4706. MR3165547 https://doi.org/10.1007/s11222-012-9362-4

Lee, S. X. and McLachlan, G. J. (2014). Finite mixtures of multivariate skew t-distributions: Some recent and new results. *Statistics and Computing* **24**, 181–202. MR3165547 https://doi.org/10.1007/s11222-012-9362-4

Lee, S. X. and McLachlan, G. J. (2016). Finite mixtures of canonical fundamental skew t-distributions: The unification of the restricted and unrestricted skew t-mixture models. *Statistics and Computing* **26**, 573–589. MR3489858 https://doi.org/10.1007/s11222-015-9545-x

Lee, S. X. and McLachlan, G. J. (2018). EMMIXcskew: An R package for the fitting of a mixture of canonical fundamental skew t-distributions. *Journal of Statistical Software* **83**, 1–32. MR3489858 https://doi.org/10.1007/s11222-015-9545-x

Lin, T. I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* **20**, 343–356. MR2725392 https://doi.org/10.1007/s11222-009-9128-9

Lin, T. I., Lee, J. C. and Hsieh, W. J. (2007). Robust mixture modeling using the skew t distribution. *Statistics and Computing* **17**, 81–92. MR2380638 https://doi.org/10.1007/s11222-006-9005-8

Lin, T. I., Lee, J. C. and Yen, S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Statistica Sinica* **17**, 909–927. MR2408641

McLachlan, G. J., Lee, S. X. and Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application* **6**, 355–378. MR3939525 https://doi.org/10.1146/annurev-statistics-031017-100325

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics* **8**, 343–366. MR1505430 https://doi.org/10.2307/2369392

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society Series A* **185**, 71–110.

Povey, D., Burget, L., Agarwal, M., Akyazi, P., Feng, K., Ghoshal, A., Glembek, O., Goel, N. K., Karafiát, M., Rastrow, A., Rose, R. C., Schwarz, P. and Thomas, S. (2010). Subspace Gaussian mixture models for speech recognition. In *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4330–4333.

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.-I., Maiera, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A., De Jagera, P. L. and Mesirov, J. P. (2009). Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences* **106**, 8519–8524.

Qian, Y., Wei, C., Lee, F. E.-H., Campbell, J., Halliley, J., Lee, J. A., Cai, J., Kong, Y. M., Sadat, E., Thomson, E., Dunn, P., Seegmiller, A. C., Karandikar, N. J., Tipton, C. M., Mosmann, T., Sanz, I. and Scheuermann, R. H. (2010). Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. *Cytometry Part B: Clinical Cytometry* **78**, S69–S82.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reynolds, D. A. and Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing* **3**, 72–83.

Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* **31**, 129–150. MR2016224 https://doi.org/10.2307/3316064

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

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 3, 613–639
https://doi.org/10.1214/22-BJPS545

On the Gompertz–Makeham law: A useful mortality model to deal with human mortality

Fredy Castellares[1,a], **Silvio Patrício**[2,b] **and Artur J. Lemonte**[3,c]

[1]*Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte/MG, Brazil,* [a]*fwcc29@gmail.com*

[2]*Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark,* [b]*silviocabralpatricio@msn.com*

[3]*Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal/RN, Brazil,* [c]*arturlemonte@gmail.com*

Abstract. The Gompertz–Makeham model was introduced as an extension of the Gompertz model in the second half of the 19th century by the British actuary William M. Makeham. Since then, this model has been successfully used in biology, actuarial science, and demography to describe mortality patterns in numerous species (including humans), determine policies in insurance, establish actuarial tables and growth models. In this paper, we derive some structural properties of the Gompertz–Makeham model in statistics, demography, and actuarial sciences, and present some other ones already introduced in the literature. All structural properties we provide are expressed in closed-form, which eliminates the need to evaluate them with numerical integration directly. In addition, we study the estimation of the Gompertz–Makeham model parameters through the discrete Poisson and Bell distributions. In particular, we verify that the recently introduced discrete Bell distribution can be an interesting alternative to the Poisson distribution, mainly because it is suitable to deal with over dispersion, unlike the Poisson distribution. On the basis of real mortality datasets, we compute the remaining life expectancy for several countries and verify that the Gompertz–Makeham model, especially under the Bell distribution, provides proper results to deal with human mortality in practice.

References

Bell, E. T. (1934a). Exponential polynomials. *Annals of Mathematics* **35**, 258–277. MR1503161 https://doi.org/10.2307/1968431

Bell, E. T. (1934b). Exponential numbers. *The American Mathematical Monthly* **41**, 411–419. MR1523147 https://doi.org/10.2307/2300300

Bierens, H. (1994). *Topics in Advanced Econometrics*. Cambridge, UK: Cambridge University Press. MR1291390 https://doi.org/10.1017/CBO9780511599279

Bowden, R. (1973). The theory of parametric identification. *Econometrica* **17**, 53–61. MR0440832 https://doi.org/10.2307/1914036

Bowers, N., Gerber, H., Hickman, J., Jones, D. and Nesbitt, C. (1997). *Actuarial Mathematics*, 2nd ed. The Society of Actuaries. MR2011229 https://doi.org/10.1080/10920277.1999.10595793

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* **42**, 693–734. MR0872958 https://doi.org/10.2307/2530689

Castellares, F., Ferrari, S. L. P. and Lemonte, A. J. (2018). On the Bell distribution and its associated regression model for count data. *Applied Mathematical Modelling* **56**, 172–185. MR3744937 https://doi.org/10.1016/j.apm.2017.12.014

Chan, B. (1982). Life annuities under Makeham's law. *Scandinavian Actuarial Journal* **1982**, 238–240.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. and Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics* **5**, 329–359. MR1414285 https://doi.org/10.1007/BF02124750

Key words and phrases. Gompertz mortality model, growth models, life expectancy, Makeham mortality model.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London, UK: Chapman and Hall. MR0370837

Dean, C. B. and Lundy, E. R. (2016). *Overdispersion*. Wiley StatsRef: Statistics Reference Online. USA: John Wiley & Sons.

Dickson, D., Hardy, M. and Waters, H. (2013). *Actuarial Mathematics for Life Contingent Risk*, 2nd ed. Cambridge, UK: Cambridge University Press. MR3112779

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis*. New York, USA: Wiley. MR0586940

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* **115**, 513–585.

Gradsteyn, I. and Ryzhik, I. (2012). *Table of Integrals, Series, and Products*. London, UK: Academic Press. MR0669666

Gurven, M. and Kaplan, H. (2007). Longevity among hunter-gatherers: A cross-cultural examination. *Population and Development Review* **33**, 321–365.

Haberman, S. (1996). Landmarks in the history of actuarial science (up to 1919). Actuarial research paper no. 84, Dept. of Actuarial Science and Statistics, City University, London.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge, UK: Cambridge University Press. MR2797563 https://doi.org/10.1017/CBO9780511973420

HMD (2012). Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at http://www.mortality.org.

Jodrá, P. (2009). A closed expression for the quantile function of the Gompertz–Makeham model. *Mathematics and Computers in Simulation* **79**, 3069–3075. MR2541318 https://doi.org/10.1016/j.matcom.2009.02.002

Jodrá, P. (2013). On order statistics from the Gompertz–Makeham distribution and the Lambert W function. *Mathematical Modelling and Analysis* **18**, 432–445. MR3175655 https://doi.org/10.3846/13926292.2013.807316

Kozusko, F. and Bajzer, Z. (2003). Combining Gompertzian growth and cell population dynamics. *Mathematical Biosciences* **185**, 153–167. MR2005948 https://doi.org/10.1016/S0025-5564(03)00094-4

Lageras, A. N. (2010). Commutation functions under Gompertz–Makeham mortality. *Scandinavian Actuarial Journal* **2010**, 161–164. MR2756764 https://doi.org/10.1080/03461230903344181

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR1639875

Lenart, A. (2014). The moments of Gompertz distribution and maximum likelihood estimation of its parameter. *Scandinavian Actuarial Journal* **2014**, 255–277. MR3196275 https://doi.org/10.1080/03461238.2012.687697

Macdonald, A. S., Currie, I. D. and Richards, S. J. (2018). *Modelling Mortality with Actuarial Applications*. Cambridge, UK: Cambridge University Press. MR3753689 https://doi.org/10.1017/9781107051386

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries* **8**, 301–310.

Makeham, W. M. (1867). On the law of mortality. *Journal of the Institute of Actuaries* **13**, 325–358.

Makeham, W. M. (1890). On the further development of Gompertz's law. *Journal of the Institute of Actuaries* **28**, 316–332.

Marshall, A. W. and Olkin, I. (2007). *Life Distributions. Structure of Nonparametric, Semiparametric, and Parametric Families*. *Springer Series in Statistics*. New York, USA: Springer. MR2344835

Milgram, M. (1985). The generalized integro-exponential function. *Mathematics of Computation* **44**, 443–458. MR0777276 https://doi.org/10.2307/2007964

Missov, T. (2013). Gamma-Gompertz life expectancy at birth. *Demographic Research* **28**, 259–270.

Missov, T. I. and Lenart, A. (2013). Gompertz–Makeham life expectancies: Expressions and applications. *Theoretical Population Biology* **90**, 29–35.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313. MR3363409 https://doi.org/10.1093/comjnl/7.4.308

Norstrom, F. (1997). The Gompertz–Makeham distribution. Unpublished master's thesis in Mathematical Statistics, UmeåUniversity.

Olshansky, S. J. and Carnes, B. A. (1997). Ever since Gompertz. *Demography* **34**, 1–15.

Olver, F., Lozier, D., Boisvert, R. and Clark, C. (2010). *Handbook of Mathematical Functions*. Cambridge, UK: Cambridge University Press. MR2723248

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Teimouri, M. and Gupta, A. K. (2012). Estimation methods for the Gompertz–Makeham distribution under progressively type-I interval censoring scheme. *National Academy Science Letters* **35**, 227–235.

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, PO Box 729, Middletown, Maryland 21769, USA. Mail concerning submissions or editorial content should be sent to the Editor at nancy@ime.unicamp.br. Mail concerning the production of this journal should be sent to: Geri Mattson at bjps@mattsonpublishing.com.

Individual and Organizational Memberships:
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/