Contents

Contributions to the Special Volume of the XVI School on Regression Models ("Escola de Modelos de Regressão") in honour of Bent Jorgensen

Articles

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

Brazilian Journal of Probability and Statistics

Volume 35 • Number 1 • February 2021

Brazilian Journal of Probability and Statistics

Volume 35 • Number 1 • February 2021

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 35, Number 1, February 2021. Published quarterly by the Brazilian Statistical Association.

POSTMASTER:

Send address changes to Brazilian Journal of Probability and Statistics, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

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

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 1
https://doi.org/10.1214/20-BJPS351PRE

Preface

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 2–20
https://doi.org/10.1214/19-BJPS458

An introduction to Bent Jørgensen's ideas

Gauss M. Cordeiro[a], Rodrigo Labouriau[b] and Denise A. Botter[c]

[a]*Universidade Federal de Pernambuco*
[b]*Aarhus University*
[c]*Universidade de São Paulo*

Abstract. We briefly expose some key aspects of the theory and use of dispersion models, for which Bent Jørgensen played a crucial role as a driving force and an inspiration source. Starting with the general notion of dispersion models, built using minimalistic mathematical assumptions, we specialize in two classes of families of distributions with different statistical flavors: exponential dispersion and proper dispersion models. The construction of dispersion models involves the solution of integral equations that are, in general, untractable. These difficulties disappear when more mathematical structure is assumed: it reduces to the calculation of a moment generating function or of a Riemann–Stieltjes integral for the exponential dispersion and the proper dispersion models, respectively. A new technique for constructing dispersion models based on characteristic functions is introduced turning the integral equations above into a tractable convolution equation and yielding examples of dispersion models that are neither proper dispersion nor exponential dispersion models. A corollary is that the cardinality of regular and non-regular dispersion models are both large.

Some selected applications are discussed including exponential families non-linear models (for which generalized linear models are particular cases) and several models for clustered and dependent data based on a latent Lévy process.

References

Artes, R. and Jørgensen, B. (2000). Longitudinal data estimating equations for dispersion models. *Scandinavian Journal of Statistics* **27**, 321–334. MR1777507 https://doi.org/10.1111/1467-9469.00192

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

Barndorff-Nielsen, O. E. (1988). *Parametric Statistical Models and Likelihhod*. *Lecture Notes in Statistics* **50**. Berlin: Springer. MR0971982 https://doi.org/10.1007/978-1-4612-3934-5

Bonat, W. H. and Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society Series C Applied Statistics* **65**, 649–675. MR3564998 https://doi.org/10.1111/rssc.12145

Bonat, W. H., Jørgensen, B., Kokonendji, C. C., Hinde, J. and Demétrio, C. G. B. (2018). Extended Poisson-Tweedie: Properties and regression models for count data. *Statistical Modelling* **18**, 24–49. MR3746503 https://doi.org/10.1177/1471082X17715718

Bonat, W. H. and Kokonendjib, C. C. (2017). Flexible Tweedie regression models for continuous data. *Journal of Statistical Computation and Simulation* **87**, 2138–2152. MR3656096 https://doi.org/10.1080/00949655.2017.1318876

Botter, D. A., Jørgensen, B. and Peres, A. A. Q. (2002). A longitudinal study of mortality and air pollution for São Paulo, Brazil. *Journal of Exposure Analysis and Environmental Epidemiology* **12**, 335–343.

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

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B, Methodological* **53**, 629–643. MR1125720

Cordeiro, G. M. and Paula, G. A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* **76**, 93–100. MR0991426 https://doi.org/10.1093/biomet/76.1.93

Key words and phrases. Dispersion models, exponential dispersion models, exponential family, non-linear models, proper dispersion models, saddlepoint approximations.

Cordeiro, G. M., Paula, G. A. and Botter, D. A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257–274.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25**, 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

Daniels, H. E. (1980). Exact sadlepoint approximation. *Biometrika* **67**, 59–63. MR0570505 https://doi.org/10.1093/biomet/67.1.59

Ferrari, S. L. P., Cordeiro, G. M. and Cribari-Neto, F. (2001). Higher-order asymptotic refinements for score tests in proper dispersion models. *Journal of Statistical Planning and Inference Applied Mathematical Modelling* **97**, 177–190. MR1851381 https://doi.org/10.1016/S0378-3758(00)00352-9

Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. London: Chapman & Hall/CRC. MR2000388

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

Jørgensen, B. (1987a). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* **49**, 127–162. MR0905186

Jørgensen, B. (1987b). Small-dispersion asymptotics. *Brazilian Journal of Probability and Statistics* **1**, 59–90. MR0932188

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall Ltd. MR1462891

Jørgensen, B. (2013). Construction of multivariate dispersion models. *Brazilian Journal of Probability and Statistics* **27**, 285–309. MR3064725 https://doi.org/10.1214/11-BJPS171

Jørgensen, B., Demétrio, C. G. B., Kristensen, E., Banta, G. T., Petersen, H. C. and Delefosse, M. (2011). Bias-corrected Pearson estimating functions for Taylor's power law applied to benthic macrofauna data. *Statistics & Probability Letters* **81**, 749–758. MR2793740 https://doi.org/10.1016/j.spl.2011.01.005

Jørgensen, B. and Labouriau, R. (2012). Exponential families and theoretical inference. In *Monografias de Matematica* **52**, 196. Rio de Janeiro, Brazil: Instituto de Matematica Pura e Aplicada (IMPA). https://pure.au.dk/portal/files/51499534/Mon52.pdf. MR1183933

Jørgensen, B., Labouriau, R. and Lundbye-Christensen, S. (1996a). Linear growth curve analysis based on exponential dispersion models. *Journal of the Royal Statistics Society B* **58**, 573–592. MR1394368

Jørgensen, B. and Lauritzen, S. (2000). Multivariate dispersion models. *Journal of Multivariate Analysis* **74**, 267–281. MR1794004 https://doi.org/10.1006/jmva.1999.1885

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. (1996b). Linear state-space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics* **24**, 385–402. MR1421864 https://doi.org/10.2307/3315747

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. I. (1999). A state space model for multivariate longitudinal count data. *Biometrika* **86**, 169–181. MR1688080 https://doi.org/10.1093/biomet/86.1.169

Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1996c). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in Medicine* **15**, 823–836.

Jørgensen, B. and Song, P. X.-K. (1997). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability* **35**, 78–92. MR1622447 https://doi.org/10.1239/jap/1032192553

Jørgensen, B. and Song, P. X.-K. (2007). Stationary state space models for longitudinal data. *Canadian Journal of Statistics* **35**, 461–483. MR2381394 https://doi.org/10.1002/cjs.5550350401

Jørgensen, B. and Tsao, M. (1999). Linear dispersion models and longitudinal data analysis. *Statistics in Medicine* **18**, 2257–2270.

Labouriau, R. (2020). Construction and extension of dispersion models. Available at arXiv:2008.05448.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal analysis using generalized linear models. *Biometrika* **73**, 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

Lucaks, E. (1970). *Characteristic Functions*, 2nd rev. ed. MR0346874

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**, 475–490. MR0569438 https://doi.org/10.2307/1426607

Ma, R. and Jørgensen, B. (2007). Nested generalized linear mixed models: An orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **69**, 625–641. MR2370072 https://doi.org/10.1111/j.1467-9868.2007.00603.x

Ma, R., Jørgensen, B. and Willms, J. D. (2009). Nested generalized linear mixed models: An orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **69**, 625–641. MR2370072 https://doi.org/10.1111/j.1467-9868.2007.00603.x

Medeiros, F. M. C., Ferrari, S. L. P. and Lemonte, A. J. (2017). Improved inference in dispersion models. *Applied Mathematical Modelling* **51**, 317–328. MR3694537 https://doi.org/10.1016/j.apm.2017.06.045

Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics* **10**, 65–80. MR0642719

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rudin, W. (1973). *Functional Analysis. TMH Edition*. Limited, New Delhi: Tata McGraw-Hill Publishing Company. MR0365062

Simas, A. and Cordeiro, G. M. (2009). Adjusted Pearson residuals in exponential family nonlinear models. *Journal of Statistical Computation and Simulation* **79**, 411–425. MR2522359 https://doi.org/10.1080/00949650701829380

Song, P. X.-K. and Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics* **56**, 496–502.

Tweedie, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proceedings of the Cambridge Philosophical Society* **49**, 41–49. MR0018381

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics Applications and New Directions* (J. K. Ghosh and J. Roy, eds.), *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, 579–604. Indian Statistical Institute. MR0786162

Ushakov, N. G. (1999). *Selected Topics in Characteristic Functions*. Utrecht, The Netherlands: VSP. MR1745554 https://doi.org/10.1515/9783110935981

Wei, B.-C. (1998). *Exponential Family Nonlinear Models*. Berlin: Springer. MR1646023

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

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 21–36
https://doi.org/10.1214/20-BJPS495

Practical issues with modeling extreme Brazilian rainfall

Paulo V. C. Pereira[a], **Isolde T. S. Previdelli**[a] **and Anthony C. Davison**[b]

[a]*Universidade Estadual de Maringá*
[b]*École Polytechnique Fédérale de Lausanne*

Abstract. Accurately quantifying extreme rainfall is important for the design of hydraulic structures, for flood mapping and zoning and for disaster management. In order to produce maps of estimates of 25-year rainfall return levels in Brazil, we selected 893 shorter and 104 longer rainfall time series from the Agência Nacional de Águas (ANA), and applied the framework of extreme value theory. Care was needed to reduce the impact of poor data. Estimates of the shape parameter of the extreme-value model fitted to rainfall data are typically biased, so we discuss an empirical correction that takes into account not only the sample-size bias, but also a so-called penultimate approximation that we use to inform a Bayesian spatial latent variable model for the annual rainfall maxima. This model accounts for subtle patterns of spatial variation in the data and provides plausible return level estimates.

References

Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. New York: Wiley. MR2108013 https://doi.org/10.1002/0470012382

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

Coles, S. G. and Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *Applied Statistics* **52**, 405–416. MR2012566 https://doi.org/10.1111/1467-9876.00413

Davis, R. A., Mikosch, T. and Cribben, I. (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* **170**, 142–152. MR2955945 https://doi.org/10.1016/j.jeconom.2012.04.003

Davison, A. C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application* **2**, 203–235.

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012). Statistical modeling of spatial extremes (with discussion). *Statistical Science* **27**, 161–186. MR2963980 https://doi.org/10.1214/11-STS376

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B* **52**, 393–442. MR1086795

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3

Dourado, F., Arraes, T. C. and Silva, M. F. (2012). The "mega-disaster" in the mountainous region of the state of Rio de Janeiro: Causes, mechanisms of mass movements and spatial allocation of investments for reconstruction post-disaster. *Anuário Do Instituto de Geociências* **35**, 43–54.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society Series C* **58**, 25–45. MR2662232 https://doi.org/10.1111/j.1467-9876.2008.00638.x

Eastoe, E. F. and Tawn, J. A. (2010). Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research* **46**, W02510.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B* **65**, 545–556.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

Key words and phrases. Bayesian hierarchical model, Data quality, penultimate approximation, return level map, small-sample bias, statistics of extremes.

Getirana, A. C. V. (2016). Extreme water deficit in Brazil detected from space. *Journal of Hydrometeorology* **17**, 591–599.

Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261. MR0797563 https://doi.org/10.2307/1269706

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., et al (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. *International Journal of Climatology* **37**, 4131–4145.

Jarraud, M. (2008). *Guide to Meteorological Instruments and Methods of Observation*. World Meteorological Organisation.

Koutsoyiannis, D. (2004a). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences Journal* **49**, 575–590.

Koutsoyiannis, D. (2004b). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal* **49**, 591–610.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer. MR0691492

Marengo, J. A., Alves, L. M., Soares, W. R., Rodriguez, D. A., Camargo, H., Riveros, M. P. and Pabló, A. D. (2013). Two contrasting severe seasonal extremes in tropical South America in 2012: Flood in Amazonia and drought in northeast Brazil. *Journal of Climate* **26**, 9137–9154.

Papalexiou, S. M. and Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research* **49**, 187–201.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131. MR0423667

Plummer, N., Allsopp, T., Lopez, J. A. and Llansó, P. (2003). *Guidelines on Climate Observation: Networks and Systems*. World Meteorological Organization.

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

Rao, V. B., Franchito, S. H., Santo, C. M. and Gan, M. A. (2016). An update on the rainfall characteristics of Brazil: Seasonal variations and trends in 1979–2011. *International Journal of Climatology* **36**, 291–302.

Resnick, S. I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer. MR2271424

Serinaldi, F. and Kilsby, C. G. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research* **50**, 336–352.

Smith, R. L. (1987). Approximations in extreme value theory. Technical report, Department of Statistics and Operations Research, University of North Carolina.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **64**, 583–639. MR2490535 https://doi.org/10.1016/S0169-7161(05)25015-0

Venema, V., Trewin, B., Wang, X., Szentimrey, T., Lakatos, M., Aguilar, E., Auer, I., Guijarro, J. A., Menne, M., Oria, C., Louamba, W. S. R. L. and Rasul, G. (2018). Guidance on the homogenization of climate station data. Available at https://eartharxiv.org/8qzrf/.

Wang, X. L., Chen, H., Wu, Y., Feng, Y. and Pu, Q. (2010). New techniques for the detection and adjustment of shifts in daily precipitation data series. *Journal of Applied Meteorology and Climatology* **49**, 2416–2436.

Wilson, P. S. and Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters* **32**, 1–4.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 37–49
https://doi.org/10.1214/20-BJPS489

A general expression for second-order covariance matrices—an application to dispersion models

Tiago M. Magalhães[a], **Denise A. Botter**[b] **and Mônica C. Sandoval**[b]

[a]*Federal University of Juiz de Fora*
[b]*University of São Paulo*

Abstract. We present a general expression that allows the calculation of both the n^{-2} asymptotic covariance matrices of the maximum likelihood estimator (MLE) and the first-order bias corrected MLE, where n is the sample size. The formula is presented in a matrix notation which has numerical advantages since it requires only simple operations on matrices and vectors. The usefulness of the formula is to construct better Wald statistics. We apply our findings to dispersion models and develop simulation studies which show that modification in the Wald statistic effectively removes size distortions of the type I error probability with no power loss. For illustrative purposes, a real data application is considered to support our theoretical results.

References

Barroso, L. P., Botter, D. A. and Cordeiro, G. M. (2013). Second-order covariance matrix formula for heteroskedastic generalized linear models. *Communications in Statistics Theory and Methods* **42**, 1618–1627. MR3041489 https://doi.org/10.1080/03610926.2011.594543

Cordeiro, G. M. (1993). General matrix formulae for computing bartlett corrections. *Statistics & Probability Letters* **16**, 11–18. MR1208493 https://doi.org/10.1016/0167-7152(93)90115-Y

Cordeiro, G. M., Barroso, L. P. and Botter, D. A. (2006). Covariance matrix formula for generalized linear models with unknown dispersion. *Communications in Statistics Theory and Methods* **35**, 113–120. MR2274004 https://doi.org/10.1080/03610920500439687

Cordeiro, G. M., Botter, D. A., Cavalcanti, A. B. and Barroso, L. P. (2014). Covariance matrix of the bias-corrected maximum likelihood estimator in generalized linear models. *Statistical Papers* **55**, 643–652. MR3227544 https://doi.org/10.1007/s00362-013-0514-1

Cordeiro, G. M. and Ferrari, S. L. P. (1996). Bartlett-type corrections for some score tests in proper dispersion models. *Communications in Statistics Theory and Methods* **25**, 29–48. MR1378945 https://doi.org/10.1080/03610929608831678

Cordeiro, G. M. and Klein, R. (1994). Bias correction in arma models. *Statistics & Probability Letters* **30**, 1317–1334. MR1278646 https://doi.org/10.1016/0167-7152(94)90100-7

Cordeiro, G. M., Paula, G. A. and Botter, D. A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257–274.

Cordeiro, G. M. and Santana, R. G. (2008). Covariance matrix formula for exponential nonlinear models. *Communications in Statistics Theory and Methods* **37**, 2724–2734. MR2458951 https://doi.org/10.1080/03610920802040407

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

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

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B* **49**, 127–162. MR0905186

Jørgensen, B. (1997a). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics* **11**, 89–140. MR1619709

Jørgensen, B. (1997b). *The Theory of Dispersion Models*. London: Chapman & Hall. MR1462891

Kosmidis, I., Kenne Pagui, E. C. and Sartori, N. (2020). Mean and median bias reduction in generalized linear models. *Statistics and Computing* **30**, 43–59. MR4057470 https://doi.org/10.1007/s11222-019-09860-6

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

Key words and phrases. Bias estimator, covariance matrix, dispersion models, Wald test.

Lemonte, A. J. (2011). Covariance matrix formula for Birnbaum-Saunders regression models. *Journal of Statistical Computation and Simulation* **81**, 899–908. MR2806933 https://doi.org/10.1080/00949650903555288

Lemonte, A. J. (2020). Covariance matrix of maximum likelihood estimators in censored exponential regression models. *Communications in Statistics—Theory and Methods*. https://doi.org/10.1080/03610926.2020.1767142

Lemonte, A. J. and Ferrari, S. L. P. (2012). Local power and size properties of the lr, Wald, score and gradient tests in dispersion models. *Statistical Methodology* **9**, 537–554. MR2925985 https://doi.org/10.1016/j.stamet.2012.03.001

Magalhães, T. M., Botter, D. A. and Sandoval, M. C. (2017). Corrigendum to: "covariance matrix formula for generalized linear models with unknown dispersion" by G. M. Cordeiro, L. P. Barroso and D. A. Botter [Communications in Statistics—Theory and Methods (2006) 35(1), 113–120]. *Communications in Statistics Theory and Methods* **46**, 10702–10703. MR3740831 https://doi.org/10.1080/03610926.2016.1242745

Magalhães, T. M., Botter, D. A. and Sandoval, M. C. (2020). Supplement to "A general expression for second-order covariance matrices—an application to dispersion models." https://doi.org/10.1214/20-BJPS489SUPP

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

Medeiros, F. M. C., Ferrari, S. L. P. and Lemonte, A. J. (2017). Improved inference in dispersion models. *Applied Mathematical Modelling* **51**, 317–328. MR3694537 https://doi.org/10.1016/j.apm.2017.06.045

Patriota, A. G. and Cordeiro, G. M. (2011). A matrix formula for the skewness of maximum likelihood estimators. *Statistics & Probability Letters* **81**, 529–537. MR2765174 https://doi.org/10.1016/j.spl.2010.12.009

Peers, H. W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *Journal of the Royal Statistical Society, Series B, Methodological* **47**, 547–554. MR0844486

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50–57.

Rasch, D., Verdooren, R. and Pilz, J. (2020). *Applied Statistics—Theory and Problem Solutions with R*. New Jersey: Wiley.

Rocha, A. V., Simas, A. B. and Cordeiro, G. M. (2010). Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models. *Statistics & Probability Letters* **80**, 718–725. MR2595152 https://doi.org/10.1016/j.spl.2009.12.030

Shenton, L. R. and Bowman, K. O. (1977). *Maximum Likelihood Estimation in Small Samples*. London: Griffin.

Simas, A. B., Cordeiro, G. M. and Rocha, A. V. (2010). Skewness of maximum likelihood estimators in dispersion models. *Journal of Statistical Planning and Inference* **140**, 2111–2121. MR2606745 https://doi.org/10.1016/j.jspi.2010.02.007

Simas, A. B., Rocha, A. V. and Barreto-Souza, W. (2011). Bias-corrected estimators for dispersion models with dispersion covariates. *Journal of Statistical Planning and Inference* **141**, 3063–3074. MR2796012 https://doi.org/10.1016/j.jspi.2011.03.028

Wald, A. (1943). Test of statistical hypotheses concerning several parameter when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482. MR0012401 https://doi.org/10.2307/1990256

Wilks, S. S. (1938). The large-sample distribution of likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**, 60–62.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 50–61
https://doi.org/10.1214/20-BJPS487

Self-organized criticality of aggregated animals attributed to Tweedie convergence

Wayne S. Kendal
University of Ottawa

In memoriam: Prof. Bent Jørgensen (1954–2015), who contributed the fundamentals on which this study was based.

Abstract. Ecologists have had an ongoing interest in a variance to mean power law that governs the clustering of individuals of animal and plant species. This same power law has been reported from disparate biological, physical and mathematical systems, and also characterizes a family of statistical distributions known as the Tweedie exponential dispersion models. Its widespread appearance can be explained by fundamental statistical convergence effects on random data that cause this, and related, power laws to emerge and provide mechanistic insight into its origin, as well as the origin of $1/f$ noise, multifractality and other phenomena attributable to self-organized criticality. A meta-analysis of ecological field data was conducted here to examine how such statistical convergence might affect the power law. These findings provided conjectural insight into a form of self-organized criticality, driven and modulated by the statistical convergence of random data, which could underlie the power law's emergence.

References

Abid, R., Kokonendji, C. C. and Masmoudi, A. (2019). Geometric dispersion models with real quadratic v-functions. *Statistics & Probability Letters* **145**, 197–204. MR3873907 https://doi.org/10.1016/j.spl.2018.09.010

Abid, R., Kokonendji, C. C. and Masmoudi, A. (2020). Geometric Tweedie regression models for continuous and semicontinuous data with variation phenomenon [in reliability]. *AStA Advances in Statistical Analysis* **104**, 33–58. MR4067219 https://doi.org/10.1007/s10182-019-00350-8

Anderson, R., Gordon, D., Crawley, M. and Hassell, M. (1982). Variability in the abundance of animal and plant species. *Nature* **296**, 245–248.

Anderson, R. M. and May, R. M. (1988). Epidemiological parameters of HIV transmission. *Nature* **333**, 514–519.

Bak, P., Tang, C. and Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A* **38**, 364–374. MR0949160 https://doi.org/10.1103/PhysRevA.38.364

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

Bardner, R. and Lofty, J. (1971). The distribution of eggs, larvae and plants within crops attacked by wheat bulb fly *Leptohylemyia coarctata* (Fall.). *Journal of Applied Ecology* **8**, 683–686.

Blackman, G. E. (1942). Statistical and ecological studies in the distribution of species in plant communities 1. Dispersion as a factor in the study of changes in plant populations. *Annals of Botany London* **6**, 351–370.

Bonat, W. H., Jørgensen, B., Kokonendji, C. C., Hinde, J. and Demetrio, C. G. B. (2018). Extended Poisson–Tweedie: Properties and regression models for count data. *Statistical Modelling* **18**, 24–49. MR3746503 https://doi.org/10.1177/1471082X17715718

Bonat, W. H. and Kokonendji, C. C. (2017). Flexible Tweedie regression models for continuous data. *Journal of Statistical Computation and Simulation* **87**, 2138–2152. MR3656096 https://doi.org/10.1080/00949655.2017.1318876

Boswell, M. T. and Patil, G. P. (1970). Chance mechanisms generating negative binomial distributions. In *Random Counts in Scientific Work* (G. P. Patil, ed.), *Random Counts in Scientific Work: Expanded from the Proceedings*

Key words and phrases. Emergent phenomena, critical phenomena, phase transitions, Taylor's power law, weak convergence.

of the Biometric Society Symposium, *Dallas, Texas*, *December 1968*, **1**, 3–22. University Park, PA: Pennsylvania State University Press. MR0287599

Burrage, R. and Gyrisco, G. (1956). The transformation of counts of European chafer larvae for analysis of variance. *Journal of Economic Entomology* **49**, 179–182.

Eisler, Z., Bartos, I. and Kertesz, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics* **57**, 89–142.

Fronczak, A. and Fronczak, P. (2010). Origins of Taylor's power law for fluctuation scaling in complex systems. *Physical Review E* **81**, 066112.

Hanski, I. (1980). Spatial patterns and movements in coprophagous beetles. *Oikos* **34**, 293–310.

Holme, N. (1950). Population-dispersion in *Tellina tenuis* da Costa. *Journal of The Marine Biological Association Plymouth* **29**, 267–280.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall. MR1462891

Jørgensen, B. and Kokonendji, C. C. (2011). Dispersion models for geometric sums. *Brazilian Journal of Probability and Statistics* **25**, 263–293. MR2832887 https://doi.org/10.1214/10-BJPS136

Jørgensen, B. and Kokonendji, C. C. (2016). Discrete dispersion models and their Tweedie asymptopics. *AStA Advances in Statistical Analysis* **100**, 43–78. MR3452285 https://doi.org/10.1007/s10182-015-0250-z

Jørgensen, B., Martínez, J. R. and Tsao, M. (1994). Asymptotic behaviour of the variance function. *Scandinavian Journal of Statistics* **213**, 223–243. MR1292637

Keeling, M. and Grenfell, B. (1999). Stochastic dynamics and a power law for measles variability. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **354**, 769–776.

Kemp, A. W. (1987). Families of discrete distributions satisfying Taylor's power law. *Biometrics* **43**, 693–699.

Kendal, W. (2017). $1/f$ noise and multifractality from bristlecone pine growth explained by the statistical convergence of random data. *Proceedings of the Royal Society A* **473**, 20160586. MR3621489 https://doi.org/10.1098/rspa.2016.0586

Kendal, W. S. (2002). Spatial aggregation of the Colorado potato beetle described by an exponential dispersion model. *Ecological Modelling* **151**, 261–269.

Kendal, W. S. (2003). An exponential dispersion model for the distribution of human single nucleotide polymorphisms. *Molecular Biology and Evolution* **20**, 579–590.

Kendal, W. S. (2004a). A scale invariant clustering of genes on human chromosome 7. *BMC Evolutionary Biology* **4**, 3.

Kendal, W. S. (2004b). Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecological Complexity* **1**, 193–209.

Kendal, W. S. (2014). Multifractality attributed to dual central limit-like convergence effects. *Physica A* **401**, 22–33. MR3294810 https://doi.org/10.1016/j.physa.2014.11.035

Kendal, W. S. (2015). Self-organized criticality attributed to a central limit-like convergence effect. *Physica A* **421**, 141–150. MR3294810 https://doi.org/10.1016/j.physa.2014.11.035

Kendal, W. S. and Jørgensen, B. (2011a). Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Physical Review E* **83**, 066115.

Kendal, W. S. and Jørgensen, B. (2011b). Tweedie convergence: A mathematical basis for Taylor's power law, $1/f$ noise and multifractality. *Physical Review E* **84**, 066120.

Kendal, W. S. and Jørgensen, B. (2015). A scale invariant distribution of the prime numbers. *Computation* **3**, 528–540.

Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society, Series B* **11**, 230–264. MR0034977

Kilpatrick, A. M. and Ives, A. R. (2003). Species interactions can explain Taylor's power law for ecological time series. *Nature* **422**, 65–68.

Mason, R. (1970). Comparison of flight aggregation in two species of southern ips (Coleoptera: Scolytidae). *Canadian Entomologist* **102**, 1036–1041.

Monro, J. (1967). The exploitation and conservation of resources by populations of insects. *Journal of Animal Ecology* **36**, 531–547.

Mora, M. (1990). Convergence of the variance functions of natural exponential families. *Annales de la Faculté des Sciences de Toulouse, Serie 5* **11**, 105–120. MR1191714

Naylor, A. (1959). An experimental analysis of dispersal in the flour beetle, *Tribolum confusum*. *Ecology* **40**, 453–465.

Nelder, J. A. and Wedderburn, R. W. M. (1972). General linearized models. *Journal of the Royal Statistical Society, Series A* **135**, 370–384.

Perry, J. N. (1994). Chaotic dynamics can generate Taylor's power law. *Proceedings of the Royal Society of London Series B, Biological Sciences* **257**, 221–226.

Petterle, R. R., Bonat, W. H., Kokonendji, C. C., Seganfredo, J. C., Morares, A. and Gomes da Silva, M. M. (2019). Double Poisson–Tweedie regression models. *The International Journal of Biostatistics* **15**, 20180119. MR3962660 https://doi.org/10.1515/ijb-2018-0119

Philippe, P. (1999). The scale-invariant spatial clustering of leukemia in San Francisco. *Journal of Theoretical Biology* **199**, 371–381.

Shan, C.-J., Cheng, W.-W., Liu, J.-B., Cheng, Y.-S. and Liu, T.-K. (2014). Scaling of geometric quantum discord close to a topological phase transition. *Scientific Reports* **4**, 1–6.

Shibuya, M. and Ouchi, Y. (1955). Pattern of spatial distribution of the soy bean pod gall midge in a soy bean field. *Oyo-Kontyu* **11**, 91–97.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature* **189**, 732–735.

Taylor, L. R. (1970). Aggregation and the transformation of counts of *Aphis fabae* Scop. on beans. *Annals of Applied Biology* **65**, 181–189.

Taylor, L. R. and Taylor, R. A. J. (1977). Aggregation, migration and population mechanics. *Nature* **265**, 415–421.

Taylor, L. R., Taylor, R. A. J., Woiwod, I. P. and Perry, J. N. (1983). Behavioural dynamics. *Nature* **303**, 801–804.

Taylor, L. R. and Woiwod, I. P. (1980). Temporal stability as a density-dependent species characteristic. *Journal of Animal Ecology* **49**, 209–224.

Taylor, L. R. and Woiwod, I. P. (1982). Comparative synoptic dynamics. I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters. *Journal of Animal Ecology* **51**, 879–906.

Taylor, L. R., Woiwod, I. P. and Perry, J. N. (1978). The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology* **47**, 383–406. MR0600375 https://doi.org/10.2307/4114

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions* (J. K. Ghosh and J. Roy, eds.), *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, 579–604. Calcutta, India: Indian Statistical Institute. MR0786162

Youdeowei, A. (1965). A note on the spatial distribution of the cocoa mirid *Sahlbergella singularis* Hagl. in a cocoa farm in western Nigeria. *Nigerian Agricultural Journal* **2**, 66–67.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 62–84
https://doi.org/10.1214/20-BJPS494

A class of asymmetric regression models for left-censored data

Helton Saulo[a], Jeremias Leão[b], Juvêncio Nobre[c] and Narayanaswamy Balakrishnan[d]

[a]*University of Brasília*
[b]*Federal University of Amazonas*
[c]*Federal University of Ceará*
[d]*McMaster University*

Abstract. A common assumption in the standard tobit model is the normality for the error distribution. However, asymmetry and bimodality may be present and alternative tobit models must be used in such cases. In this paper, we propose a tobit model based on the class of log-symmetric distributions, which includes as special cases heavy/light tailed distributions and bimodal distributions. We implement a likelihood-based approach for parameter estimation and consider a type of residual. We then discuss the problem of performing hypothesis tests within the proposed class by using the likelihood ratio and gradient statistics, which are particularly convenient for tobit models, as they do not require the information matrix. An elaborate Monte Carlo study is carried out for evaluating the performance of the maximum likelihood estimates, the likelihood ratio and gradient tests and the empirical distribution of the residuals. Finally, we illustrate the proposed methodology with the use of a real data set.

References

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics* **24**, 3–61. MR0739428 https://doi.org/10.1016/0304-4076(84)90074-5

Arellano, R., Castro, L. M., Gonzalez, G. and Muñoz, K. A. (2012). Student-t censored regression model: Properties and inference. *Statistical Methods and Applications* **21**, 453–473. MR2992913 https://doi.org/10.1007/s10260-012-0199-y

Balakrishnan, N. and Kundu, D. (2019). Birnbaum–Saunders distribution: A review of models, analysis, and applications. *Applied Stochastic Models in Business and Industry* **35**, 4–132 (with discussions). MR3915800 https://doi.org/10.1002/asmb.2348

Barros, M., Galea, M., Gonzalez, M. and Leiva, V. (2010). Influence diagnostics in the tobit censored response model. *Statistical Methods and Applications* **19**, 379–397. MR2673350 https://doi.org/10.1007/s10260-010-0135-y

Barros, M., Galea, M., Leiva, V. and Santos-Neto, M. (2018). Generalized tobit models: Diagnostics and application in econometrics. *Journal of Applied Statistics* **45**, 145–167. MR3736863 https://doi.org/10.1080/02664763.2016.1268572

Barros, M., Paula, G. and Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: Robustness and diagnostics. *Lifetime Data Analysis* **14**, 316–332. MR2516848 https://doi.org/10.1007/s10985-008-9085-1

Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327. MR0253493 https://doi.org/10.2307/3212003

Crow, E. L. and Shimizu, K. (1988). *Lognormal Distributions: Theory and Applications*. New York: Marcel Dekker. MR0939191

Desousa, M. F., Saulo, H., Leiva, V. and Scalco, P. (2018). On a tobit-Birnbaum–Saunders model with an application to antibody response to vaccine. *Journal of Applied Statistics* **45**, 932–955. MR3772119 https://doi.org/10.1080/02664763.2017.1322559

Díaz-García, J. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions. *Journal of Statistical Planning and Inference* **128**, 445–457. MR2102769 https://doi.org/10.1016/j.jspi.2003.11.007

Key words and phrases. Log-symmetric distributions, likelihood ratio test, gradient test, R software, tobit models.

Díaz-García, J. and Leiva, V. (2007). Erratum to a new family of life distributions based on the elliptically contoured distributions [Journal of Statistical Planning and Inference 128(2) (2005) 445–457]. *Journal of Statistical Planning and Inference* **137**, 1512–1513. MR2312807 https://doi.org/10.1016/j.jspi.2006.06.040

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

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed vs. expected Fisher information. *Biometrika* **65**, 457–487. MR0521817 https://doi.org/10.1093/biomet/65.3.457

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London, UK: Chapman and Hall. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

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

Helsel, D. R. (2011). *Statistics for Censored Environmental Data Using Minitab and R*. Hoboken, New Jersey: John Wiley & Sons.

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

Johnson, N., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1*, 2nd ed. New York: John Wiley & Sons. MR1326603

Johnson, N., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*, 2nd ed. New York: John Wiley & Sons. MR1326603

Jones, M. C. (2008). On reciprocal symmetry. *Journal of Statistical Planning and Inference* **138**, 3039–3043. MR2526220 https://doi.org/10.1016/j.jspi.2007.11.006

Kano, Y., Berkane, M. and Bentler, P. M. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association* **88**, 135–143. MR1212483

Leiva, V., Barros, M., Paula, G. and Galea, M. (2007). Influence diagnostics in log-Birnbaum–Saunders regression models with censored data. *Computational Statistics and Data Analysis* **51**, 5694–5707. MR2407671 https://doi.org/10.1016/j.csda.2006.09.020

Lemonte, A. (2016). *The Gradient Test: Another Likelihood-Based Test*. London, UK: Academic Press. MR3618685

Lemonte, A. and Ferrari, S. (2011). Testing hypotheses in the Birnbaum–Saunders distribution under type-II censored samples. *Computational Statistics and Data Analysis* **55**, 2388–2399. MR2786997 https://doi.org/10.1016/j.csda.2011.02.005

Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics: Theory and Methods* **41**, 1165–1182. MR1450228 https://doi.org/10.1080/03610929708831974

Marshall, A. and Olkin, I. (2007). *Life Distributions*. New York: Springer. MR2344835

Martínez-Flores, G., Bolfarine, H. and Gómez, H. W. (2013a). The alpha-power tobit model. *Communications in Statistics: Theory and Methods* **42**, 633–643. MR3211940 https://doi.org/10.1080/03610926.2011.630770

Martínez-Flores, G., Bolfarine, H. and Gómez, H. W. (2013b). Asymmetric regression models with limited responses with an application to antibody response to vaccine. *Biometrical Journal* **55**, 156–172. MR3045838 https://doi.org/10.1002/bimj.201100116

Massuia, M. B., Cabral, C. R. B., Matos, L. A. and Lachos, V. H. (2015). Influence diagnostics for student-t censored linear regression models. *Statistics* **49**, 1074–1094. MR3378032 https://doi.org/10.1080/02331888.2014.958489

Medeiros, 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

Mittelhammer, R. C., Judge, G. G. and Miller, D. J. (2000). *Econometric Foundations*. New York: Cambridge University Press. MR1789434

Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics* **51**, 1570–1578.

Ng, H. K. T., Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics and Data Analysis* **43**, 283–298. MR1996813 https://doi.org/10.1016/S0167-9473(02)00254-2

Ott, W. R. (1990). A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air and Waste Management Association* **40**, 1378–1383.

Podlaski, R. (2008). Characterization of diameter distribution data in near-natural forests using the Birnbaum–Saunders distribution. *Canadian Journal of Forest Research* **18**, 518–527.

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

Rieck, J. and Nedelman, J. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics* **3**, 51–60.

Saulo, H., Leão, J., Leiva, V. and Aykroyd, R. G. (2019). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data. *Statistical Papers* **60**, 1605–1629. MR4017025 https://doi.org/10.1007/s00362-017-0888-6

Saulo, H., Vila, R., Vilca, F. and Martínez, J. L. (2020). On asymmetric regression models with allowance for temporal dependence. *Journal of Statistical Theory and Practice* **14**, 40. MR4110661 https://doi.org/10.1007/s42519-020-00104-9

Terrell, G. (2002). The gradient statistic. *Computing Science and Statistics* **34**, 206–215.

Thorarinsdottir, T. L. and Gneiting, T. (2010). Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society, Series A* **173**, 371–388. MR2751882 https://doi.org/10.1111/j.1467-985X.2009.00616.x

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36. MR0090462 https://doi.org/10.2307/1907382

Vanegas, L. H. and Paula, G. A. (2015). A semiparametric approach for joint modeling of median and skewness. *Test* **24**, 110–135. MR3314576 https://doi.org/10.1007/s11749-014-0401-7

Vanegas, L. H. and Paula, G. A. (2016a). An extension of log-symmetric regression models: R codes and applications. *Journal of Statistical Simulation and Computation* **86**, 1709–1735. MR3473841 https://doi.org/10.1080/00949655.2015.1081689

Vanegas, L. H. and Paula, G. A. (2016b). 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

Ventura, M., Saulo, H., Leiva, V. and Monsueto, S. E. (2019). Log-symmetric regression models: Information criteria and application to movie business and industry data. *Applied Stochastic Models in Business and Industry* **35**, 963–977. MR3994368 https://doi.org/10.1002/asmb.2433

Villegas, C., Paula, G. and Leiva, V. (2011). Birnbaum–Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability* **60**, 748–758.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**, 60–62.

Yenilmez, I., Mert Kantar, Y. and Acitaş, C. (2018). Estimation of censored regression model in the case of non-normal error. *Sigma Journal of Engineering and Natural Sciences* **36**, 513–521.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 85–100
https://doi.org/10.1214/20-BJPS488

Regression models for change point data in extremes

Fernando Ferraz do Nascimento[a] **and Alan da Silva Assunção**[b]

[a]*Universidade Federal do Piauí*
[b]*Universidade Federal do Ceará*

Abstract. Many extreme events are characterized by being always susceptible to outside influences that will modify their behavior at some point in time. The change point tool has been used in statistical models to detect when these changes occur. This paper presents a model based on a Bayesian approach that describes the behavior of extreme data regarding river quota, which may present more than one change point. In each one of the regimes, the GEV distribution is adjusted and each GEV parameter of each regime is written in function of presence of covariates. In the applications proposed here, the results showed that the model was able to accurately estimate the actual amount of change points in the series, and also showed that it was extremely important to consider them in the analysis, since it was verified that after the change of regime, the levels of return have changed considerably. The results were also able to show which months the occurrence of an extreme event is greater.

References

Barry, D. and Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association* **88**, 309–319. MR1212493

Beaulieu, C., Chen, J. and Sarmiento, J. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **370**, 1228–1249. 1962.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian analysis to change point problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **41**, 389–405. MR1278223

Castellanos, M. E. and Cabras, S. (2007). A default Bayesian procedure for the generalized Pareto distribution. *Journal of Statistical Planning and Inference* **137**, 473–483. MR2298951 https://doi.org/10.1016/j.jspi.2006.01.006

Chaves-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **54**, 207–222. MR2134607 https://doi.org/10.1111/j.1467-9876.2005.00479.x

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics* **86**, 221–241. MR1649222 https://doi.org/10.1016/S0304-4076(97)00115-2

Coles, S. (2001). *Introduction to Statistical Modelling of Extreme Values*. Berlin: Springer. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

Csörgõ, M. and Horvát, L. (1997). *Limit Theorems in Change Point Analysis*. Chishester: Wiley. MR2743035

Dupuis, D. J., Sun, Y. and Wang, H. J. (2015). Detecting change-points in extremes. *Statistics and Its Interface* **8**, 19–31. MR3320386 https://doi.org/10.4310/SII.2015.v8.n1.a3

Fisher, R. A. and Tippet, L. H. C. (1928). On the estimation of the frequency distributions of the largest and smallest sumber of a sample. *Proceedings of the Cambridge Philosophycal Society* **24**, 180–190.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Baton Rouge: Chapman and Hall/CRC. MR2260716

Jarusková, D. and Rencová, M. (2008). Analysis of annual maximal and minimal temperatures for some European cities by change point methods. *Environmetrics* **19**, 221–233. MR2420467 https://doi.org/10.1002/env.865

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society* **81**, 158–171.

Katz, R. W. and Brown, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change* **21**, 289–302.

Ko, S. I. M., Chong, T. L. C. and Ghosh, P. (2015). Dirichlet Process Hidden Markov Multiple Change-point Model. *Bayesian Analysis* **10**, 275–296. MR3420883 https://doi.org/10.1214/14-BA910

Key words and phrases. Extreme events, change points, regression structure, Bayesian approach, river quota.

Mendes, B. V. M. (2004) *Introdução a análise de eventos extremos*. *Rio de Janeiro, E-papers*. MR2117963

Nascimento, F. F. and Bourguignon, M. (2020). Bayesian time-varying quantile regression to extremes. *Environmetrics* **31**, e2596. MR4075862 https://doi.org/10.1002/env.2596

Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2011). Regression models for exceedance data via the full likelihood. *Environmental and Ecological Statistics* **18**, 495–512. MR2832907 https://doi.org/10.1007/s10651-010-0148-6

Nascimento, F. F. and Silva, W. V. M. (2017). A Bayesian model for multiple change point to extremes, with application to environmental and financial data. *Journal of Applied Statistics* **44**, 2410–2426. MR3683052 https://doi.org/10.1080/02664763.2016.1254733

Parmesan, C., Root, T. L. and Willig, M. R. (2000). Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society* **81**, 443.

Stephens, D. A. (1994). Bayesian retrospective multiple-changepoint identification. *Applied Statistics* **43**, 159–178.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**, 407–426. MR2749848 https://doi.org/10.1007/s10651-007-0078-0

Silva, W. V. M., Nascimento, F. F. and Bourguignon, M. (2020). A change-point model for the r-largest order statistics with application to river quota. *Applied Mathematical Modelling*, 666–679. MR4070321 https://doi.org/10.1016/j.apm.2020.01.064

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Papers, Vol. II*, 271–294. Providence, R.I.: American Mathematical Society. Reprinted in Selected.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 101–118
https://doi.org/10.1214/20-BJPS497

The snap, crackle and pop of solar flares explained

Wayne S. Kendal
University of Ottawa

In memoriam: Prof. Bent Jørgensen (1954–2015), who contributed the fundamentals on which this study was based.

Abstract. The irregular fluctuations of solar flare emissions, as determined from terrestrial neutron monitors, remains poorly understood. These records empirically revealed a temporally-related variance to mean power law, $1/f$ noise and a non-Gaussian distribution, all features indicative of self-organized criticality, a theory of how derministic dynamical systems can spontaneously evolve to unstable states that express erratic changes. The non-Gaussian distribution found here approximated a Tweedie compound Poisson exponential dispersion model, a statistical distribution characterized by a variance to mean power law that itself can imply $1/f$ noise. Tweedie exponential dispersion models serve a primary role in statistical theory as foci for weak convergence for a wide range of random distributions, a role which supports an alternative conjecture to explain the solar flare fluctuations as being based on random processes rather than a deterministic system.

References

Abraham, R. H. (1996). Visualization techniques for cellular dynamata. In *Introduction to Nonlnear Physics* (L. Li, ed.) 296–307. New York: Springer.

Anderson, R. M. and May, R. M. (1988). Epidemiological parameters of HIV transmission. *Nature* **333**, 514–519.

Bak, P. and Chen, K. (1991). Self-organized criticality. *Scientific American* **264**, 46–53.

Bak, P., Tang, C. and Wiesenfeld, K. (1987). Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters* **59**, 381–384.

Bak, P., Tang, C. and Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A* **38**, 364–374. MR0949160 https://doi.org/10.1103/PhysRevA.38.364

Bury, K. (1999). *Statistical Distributions in Engineering*. Cambridge: Cambridge University Press.

Dutta, P. and Horn, P. M. (1981). Low-frequency fluctuations in solids: $1/f$ noise. *Reviews of Modern Physics* **53**, 497–516.

Eisler, Z., Bartos, I. and Kertesz, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics* **57**, 89–142.

Eubank, S. G. and Farmer, J. D. (1996). Introduction to dynamical systems. In *Introduction to Nonlnear Physics* (L. Li, ed.) 55–105. New York: Springer. MR1431055

Fischer, H. (2011). *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. New York: Springer. MR2743162 https://doi.org/10.1007/978-0-387-87857-7

Frigg, R. (2003). Self-organized criticality—what it is and what it isn't. *Studies in History and Philosophy of Science* **34**, 613–632.

Garcia Canal, C., Hojvat, C. and Taritina, T. (2012). Scaler mode of the Auger observatory and sunspots. *The Astrophysical Journal Supplement Series* **202**, 1–6.

Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review* **46**, 269–282. MR2114455 https://doi.org/10.1137/S0036144501394387

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G., Peng, C.-K. and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiological signals. *Circulation* **101**, e215–e220.

Hayes, B. (2003). The spectrum of Riemannium. *American Scientist* **91**, 296–300.

Key words and phrases. Solar cosmic rays, emergent property, Taylor's power law, multifractals, weak convergence.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall. MR1462891
Jørgensen, B. and Kokonendji, C. C. (2011). Dispersion models for geometric sums. *Brazilian Journal of Probability and Statistics* **25**, 263–293. MR2832887 https://doi.org/10.1214/10-BJPS136
Jørgensen, B., Martínez, J. R. and Tsao, M. (1994). Asymptotic behaviour of the variance function. *Scandinavian Journal of Statistics* **213**, 223–243. MR1292637
Kantelhardt, J., Koscielny-Bunde, E., Rego, H., Havlin, S. and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A* **295**, 441–454.
Keeling, M. and Grenfell, B. (1999). Stochastic dynamics and a power law for measles variability. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **354**, 769–776.
Kendal, W. (2017). $1/f$ noise and multifractality from bristlecone pine growth explained by the statistical convergence of random data. *Proceedings—Royal Society Mathematical, Physical and Engineering Sciences* **473**, 20160586. MR3621489 https://doi.org/10.1098/rspa.2016.0586
Kendal, W. S. (2001). A stochastic model for the self-similar heterogeneity of regional organ blood flow. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 837–841. MR1814761 https://doi.org/10.1073/pnas.021347898
Kendal, W. S. (2014). Multifractality attributed to dual central limit-like convergence effects. *Physica A* **401**, 22–33. MR3294810 https://doi.org/10.1016/j.physa.2014.11.035
Kendal, W. S. (2015). Self-organized criticality attributed to a central limit-like convergence effect. *Physica A* **421**, 141–150. MR3294810 https://doi.org/10.1016/j.physa.2014.11.035
Kendal, W. S. and Jørgensen, B. (2011). Tweedie convergence: A mathematical basis for Taylor's power law, $1/f$ noise and multifractality. *Physical Review E* **84**, 066120.
Kendal, W. S. and Jørgensen, B. (2015). A scale invariant distribution of the prime numbers. *Computation* **3**, 528–540.
Kozubowski, T. J. and Podgórski, K. (2001). Asymmetric Laplace laws and modeling financial data. *Mathematical and Computer Modelling* **34**, 1003–1021. MR1858834 https://doi.org/10.1016/S0895-7177(01)00114-5
Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V. (1994). On the self-similar nature of ethernet traffic. *IEEE/ACM Transactions on Networking* **2**, 1–15.
Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**, 422–437. MR0242239 https://doi.org/10.1137/1010093
McQuarrie, D. A. (1976). *Statistical Mechanics*. New York: Harper & Row.
Musha, T. and Yamamoto, M. (1997). $1/f$ fluctuations in biological systems. In *Proceedings—19th International Conference—IEEE/EMBS*, 2692–2697. *Chicago, IL*.
Muzy, J. F., Bacry, E. and Arnedo, A. (1993a). Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E* **47**, 875–884.
Muzy, J. F., Bacry, E. and Arnedo, A. (1993b). Multifractal formalism for fractal signals: The struture-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E. Statistical, Nonlinear, and Soft Matter Physics* **47**, 875–884.
Nurujjaman, M. and Sekar Iyengar, A. (2007). Realization of SOC behavior in a dc glow discharge plasma. *Physics Letters A* **360**, 717–721.
Pathria, R. K. (1972). *Statistical Mechanics*. Oxford: Pegamon Press.
Petterle, R. R., Bonat, W. H., Kokonendji, C. C., Seganfredo, J. C., Morares, A. and Gomes da Silva, M. M. (2019). Double Poisson–Tweedie regression models. *The International Journal of Biostatistics* **15**, 20180119. MR3962660 https://doi.org/10.1515/ijb-2018-0119
Philippe, P. (1999). The scale-invariant spatial clustering of leukemia in San Francisco. *Journal of Theoretical Biology* **199**, 371–381.
Press, W. (1978). Flicker noises in astronomy and elsewhere. *Comments on Astrophysics* **7**, 103–119.
Shan, C.-J., Cheng, W.-W., Liu, J.-B., Cheng, Y.-S. and Liu, T.-K. (2014). Scaling of geometric quantum discord close to a topological phase transition. *Scientific Reports* **4**, 1–6.
Stanley, H. E. and Meakin, P. (1988). Multifractal phenomena in physics and chemistry. *Nature* **335**, 405–409.
Susskind, L. and Hrabovsky, G. (2013). *The Theoretical Minimum*. New York: Basic Books.
Szendro, P., Vincze, G. and Szasz, A. (2001). Pink-noise behaviour of biosystems. *European Biophysics Journal* **30**, 227–231.
Takahashi, K. (1989). Detection of solar neutrons by ground-based neutron monitor. *Space Science Reviews* **51**, 123–146.
Taylor, L. R., Taylor, R. A. J., Woiwod, I. P. and Perry, J. N. (1983). Behavioural dynamics. *Nature* **303**, 801–804.
Tsybakov, B. and Georganas, N. D. (1997). On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution. *IEEE/ACM Transactions on Networking* **5**, 397–409. MR1664070 https://doi.org/10.1109/18.705538

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions* (J. K. Ghosh and J. Roy, eds.). *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, 579–604. Calcutta, India: Indian Statistical Institute. MR0786162

Usoskin, I., Kovaltsov, G., Kananen, H. and Tanskanen, P. (1997). The world neutron monitor network as a tool for the study of solar neutrons. *Annales Geophysicae* **15**, 375–386.

Vespignani, A. and Zapperi, S. (1998). How self-organized criticality works: A unified mean-field picture. *Physical Review E* **57**, 6345–6362. MR1628210 https://doi.org/10.1103/PhysRevE.57.6345

Werndl, C. (2009). What are the new implications of chaos for unpredictability? *British Journal for the Philosophy of Science* **60**, 195–220. MR2485508 https://doi.org/10.1093/bjps/axn053

Wigner, E. P. (1967). Random matrices in physics. *SIAM Review* **9**, 1–23.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 119–137
https://doi.org/10.1214/20-BJPS490

A bivariate fatigue-life regression model and its application to fracture of metallic tools

Helton Saulo[a], Jeremias Leão[b], Víctor Leiva[c], Roberto Vila[a] and Vera Tomazella[d]

[a]*Universidade de Brasília*
[b]*Universidade Federal do Amazonas*
[c]*Pontificia Universidad Católica de Valparaíso*
[d]*Universidade Federal de São Carlos*

Abstract. The Birnbaum–Saunders distribution has been widely used to model reliability and fatigue data. In this paper, we propose a regression of generalized linear models type based on a new bivariate Birnbaum–Saunders distribution. This is parameterized in terms of its means and allows data to be described in their original scale. We estimate the model parameters and carry out inference with the maximum likelihood method. A case study with real-world reliability data is conducted for motivating our investigation, illustrating the potential applications of the proposed results. We obtain a predictive model which can be a useful addition to the tool-kit of diverse practitioners, reliability engineers, applied statisticians, and data scientists.

References

Athayde, E. (2017). A characterization of the Birnbaum–Saunders distribution. *REVSTAT Statistical Journal* **15**, 333–354. MR3684667

Aykroyd, R. G., Leiva, V. and Marchant, C. (2018). Multivariate Birnbaum–Saunders distributions: Modelling and applications. *Risks* **6**, article 21, 1–25. MR3430824 https://doi.org/10.1016/B978-0-12-803769-0.00001-7

Balakrishnan, N. and Kundu, D. (2019). Birnbaum–Saunders distribution: A review of models, analysis, and application. *Applied Stochastic Models in Business and Industry* **35**, 4–49. MR3915800 https://doi.org/10.1002/asmb.2348

Barros, M., Leiva, V., Ospina, R. and Tsuyuguchi, A. (2014). Goodness-of-fit tests for the Birnbaum–Saunders distribution with censored reliability data. *IEEE Transactions on Reliability* **63**, 543–554.

Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327. MR0253493 https://doi.org/10.2307/3212003

Carrasco, J. M. F., Figueroa-Zuniga, J. I., Leiva, V., Riquelme, M. and Aykroyd, R. G. (2020). An errors-in-variables model based on the Birnbaum–Saunders and its diagnostics with an application to earthquake data. *Stochastic Environmental Research and Risk Assessment* **34**, 369–380.

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

Dasilva, A., Dias, R., Leiva, V., Marchant, C. and Saulo, H. (2020). Birnbaum–Saunders regression models: A comparative evaluation of three approaches. *Journal of Statistical Computation and Simulation* **90**, 2552–2570. MR4145355 https://doi.org/10.1080/00949655.2020.1782912

Desousa, M., Saulo, H., Leiva, V. and Santos-Neto, M. (2020). On a new mixture-based regression model: Simulation and application to data with high censoring. *Journal of Statistical Computation and Simulation* **90**, 2861–2877. https://doi.org/10.1080/00949655.2020.1790560

Ferreira, M. (2013). A study of exponential-type tails applied to Birnbaum–Saunders models. *Chilean Journal of Statistics* **4**, 87–97. MR3054226

Garcia-Papani, F., Leiva, V., Uribe-Opazo, M. A. and Aykroyd, R. G. (2018). Birnbaum–Saunders spatial regression models: Diagnostics and application to chemical data. *Chemometrics and Intelligent Laboratory Systems* **177**, 114–128.

Garcia-Papani, F., Uribe-Opazo, M. A., Leiva, V. and Aykroyd, R. G. (2017). Birnbaum–Saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment* **131**, 105–124.

Key words and phrases. Bivariate Birnbaum–Saunders distribution, fatigue data, maximum likelihood method, predictive models, R software.

Huerta, M., Leiva, V., Liu, S., Rodriguez, M. and Villegas, D. (2019). On a partial least squares regression model for asymmetric data with a chemical application in mining. *Chemometrics and Intelligent Laboratory Systems* **190**, 55–68.

Ibacache-Pulgar, G., Paula, G. A. and Galea, M. (2014). On influence diagnostics in elliptical multivariate regression models with equicorrelated random errors. *Statistical Modelling* **16**, 14–21. MR3110885 https://doi.org/10.1016/j.stamet.2013.06.001

Khosravi, M., Kundu, D. and Jamalizadeh, A. (2015). On bivariate and mixture of bivariate Birnbaum–Saunders distributions. *Statistical Methodology* **23**, 1–17. MR3278798 https://doi.org/10.1016/j.stamet.2014.07.001

Kocherlakota, S. (1986). The bivariate inverse Gaussian distribution: An introduction. *Communications in Statistics Theory and Methods* **15**, 1081–1112. MR0836585 https://doi.org/10.1080/03610928608829171

Kundu, D. (2015a). Bivariate log-Birnbaum–Saunders distribution. *Statistics* **49**, 900–917. MR3367730 https://doi.org/10.1080/02331888.2014.915840

Kundu, D. (2015b). Bivariate sinh-normal distribution and a related model. *Brazilian Journal of Probability and Statistics* **20**, 590–607. MR3355749 https://doi.org/10.1214/13-BJPS235

Kundu, D., Balakrishnan, N. and Jamalizadeh, A. (2010). Bivariate Birnbaum–Saunders distribution and associated inference. *Journal of Multivariate Analysis* **101**, 113–125. MR2557622 https://doi.org/10.1016/j.jmva.2009.05.005

Leiva, V. (2016). *The Birnbaum–Saunders Distribution*. New York, US: Academic Press. MR3430824 https://doi.org/10.1016/B978-0-12-803769-0.00001-7

Leiva, V. (2019). An interview with Sam C. Saunders. *Applied Stochastic Models in Business and Industry* **35**, 133–137. MR3915818 https://doi.org/10.1002/asmb.2429

Leiva, V., Aykroyd, R. G. and Marchant, C. (2019). Discussion of "Birnbaum–Saunders distribution: A review of models, analysis, and applications" and a novel multivariate data analytics for an economics example in the textile industry. *Applied Stochastic Models in Business and Industry* **35**, 112–117. MR3915814 https://doi.org/10.1002/asmb.2401

Leiva, V., Sánchez, L., Galea, M. and Saulo, H. (2020). Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression. *Stochastic Environmental Research and Risk Assessment* **34**, 1457–1471.

Leiva, V., Santos-Neto, M., Cysneiros, F. J. A. and Barros, M. (2014a). Birnbaum–Saunders statistical modelling: A new approach. *Statistical Modelling* **14**, 21–48. MR3179546 https://doi.org/10.1177/1471082X13494532

Leiva, V., Saulo, H., Leao, J. and Marchant, C. (2014b). A family of autoregressive conditional duration models applied to financial data. *Computational Statistics & Data Analysis* **79**, 175–191. MR3227995 https://doi.org/10.1016/j.csda.2014.05.016

Leiva, V., Saulo, H., Souza, R., Aykroyd, R. G. and Vila, R. (2021). A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *Journal of Forecasting*. To appear. https://doi.org/10.1002/for.2718

Leiva, V. and Saunders, S. C. (2015). Cumulative damage models. In *Wiley StatsRef: Statistics Reference Online* (N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels, eds.) 1–10. https://doi.org/10.1002/9781118445112.stat02136.pub2

Leiva, V. and Vivanco, J. F. (2017). Fatigue models. In *Wiley StatsRef: Statistics Reference Online* (N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels, eds.) 1–10. https://doi.org/10.1002/9781118445112.stat02141.pub2

Lepadatu, D., Kobi, A., Hambli, R. and Barreau, A. (2005). Lifetime multiple response optimization of metal extrusion die. In *Proceedings of the Annual Reliability and Maintainability Symposium*, 37–42. Institute of Electrical and Electronics Engineers.

Marchant, C., Leiva, V. and Cysneiros, F. J. A. (2016). A multivariate log-linear model for Birnbaum–Saunders distributions. *IEEE Transactions on Reliability* **65**, 816–827. MR3430824 https://doi.org/10.1016/B978-0-12-803769-0.00001-7

Martinez, S., Giraldo, R. and Leiva, V. (2019). Birnbaum–Saunders functional regression models for spatial data. *Stochastic Environmental Research and Risk Assessment* **30**, 1765–1780.

Ng, H. K. T., Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis* **43**, 283–298. MR1996813 https://doi.org/10.1016/S0167-9473(02)00254-2

Sánchez, L., Leiva, V., Galea, M. and Saulo, H. (2021). Birnbaum–Saunders quantile regression and its diagnostics with application to economic data. *Applied Stochastic Models in Business and Industry*. To appear. MR4171608 https://doi.org/10.1080/03610926.2019.1626425

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V. and Ahmed, S. (2012). On new parameterizations of the Birnbaum–Saunders distribution. *Pakistan Journal of Statistics* **28**, 1–26. MR2931825

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V. and Barros, M. (2014). A reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications. *REVSTAT Statistical Journal* **12**, 247–272. MR3301849

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V. and Barros, M. (2016). Reparameterized Birnbaum–Saunders regression models with varying precision. *Electronic Journal of Statistics* **10**, 2825–2855. MR3553913 https://doi.org/10.1214/16-EJS1187

Saulo, H., Leão, J., Leiva, V. and Aykroyd, R. G. (2019). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data. *Statistical Papers* **60**, 1605–1629. MR4017025 https://doi.org/10.1007/s00362-017-0888-6

Saulo, H., Leão, J., Leiva, V., Vila, R. and Tomazella, V. (2020). On mean-based bivariate Birnbaum–Saunders distributions: Properties, inference and application. *Communications in Statistics Theory and Methods* **49**, 6032–6056. MR4171608 https://doi.org/10.1080/03610926.2019.1626425

Vila, R., Leão, J., Saulo, H., Shahzad, M. N. and Santos-Neto, M. (2020). On a bimodal Birnbaum–Saunders distribution with applications to lifetime data. *Brazilian Journal of Probability and Statistics* **34**, 495–518. MR4124538 https://doi.org/10.1214/19-BJPS448

Vilca, F. (2019). Discussion of "Birnbaum–Saunders distribution: A review of models, analysis, and applications". *Applied Stochastic Models in Business and Industry* **35**, 100–103. MR3915811 https://doi.org/10.1002/asmb.2411

Vilca, F., Balakrishnan, N. and Zeller, C. (2014a). A robust extension of the bivariate Birnbaum–Saunders distribution and associated inference. *Journal of Multivariate Analysis* **124**, 418–435. MR3147335 https://doi.org/10.1016/j.jmva.2013.11.005

Vilca, F., Balakrishnan, N. and Zeller, C. (2014b). The bivariate sinh-elliptical distribution with applications to Birnbaum–Saunders distribution and associated regression and measurement error models. *Computational Statistics & Data Analysis* **80**, 1–16. MR3240471 https://doi.org/10.1016/j.csda.2014.06.001

Vilca, F., Romeiro, R. G. and Balakrishnan, N. (2016). A bivariate Birnbaum–Saunders regression model. *Computational Statistics & Data Analysis* **97**, 169–183. MR3447043 https://doi.org/10.1016/j.csda.2015.12.003

Villegas, C., Paula, G. A. and Leiva, V. (2011). Birnbaum–Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability* **60**, 748–758.

Yap, B. W. and Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation* **81**, 2141–2155. MR2864195 https://doi.org/10.1080/00949655.2010.520163

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 138–157
https://doi.org/10.1214/20-BJPS469

Confidence intervals of the index C_{pk} for normally distributed quality characteristics using classical and Bayesian methods of estimation

Mahendra Saha[a], Sanku Dey[b], Abhimanyu Singh Yadav[c] and Sajid Ali[d]

[a]*Central University of Rajasthan*
[b]*St. Anthony's College*
[c]*Banaras Hindu University*
[d]*Quaid-I-Azam University*

Abstract. One of the indicators for evaluating the capability of a process potential and performance in an effective way is the process capability index (PCI). It is of great significance to quality control engineers as it quantifies the relation between the actual performance of the process and the pre-set specifications of the product. Most of the traditional PCIs performed well when process follows the normal behaviour. In this article, we consider a process capability index, C_{pk}, suggested by Kane (*Journal of Quality Technology* **18** (1986) 41–52) which can be used for normal random variables. The objective of this article is three fold: First, we address different methods of estimation of the process capability index C_{pk} from frequentist approaches for the normal distribution. We briefly describe different frequentist approaches, namely, maximum likelihood estimators, least squares and weighted least squares estimators, maximum product of spacings estimators, Cramèr–von-Mises estimators, Anderson–Darling estimators and Right-Tail Anderson–Darling estimators and compare them in terms of their mean squared errors using extensive numerical simulations. Second, we compare three parametric bootstrap confidence intervals (BCIs) namely, standard bootstrap, percentile bootstrap and bias-corrected percentile bootstrap. Third, we consider Bayesian estimation under squared error loss function using normal prior for location parameter and inverse gamma for scale parameter for the considered model. Monte Carlo simulation study has been carried out to compare the performances of the classical BCIs and highest posterior density (HPD) credible intervals of C_{pk} in terms of average widths and coverage probabilities. Finally, two real data sets have been analyzed for illustrative purposes.

References

Ali, S. and Riaz, M. (2014). On the generalized process capability under simple and mixture models. *Journal of Applied Statistics* **41**, 832–852. MR3291789 https://doi.org/10.1080/02664763.2013.856386

Alkasasbeh, M. R. and Raqab, M. Z. (2009). Estimation of the generalized logistic distribution parameters: Comparative study. *Statistical Methodology* **2009**, 262–279. MR2649656 https://doi.org/10.1016/j.stamet.2008.10.001

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics* **23**, 193–212. MR0050238 https://doi.org/10.1214/aoms/1177729437

Barnett, N. S. (1988). Process control and product quality: The C_p and C_{pk} revisited. *International Journal of Quality and Reliability Management* **7**, 34–43.

Chan, L. K., Cheng, S. W. and Spiring, F. A. (1988). A new measure of process capability: C_{pm}. *Journal of Quality Technology* **30**, 162–175.

Key words and phrases. Process capability index, normal distribution, different classical methods of estimation, bootstrap confidence intervals, sensitivity analysis.

Chen, J. P. and Chen, K. S. (2004). Comparison of two process capabilities by using indices Cpm: An application to a color STN display. *International Journal of Quality and Reliability Management* **21**, 90–101.

Chen, J. P. and Tong, L. I. (2003). Bootstrap confidence interval of the difference between two process capability indices. *The International Journal of Advanced Manufacturing Technology* **21**, 249–256.

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

Cheng, R. C. H. and Amin, N. A. K. (1979). Maximum product-of-spacings estimation with applications to the log-normal distribution. Math Report, Department of Mathematics, UWIST, Cardi.

Dennis, J. E. and Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Non-linear Equations*. Englewood Cliffs, NJ: Prentice-Hall. MR0702023

Dey, S., Ali, S. and Park, C. (2015). Weighted exponential distribution: Properties and different methods of estimation. *Journal of Statistical Computation and Simulation* **85**, 3641–3661. MR3401100 https://doi.org/10.1080/00949655.2014.992346

Dey, S., Alzaatreh, A., Zhang, C. and Kumar, D. (2017b). A new extension of generalized exponential distribution with application to ozone data. *Ozone Science and Engineering* **39**, 273–285.

Dey, S., Dey, T. and Kundu, D. (2014). Two-parameter Rayleigh distribution: Different methods of estimation. *American Journal of Mathematical and Management Sciences* **33**, 55–74.

Dey, S., Josmar, M. J. and Nadarajah, S. (2017c). Kumaraswamy distribution: Different methods of estimation. *Computational and Applied Mathematics*. https://doi.org/10.1007/s40314-017-0441-1

Dey, S., Kumar, D., Ramos, P. L. and Louzada, F. (2017a). Exponentiated Chen distribution: Properties and estimation. *Communications in Statistics Simulation and Computation* **46**, 8118–8139. MR3765022 https://doi.org/10.1080/03610918.2016.1267752

Dey, S., Moala, F. A. and Kumar, D. (2018). Statistical properties and different methods of estimation of Gompertz distribution with application. *Journal of Statistics & Management Systems* **21**, 839–876.

Dey, S. and Saha, M. (2019). Bootstrap confidence intervals of generalized process capability index Cpyk using different methods of estimation. *Journal of Applied Statistics* **46**, 1843–1869. MR3953598 https://doi.org/10.1080/02664763.2019.1572721

Dey, S., Saha, M., Maiti, S. S. and Jun, H. C. (2017). Bootstrap confidence intervals of generalized process capability Cpyk for Lindley and power Lindley distributions. *Communications in Statistics-Simulation and Computation*. MR3765049 https://doi.org/org/10.1080/03610918.2017.1280166

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Re-Sampling Plans*. *SIAM, CBMS-NSF Monograph.* **38**. Pennsylvania: SIAM: Philadelphia. MR0659849

Flaig, J. J. (1999). Process capability sensitivity analysis. *Quality Engineering* **11**, 587–592.

Gensidy, A. (1985). C_p and C_{pk}. *Quality Progress* **18**, 7–8.

Hsiang, T. C. and Taguchi, G. (1985). A tutorial on quality control and assurance. In *Annual Meeting on the American Statistical Association. Las Vegas, NV*. Unpublished presentation.

Huiming, Z. Y., Jun, Y. and Liya, H. (2007). Bayesian evaluation approach for process capability based on sub samples. In *IEEE International Conference on Industrial Engineering and Engineering Management*, 1200–1203. *Singapore*.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314.

Juran, J. M. (1974). *Juran's Quality Control Handbook*, 3rd ed. New York, USA: McGraw-Hill.

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology* **18**, 41–52.

Kargar, M., Mashinchi, M. and Parchami, A. (2014). A Bayesian approach to capability testing based on Cpk with multiple samples. *Quality and Reliability Engineering International* **30**, 615–621.

Kashif, M., Aslam, M., Al-Marshadi, A. H. and Jun, C. H. (2016). Capability indices for non-normal distribution using Gini's mean difference as measure of variability. *IEEE Access* **4**, 7322–7330.

Kashif, M., Aslam, M., Rao, G. S., Al-Marshadi, A. H. and Jun, C. H. (2017). Bootstrap confidence intervals of the modified process capability index for Weibull distribution. *Arab Journal of Science and Engineering*. https://doi.org/10.1007/s13369-017-2562-7

Kotz, S. and Johnson, N. L. (1999). Delicate relations among the basic process capability indices C_p, C_{pk}, C_{pm} and their modifications. *Communications in Statistics Theory and Methods* **26**, 849–861. MR1713633 https://doi.org/10.1080/03610919908813581

Kotz, S. and Johnson, N. L. (2002). Process capability indices—A review. *Journal of Quality Technology* **34**, 2–19.

Kotz, S. and Lovelace, C. (1998). *Introduction to Process Capability Indices*. London, UK: Arnold.

Kundu, D. and Raqab, M. Z. (2005). Generalized Rayleigh distribution: Different methods of estimation. *Computational Statistics & Data Analysis* **49**, 187–200. MR2129172 https://doi.org/10.1016/j.csda.2004.05.008

Leiva, V., Marchant, C. and Saulo, H. (2014). Capability index for Birnbaum–Saunders process applied to electronic and food industries. *Journal of Applied Statistics* **41**, 1881–1902. MR3292616 https://doi.org/10.1080/02664763.2014.897690

Lin, T. Y., Wu, C. W., Chen, J. C. and Chiou, Y. H. (2011). Applying Bayesian approach to assess process capability for asymmetric tolerances based on Cpmk index. *Applied Mathematical Modelling* **35**, 4473–4489. MR2801968 https://doi.org/10.1016/j.apm.2011.03.011

MacDonald, P. D. M. (1971). Comment on an estimation procedure for mixtures of distributions by Choi and Bulgren. *Journal of the Royal Statistical Society, Series B* **33**, 326–329. MR0246400

Maiti, S. S. and Saha, M. (2012). Bayesian estimation of generalized process capability indices. *Journal of Probability and Statistics*. MR2885989 https://doi.org/10.1155/2012/819730

Maiti, S. S., Saha, M. and Nanda, A. K. (2010). On generalizing process capability indices. *Journal of Quality Technology and Quantitative Management* **7**, 279–300. MR3204873

Miao, R., Zhang, X., Yang, D., Zhao, Y. and Jiang, Z. (2011). A conjugate Bayesian approach for calculating process capability indices. *Expert Systems with Applications* **38**, 8099–8104.

Ouyang, L. Y., Wu, C. C. and Kuo, H. L. (2002). Bayesian assessment for some process capability indices. *International Journal of Information and Management Sciences* **13**, 1–8. MR1923493

Pearn, W. L., Kotz, S. and Johnson, N. L. (1992). Distributional and inferential properties of process capability indices. *Journal of Quality Technology* **24**, 216–231. MR1379982

Pearn, W. L., Lin, C. H. and Chen, H. S. (1998). Distributional and inferential properties of the process accuracy and process precision indices. *Communications in Statistics Theory and Methods* **27**, 985–1000. MR1613513 https://doi.org/10.1080/03610929808832139

Pearn, W. L., Tai, Y. T., Hsiao, I. F. and Ao, Y. P. (2014). Approximately unbiased estimator for non-normal process capability index C_{Npk}. *Journal of Testing and Evaluation* **42**, 1408–1417.

Pearn, W. L., Tai, Y. T. and Wang, H. T. (2016). Estimation of a modified capability index for non-normal distributions. *Journal of Testing and Evaluation* **44**, 1998–2009.

Peng, C. (2010a). Parametric lower confidence limits of quantile-based process capability indices. *Journal of Quality Technology and Quantitative Management* **7**, 199–214.

Peng, C. (2010b). Estimating and testing quantile-based process capability indices for processes with skewed distributions. *Journal of Data Science* **8**, 253–268.

Pina-Monarrez, M. R., Ortiz-Yañez, J. F. and Rodríguez-Borbón, M. I. (2016). Non-normal capability indices for the Weibull and log-normal distributions. *Quality and Reliability Engineering International* **32**, 1321–1329.

Rao, G. S., Aslam, M. and Kantam, R. R. L. (2016). Bootstrap confidence intervals of C_{Npk} for inverse Rayleigh and log-logistic distributions. *Journal of Statistical Computation and Simulation* **86**, 862–873. MR3437062 https://doi.org/10.1080/00949655.2015.1040799

Saha, M., Dey, S. and Maiti, S. S. (2018). Parametric and non-parametric bootstrap confidence intervals of C_{Npk} for exponential power distribution. *Journal Industrial and Production Engineering* **35**, 160–169.

Saxena, S. and Singh, H. P. (2006). A Bayesian estimator of process capability index. *Journal of Statistics & Management Systems* **9**, 269–283. MR2245138 https://doi.org/10.1080/09720510.2006.10701206

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of non-central distributions in computing intervals. *Educational and Psychological Measurement* **61**, 605–632. MR1842990 https://doi.org/10.1177/00131640121971392

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods* **9**, 164–182.

Swain, J., Venkatraman, S. and Wilson, J. (1988). Least squares estimation of distribution function in Johnsons translation system. *Journal of Statistical Computation and Simulation* **29**, 271–297.

Teimouri, M., Hoseini, S. M. and Nadarajah, S. (2013). Comparison of estimation methods for the Weibull distribution. *Statistics* **47**, 93–109. MR3023017 https://doi.org/10.1080/02331888.2011.559657

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher* **31**, 25–32.

Weber, S., Ressurreio, T. and Duarte, C. (2016). Yield prediction with a new generalized process capability index applicable to non-normal data. *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems* **35**, 931–942.

Wu, C. W. and Lin, T. Y. (2009). A Bayesian procedure for assessing process performance based on the third-generation capability index. *Journal of Applied Statistics* **6**, 1205–1223. MR2744392 https://doi.org/10.1080/02664760802582298

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 158–171
https://doi.org/10.1214/20-BJPS470

Sample size for estimating organism concentration in ballast water: A Bayesian approach

Eliardo G. Costa[a], **Carlos Daniel Paulino**[b] **and Julio M. Singer**[c]

[a]*Universidade Federal do Rio Grande do Norte*
[b]*CEAUL, Universidade de Lisboa*
[c]*Universidade de São Paulo*

Abstract. Estimation of microorganism concentration in ballast water tanks is important to evaluate and possibly to prevent the introduction of invasive species in stable ecosystems. For such purpose, the number of organisms in ballast water aliquots must be counted and used to estimate their concentration with some precision requirement. Poisson and negative binomial models have been employed to describe the organism distribution in the tank, but determination of sample sizes required to generate estimates with pre-specified precision is still not well established. A Bayesian approach is a flexible alternative to accommodate adequate models that account for the heterogeneous distribution of the organisms and may provide a sequential way of enhancing the estimation procedure by updating the prior distribution along the ballast water discharging process. We adopt such an approach to compute sample sizes required to construct credible intervals obtained via two optimality criteria that have not been employed in this context. Such intervals may be used in the decision with respect to compliance with the D-2 standard of the Ballast Water Management Convention. We also conduct a simulation study to verify whether the credible intervals obtained with the proposed sample sizes satisfy the precision criteria.

References

Adcock, C. J. (1987). A Bayesian approach to calculating sample sizes for multinomial sampling. *Journal of the Royal Statistical Society Series D The Statistician* **36**, 155–159.

Adcock, C. J. (1988). A Bayesian approach to calculating sample sizes. *Journal of the Royal Statistical Society Series D The Statistician* **37**, 433–439.

Bailey, S. A. (2015). An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. *Aquatic Ecosystem Health & Management* **18**, 261–268.

Basurko, O. C. and Mesbahi, E. (2011). Statistical representativeness of ballast water sampling. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 183–190.

Carney, K. J., Basurko, O. C., Pazouki, K., Marsham, S., Delany, J. E., Desai, D. V., Anil, A. C. and Mesbahi, E. (2013). Difficulties in obtaining representative samples for compliance with the ballast water management convention. *Marine Pollution Bulletin* **68**, 99–105.

Casas-Monroy, O., Rajakaruna, H. and Bailey, S. A. (2020). Improving estimation of phytoplankton abundance and distribution in ballast water discharges. *Journal of Applied Phycology*. https://doi.org/10.1007/s10811-019-02034-x

Costa, E. G. (2017). Tamanho amostral para estimar a concentração de organismos em água de lastro: uma abordagem bayesiana. Ph.D. thesis, Departamento de Estatística, Universidade de São Paulo, São. Paulo. In Portuguese.

Costa, E. G., Lopes, R. M. and Singer, J. M. (2015). Implications of heterogeneous distributions of organisms on ballast water sampling. *Marine Pollution Bulletin* **91**, 280–287.

Costa, E. G., Lopes, R. M. and Singer, J. M. (2016). Sample size for estimating the mean concentration of organisms in ballast water. *Journal of Environmental Management* **180**, 433–438.

Costa, E. G., Paulino, C. D. and Singer, J. M. (2020). Supplement to "Sample size for estimating organism concentration in ballast water: A Bayesian approach." https://doi.org/10.1214/20-BJPS470SUPP

Key words and phrases. Average coverage criterion, average length criterion, Poisson/gamma distribution, negative binomial/Pearson Type VI distribution.

De Santis, F. (2007). Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society Series A Statistics in Society* **170**, 95–113. MR2339100 https://doi.org/10.1111/j.1467-985X.2006.00438.x

First, M. R., Robbins-Wamsley, S. H., Riley, S. C., Moser, C. S., Smith, G. E., Tamburri, M. N. and Drake, L. A. (2013). Stratification of living organisms in ballast tanks: How do organism concentrations vary as ballast water is discharged? *Environmental Science & Technology* **47**, 4442–4448.

Frazier, M., Miller, A. W., LeeII, H. and Reusser, D. A. (2013). Counting at low concentrations: The statistical challenges of verifying ballast water discharge standards. *Ecological Applications* **23**, 339–351.

Gollasch, S. and David, M. (2017). Recommendations for representative ballast water sampling. *Journal of Sea Research* **123**, 1–15.

IMO (2004). International convention for the control and management of ship ballast water and sediments. http://www.imo.org/.

Inoue, L. Y., Berry, D. A. and Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *American Statistician* **59**, 79–87. MR2113200 https://doi.org/10.1198/000313005X21069

Jang, P. G., Kim, J. H. and Choi, K. H. (2019). Testing of organisms $\geq$ 10 μm and $\leq$ 50 μm for ballast water management systems. *Advances in Oceanography & Marine Biology* **1**.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994a). *Continuous Univariate Distributions, Vol. 1*, 2nd ed. New York: John Wiley & Sons. MR1299979

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994b). *Continuous Univariate Distributions, Vol. 2*, 2nd ed. New York: John Wiley & Sons. MR1326603

Joseph, L. and Bélisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society Series D The Statistician* **46**, 209–226.

Joseph, L., Berger, R. D. and Bélisle, P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine* **16**, 769–781.

Joseph, L. and Wolfson, D. B. (1997). Interval-based versus decision theoretic criteria for the choice of sample size. *Journal of the Royal Statistical Society Series D The Statistician* **46**, 145–149.

Joseph, L., Wolfson, D. B. and Berger, R. D. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Journal of the Royal Statistical Society Series D The Statistician* **44**, 143–154.

Ludwig, J. A. and Reynolds, J. F. (1988). *Statistical Ecology: A Primer on Methods and Computing, Vol. 1*, 1st ed. New York: John Wiley & Sons.

Marbuah, G., Gren, I.-M. and McKie, B. (2014). Economics of harmful invasive species: A review. *Diversity* **6**, 500–523.

Miller, A. W., Frazier, M., Smith, G. E., Perry, E. S., Ruiz, G. M. and Tamburri, M. N. (2011). Enumerating sparse organisms in ships' ballast water: Why counting to 10 is not easy. *Environmental Science & Technology* **45**, 3539–3546.

M'Lan, C. E., Joseph, L. and Wolfson, D. B. (2006). Bayesian sample size determination for case-control studies. *Journal of the American Statistical Association* **101**, 760–772. MR2256186 https://doi.org/10.1198/016214505000001023

M'Lan, C. E., Joseph, L. and Wolfson, D. B. (2008). Bayesian sample size determination for binomial proportions. *Bayesian Analysis* **3**, 269–296. MR2407427 https://doi.org/10.1214/08-BA310

Murphy, K. R., Ritz, D. and Hewitt, C. L. (2002). Heterogeneous zooplankton distribution in a ship's ballast tanks. *Journal of Plankton Research* **24**, 729–734.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rahme, E., Joseph, L. and Gyorkos, T. W. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Journal of the Royal Statistical Society Series C Applied Statistics* **49**, 119–128. MR1817878 https://doi.org/10.1111/1467-9876.00182

Rajakaruna, H., VandenByllaardt, J., Kydd, J. and Bailey, S. (2018). Modeling the distribution of colonial species to improve estimation of plankton concentration in ballast water. *Journal of Sea Research* **133**, 166–176.

Stamey, J. D., Young, D. M. and Bratcher, T. L. (2006). Bayesian sample-size determination for one and two Poisson rate parameters with applications to quality control. *Journal of Applied Statistics* **33**, 583–594. MR2240928 https://doi.org/10.1080/02664760600679643

Wang, F. and Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* **17**, 193–208. MR1925941 https://doi.org/10.1214/ss/1030550861

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 172–186
https://doi.org/10.1214/20-BJPS471

Yang and Prentice model with piecewise exponential baseline distribution for modeling lifetime data with crossing survival curves

Fábio N. Demarqui and Vinícius D. Mayrink

Universidade Federal de Minas Gerais

Abstract. Proportional hazards (PH), proportional odds (PO) and accelerated failure time (AFT) models have been widely used to deal with survival data in different fields of knowledge. Despite their popularity, such models are not suitable to handle survival data with crossing survival curves. Yang and Prentice (2005) proposed a semiparametric two-sample approach, denoted here as the YP model, allowing the analysis of crossing survival curves and including the PH and PO configurations as particular cases. In a general regression setting, the present work proposes a fully likelihood-based approach to fit the YP model. The main idea is to model the baseline hazard via the piecewise exponential (PE) distribution. The approach shares the flexibility of the semiparametric models and the tractability of the parametric representations. An extensive simulation study is developed to evaluate the performance of the proposed model. We demonstrate how useful is the new method through the analysis of survival times related to patients enrolled in a cancer clinical trial. Finally, an `R` package called `YPPE` was developed to fit the proposed model. The simulation results indicate that our model performs well for moderate sample sizes in the general regression setting. A superior performance is also observed with respect to the original YP model designed for the two-sample scenario.

References

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica* **4**, 505–524. MR1309427

Brent, R. P. (1973). *Algorithms for Minimization Without Derivatives*. Englewood Cliffs: Prentice-Hall. MR0339493

Breslow, N. (1972). Discussion on regression models and life-tables (by D. R. Cox). *Journal of the Royal Statistical Society, Series B* **34**, 216–217. MR0341758

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* **76**.

Clark, D. E. and Ryan, L. M. (2002). Concurrent prediction of hospital mortality and length of stay from risk factors on admission. *Health Services Research* **37**, 631–645.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **34**, 187–220. MR0341758

Demarqui, F. N., Dey, D. K., Loschi, R. H. and Colosimo, E. A. (2011). Modeling survival data using the piecewise exponential model with random time grid. In *Recent Advances in Biostatistics, Vol. 4* (M. Bhattacharjee, S. K. Dhar and S. Subramanian, eds.) 109–122. Singapore: World Scientific. https://doi.org/10.1142/9789814329804_0006

Demarqui, F. N., Dey, D. K., Loschi, R. H. and Colosimo, E. A. (2014). Fully semiparametric Bayesian approach for modeling survival data with cure fraction. *Biometrical Journal* **56**, 198–218.

Demarqui, F. N., Loschi, R. H., Dey, D. K. and Colosimo, E. A. (2012). A class of dynamic piecewise exponential models with random time grid. *Journal of Statistical Planning and Inference* **142**, 728–742. MR2853579 https://doi.org/10.1016/j.jspi.2011.09.006

Diao, G., Zeng, D. and Yang, S. (2013). Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics* **69**, 840–849. MR3146780 https://doi.org/10.1111/biom.12097

Key words and phrases. Survival analysis, short-term and long-term hazard ratios, semiparametric modeling, maximum likelihood estimation.

Egge, K. and Zahl, P. H. (1999). Survival of glaucoma patients. *Acta Ophthalmologica Scandinavica* **77**, 397–401.
Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society Series C Applied Statistics* **40**, 63–79.
Gastrointestinal Tumor Study Group (1982). A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma. *Cancer* **49**, 1771–1777.
Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer. MR1876598 https://doi.org/10.1007/978-1-4757-3447-8
Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B, Methodological* **40**, 214–221. MR0517442
Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267–278. MR0326939 https://doi.org/10.1093/biomet/60.2.267
Lee, S. H. (2011). Maximum of the weighted Kaplan–Meier tests for the two-sample censored data. *Journal of Statistical Computation and Simulation* **81**, 1017–1026. MR2820063 https://doi.org/10.1080/00949651003627753
Nieto-Barajas, L. E. (2014). Bayesian semiparametric analysis of short- and long-term hazard ratios with covariates. *Computational Statistics & Data Analysis* **71**, 477–490. MR3131984 https://doi.org/10.1016/j.csda.2013.03.012
Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*, 2nd ed. New York: Springer. MR2244940
Putter, H., Sasako, M., Hartgrink, H. H., van-de-Velde, C. J. H. and van-Houwelingen, J. C. (2005). Long-term survival with non-proportional hazards: Results from the Dutch gastric cancer trial. *Statistics in Medicine* **24**, 2807–2821. MR2201852 https://doi.org/10.1002/sim.2143
R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
Sahu, S. K., Dey, D. K., Aslanidou, H. and Sinha, D. (1997). A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis* **3**, 123–137.
Schneider, S., Demarqui, F. N., Colosimo, E. A. and Mayrink, V. D. (2020). An approach to model clustered survival data with dependent censoring. *Biometrical Journal* **62**, 157–174. MR2635103 https://doi.org/10.1002/bimj.201800391
Shyur, H. J., Elsayed, E. A. and Luxhoj, J. T. (1999). A general model for accelerated life testing with time-dependent covariates. *Naval Research Logistics* **49**, 303–321. MR1677541 https://doi.org/10.1002/(SICI)1520-6750(199904)46:3\<303::AID-NAV4\>3.3.CO;2-W
Sinha, D., Chen, M. H. and Ghosh, S. K. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics* **55**, 585–590. MR1705161 https://doi.org/10.1111/j.0006-341X.1999.00585.x
Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association* **92**, 1195–1212.
Tong, X., Zhu, C. and Sun, J. (2007). Semiparametric regression analysis of two-sample current status data, with applications to tumorigenicity experiments. *Canadian Journal of Statistics* **35**, 575–584. MR2416857 https://doi.org/10.1002/cjs.5550350408
Yang, S. (2018). Improving testing and description of treatment effect in clinical trials with survival outcomes. *Statistics in Medicine* **38**, 530–544. MR3902596 https://doi.org/10.1002/sim.7676
Yang, S. and Prentice, R. L. (2005). Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika* **92**, 1–17. MR2158606 https://doi.org/10.1093/biomet/92.1.1
Yang, S. and Prentice, R. L. (2010). Improved logrank-type tests for survival data using adaptive weights. *Biometrics* **66**, 30–38. MR2756688 https://doi.org/10.1111/j.1541-0420.2009.01243.x
Yang, S. and Prentice, R. L. (2011). Estimation of the 2-sample hazard ratio function using a semiparametric model. *Biostatistics* **12**, 354–368.
Yang, S. and Prentice, R. L. (2015). Assessing potentially time-dependent treatment effect from clinical trials and observational studies for survival data, with applications to the Women's Health Initiative combined hormone therapy trial. *Statistics in Medicine* **34**, 1801–1817. MR3334693 https://doi.org/10.1002/sim.6453
Yang, S. and Zhao, Y. (2012). Checking the short-term and long-term hazard ratio model for survival data. *Scandinavian Journal of Statistics* **39**, 554–567. MR2971638 https://doi.org/10.1111/j.1467-9469.2012.00804.x
Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B* **69**, 507–564. MR2370068 https://doi.org/10.1111/j.1369-7412.2007.00606.x
Zhang, H., Wang, P. and Sun, J. (2018). Regression analysis of interval-censored failure time data with possibly crossing hazards. *Statistics in Medicine* **37**, 768–775. MR3760447 https://doi.org/10.1002/sim.7538

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 187–204
https://doi.org/10.1214/20-BJPS468

The random deterioration rate model with measurement error based on the inverse Gaussian distribution

Lia H. M. Morita[a], Vera L. D. Tomazella[b], Pedro L. Ramos[c], Paulo H. Ferreira[c] and Francisco Louzada[c]

[a]*Federal University of Mato Grosso*
[b]*Federal University of São Carlos*
[c]*University of São Paulo*

Abstract. In this paper, we introduce the random deterioration rate model with measurement error in order to incorporate the variability among different components. The motivation behind the random variable model is to capture the randomness in the individual differences across the population. This model incorporates only sample uncertainty of the degradation, and no temporal variability is included. The measurement error models appear to overcome this problem. The random rate analysis is based on repeated measurements of failure sizes generated by a degradation process over time in a components population. Some characteristics of the random deterioration rate model based on the inverse Gaussian distribution and subject to measurement error, are examined. We carry out simulation studies to (i) assess the performance of the maximum likelihood estimates obtained through the Gaussian quadrature along with Quasi-Newton optimization method; and (ii) examine the effects of model misspecification on the model selection criteria's performance, as well as on the lifetime prediction's accuracy and precision. The potentiality of the proposed model is illustrated through two real data sets.

References

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association* **49**, 765–769. MR0069459

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications* **6**, 76–90. MR0433870

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. *New York, NY*. MR1919620

Carrasco, J. M. F., Ferrari, S. L. P. and Arellano-Valle, R. B. (2014). Errors-in-variables beta regression models. *Journal of Applied Statistics* **41**, 1530–1547. MR3292031 https://doi.org/10.1080/02664763.2014.881784

Crowther, M. J., Andersson, T. M.-L., Lambert, P. C., Abrams, K. R. and Humphreys, K. (2016). Joint modelling of longitudinal and survival data: Incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine* **35**, 1193–1209. MR3476536 https://doi.org/10.1002/sim.6779

Doornik, J. A. (2009). An object-oriented matrix programming language ox 6.

EPRI (2006). Steam generator integrity assessment guidelines: REVISION 2. Technical Report TR1012987.

Fenyvesi, L., Lu, H. and Jack, T. R. (2004). Prediction of corrosion defect growth on operating pipelines. In *2004 International Pipeline Conference*, 225–230. American Society of Mechanical Engineers.

Freitas, M. A., Luíza, M., de Toledo, G., Colosimo, E. A. and Pires, M. C. (2009). Using degradation data to assess reliability: A case study on train wheel degradation. *Quality and Reliability Engineering International* **25**, 607–629.

Freitas, M. A., Colosimo, E. A., dos Santos, T. R. and Pires, M. C. (2010). Reliability assessment using degradation models: Bayesian and classical approaches. *Pesquisa Operacional* **30**, 194–219.

Huyse, L. and Roodselaar, V. (2010). Effects of inline inspection sizing uncertainties on the accuracy of the largest features and corrosion rate statistics. In *2010 8th International Pipeline Conference (IPC2010). Calgary, Alberta, Canada.*

Key words and phrases. Degradation analysis, inverse Gaussian distribution, LASER data, locomotive wheels data, random deterioration rate.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association* **84**, 717–726. MR1132587

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: An example. *Journal of the Royal Statistical Society Series C Applied Statistics* **50**, 325–335. MR1856328 https://doi.org/10.1111/1467-9876.00237

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* **27**, 3105–3124. MR2522152 https://doi.org/10.1002/sim.3153

Lu, C. J. and Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics* **35**, 161–174. MR1225093 https://doi.org/10.2307/1269661

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York, NY: John Wiley & Sons. Chapter 13.

Meeker, W. Q., Escobar, L. A. and Lu, C. J. (1998). Accelerated degradation tests: Modeling and analysis. *Technometrics* **40**, 89–99.

Morita, L. H. M., Tomazella, V. L. D., Ramos, P. L., Ferreira, P. H. and Louzada, F. (2020). Supplement to "The random deterioration rate model with measurement error based on the inverse Gaussian distribution." https://doi.org/10.1214/20-BJPS468SUPP

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M. and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* **15**, 39–57. MR2269362 https://doi.org/10.1198/106186006X96854

Pandey, M. D. and Lu, D. (2013). Estimation of parameters of degradation growth rate distribution from noisy measurement data. *Structural Safety* **43**, 60–69.

Peng, C. Y. and Tseng, S. T. (2009). Mis-specification analysis of linear degradation models. *IEEE Transactions on Reliability* **58**, 444–455.

Peng, W., Li, Y.-F., Yang, Y.-J., Huang, H.-Z. and Zuo, M. J. (2014). Inverse Gaussian process models for degradation analysis: A Bayesian perspective. *Reliability Engineering & Systems Safety* **130**, 175–189.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W. and Kahaner, D. K. (2012). *Quadpack: A Subroutine Package for Automatic Integration, Vol. 1*. Berlin: Springer. MR0712135 https://doi.org/10.1007/978-3-642-61786-7

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

R Core Team R (2018). *A Language and Environment for Statistical Computing; 2015*.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111–163.

Rondon, L. M. and Bolfarine, H. (2017). Bayesian analysis of flexible measurement error models. *Brazilian Journal of Probability and Statistics* **31**, 618–639. MR3693983 https://doi.org/10.1214/16-BJPS326

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics, Vol. 1*, 1st ed. Netherlands: Springer. MR0876486

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

Seshadri, V. (2012). *The Inverse Gaussian Distribution: Statistical Theory and Applications, Vol. 137*. Berlin: Springer. MR1622488 https://doi.org/10.1007/978-1-4612-1456-4

van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & Systems Safety* **94**, 2–21.

Villaseñor, J. A. and Gonzalez-Estrada, E. (2015). Tests of fit for inverse Gaussian distributions. *Statistics & Probability Letters* **105**, 189–194. MR3371998 https://doi.org/10.1016/j.spl.2015.06.017

Wang, X. and Xu, D. (2010). An inverse Gaussian process model for degradation data. *Technometrics* **52**, 188–197. MR2676425 https://doi.org/10.1198/TECH.2009.08197

Xiao, N.-C., Huang, H.-Z., Wang, Z., Liu, Y. and Zhang, X.-L. (2012). Unified uncertainty analysis by the mean value first order saddlepoint approximation. *Structural and Multidisciplinary Optimization* **46**, 803–812.

AIMS AND SCOPE

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

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

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

GENERAL INFORMATION

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

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

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

Correspondence. Mail concerning membership, subscriptions, nonreceipt claims, copyright permissions, advertising or back issues should be sent to the IMS Dues and Subscription Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998. Mail concerning submissions or editorial content should be sent to the Editor at nancy@ime.unicamp.br. Mail concerning the production of this journal should be sent to: Geri Mattson at bjps@mattsonpublishing.com.

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

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

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

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