

# Contents

## **A. AGRESTI**

The foundations of statistical science: A history of textbook presentations ..... 657

## **B. FAVA and H. F. LOPES**

The illusion of the illusion of sparsity: An exercise in prior sensitivity ..... 699

## **M. TOUNSI and M. ZITOUNI**

Matrix-variate Lindley distributions and its applications ..... 721

## **A. R. S. SILVA, C. L. N. AZEVEDO, J. L. BAZÁN and J. S. NOBRE**

Bayesian inference for zero-and/or-one augmented beta rectangular regression models ..... 749

## **Z. LIU, C. TU and Y. PAN**

Model-assisted SCAD calibration for non-probability samples ..... 772

## **D. I. GALLARDO and M. BOURGUIGNON**

A note on the Nielsen distribution ..... 788

## **R. MAMIN and L. MYTNIK**

Absolute continuity of the super-Brownian motion with infinite mean ..... 791

## **C. LATTANZI and M. LEONELLI**

A change-point approach for the identification of financial extreme regimes... 811

## **S. NAQVI, N. MISRA and P. S. CHAN**

Some preservation properties of shifted stochastic orders ..... 838

## **V. S. O. RIBEIRO, J. S. NOBRE, J. R. S. DOS SANTOS and C. L. N. AZEVEDO**

Beta rectangular regression models to longitudinal data ..... 851

## **E. R. PINTO and L. A. PEREIRA**

On variable selection in joint modeling of mean and dispersion ..... 875





# Brazilian Journal of Probability and Statistics

---

Volume 35 • Number 4 • November 2021

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 35, Number 4, November 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, PO Box 729, Middletown, Maryland 21769, USA.

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

Copyright © 2021 by the Brazilian Statistical Association.

Printed in the United States of America



Partial financial support:  
CNPq and CAPES (Brazil).



## The foundations of statistical science: A history of textbook presentations

Alan Agresti

*Department of Statistics, University of Florida, Gainesville, Florida 32611, USA, [agresti@ufl.edu](mailto:agresti@ufl.edu)*

**Abstract.** We discuss how the foundations of statistical science have been presented historically in textbooks, with focus on the first half of the twentieth century after the field had become better defined by advances due to Francis Galton, Karl Pearson, and R. A. Fisher. Our main emphasis is on books that presented the theory underlying the subject, often identified as *mathematical statistics*, with primary focus on books authored by G. Udny Yule, Maurice Kendall, Samuel Wilks, and Harald Cramér. We also discuss influential books on statistical methods by R. A. Fisher and George Snedecor that showed scientists how to implement the theory. We then survey textbooks published in the quarter century after World War 2, as Statistics gathered more visibility as an academic subject and Departments of Statistics were formed at many universities. We also summarize how Bayesian presentations of Statistics emerged. In each section, we describe how the books were evaluated in reviews shortly after their publications. We conclude by briefly discussing the recent past, the present, and the future of textbooks on the foundations of statistical science and include comments about this by several notable statisticians.

### References

- Agresti, A. and Kateri, M. (2022). *Foundations of Statistics for Data Scientists, with R and Python*. Boca Raton: CRC Press.
- Agresti, A. and Meng, X.-L. (2013). *Strength in Numbers: The Rising of Academic Statistics Departments in the U.S.* New York: Springer. MR3077328 <https://doi.org/10.1007/978-1-4614-3649-2>
- Aitken, A. C. (1939). *Statistical Mathematics*. Edinburgh: Oliver and Boyd. MR0001493
- Aldrich, J. (2005). The statistical education of Harold Jeffreys. *International Statistical Review* **73**, 289–307.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. MR0091588
- Bartlett, M. S. (1940). The present position of mathematical statistics (with discussion). *Journal of the Royal Statistical Society* **103**, 1–29.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley. MR1274699 <https://doi.org/10.1002/9780470316870>
- Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day. MR0443141
- Bickel, P. J. and Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume II*. London: Chapman and Hall / CRC. MR3287337
- Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis*. Cambridge: MIT Press. MR0381130
- Blackwell, D. (1969). *Basic Statistics*. New York: McGraw-Hill. MR1535965 <https://doi.org/10.2307/2316759>
- Blalock, H. M. (1960). *Social Statistics*. New York: McGraw-Hill.
- Blom, G. (1987). Harald Cramér 1893–1985. *The Annals of Statistics* **15**, 1335–1350. MR0913560 <https://doi.org/10.1214/aos/1176350596>
- Bowley, A. L. (1901). *Elements of Statistics*. London: P. S. King & Son.
- Box, J. F. (1978). *R. A. Fisher: The Life of a Scientist*. New York: Wiley. MR0500579
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading: Addison-Wesley. MR0418321

---

*Key words and phrases.* History of statistics, mathematical statistics, statistical theory, G. Udny Yule, R. A. Fisher, Maurice Kendall, Samuel Wilks, Harald Cramér, George Snedecor.

- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science* **16**, 199–231. [MR1874152](https://doi.org/10.1214/ss/1009213726)  
<https://doi.org/10.1214/ss/1009213726>
- Brownlee, K. A. (1960). *Statistical Theory and Methodology in Science and Engineering*. New York: Wiley. [MR0119268](#)
- Brunk, H. D. (1960). *An Introduction to Mathematical Statistics*. Boston: Ginn and Company. [MR0114257](#)
- Burgess, R. W. (1927). *The Mathematics of Statistics*. Boston: Houghton Mifflin.
- Casella, G. and Berger, R. (1990). *Statistical Inference*. Wadsworth & Brooks/Cole. [MR1051420](#)
- Cobb, G. W. (2011). Teaching statistics: Some important tensions. *Chilean Journal of Statistics* **2**, 31–62. [http://soche.cl/chjs/volumes/02/01/Cobb\(2011\).pdf](http://soche.cl/chjs/volumes/02/01/Cobb(2011).pdf). [MR2820017](#)
- Cochran, W. G. (1953). *Sampling Techniques*. New York: Wiley. [MR0054199](#)
- Cochran, W. G. and Cox, G. (1950). *Experimental Designs*. New York: Wiley. [MR0034558](#)
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall. [MR0370837](#)
- Cramér, H. (1937). *Random Variables and Probability Distributions*. *Cambridge Tract in Mathematics* **36**. Cambridge. [MR0165599](#)
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press. [MR0016588](#)
- de Finetti, B. (1974, 1975). *Theory of Probability*. New York: Wiley. Two volumes translated by A. Machi and A. F. M. Smith. [MR0440641](#)
- DeGroot, M. (1975). *Probability and Statistics*. Reading, MA: Addison Wesley. [MR0373075](#)
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics* **26**, 745–766. [MR3765335](https://doi.org/10.1080/10618600.2017.1384734) <https://doi.org/10.1080/10618600.2017.1384734>
- Draper, N. R. and Smith, H. (1967). *Applied Regression Analysis*. New York: Wiley. [MR0211537](#)
- Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press. [MR3523956](https://doi.org/10.1017/CBO9781316576533) <https://doi.org/10.1017/CBO9781316576533>
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd. (14th ed. 1970.) [MR0346954](#)
- Fisher, R. A. (1990). *Statistical Methods, Experimental Design, and Scientific Inference*. London: Oxford University Press. [MR0346955](#)
- Fisz, M. (1963). *Probability Theory and Mathematical Statistics*. New York: Wiley. [MR0164358](#)
- Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. New York: Wiley. [MR0083868](#)
- Freund, J. E. (1962). *Mathematical Statistics*. New York: Prentice-Hall. [MR0166853](#)
- Fry, T. C. (1928). *Probability and Its Engineering Uses*. Princeton: D. Van Nostrand Company, Inc. [MR0181027](#)
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall. [MR1385925](#)
- Good, I. J. (1950). *Probability and the Weighing of Evidence*. New York: Hafner. [MR0041366](#)
- Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge, MA: MIT Press. [MR0185724](#)
- Hardin, J., Horton, N. J., Nolan, D. and Temple Lang, D. (2021). Computing in the statistics curricula: A 10-year retrospective. *Journal of Statistics and Data Science Education* **29**(sup1), S4–S6. <https://doi.org/10.1080/10691898.2020.1862609>
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer. [MR2722294](https://doi.org/10.1007/978-0-387-84858-7) <https://doi.org/10.1007/978-0-387-84858-7>
- Hill, A. B. (1937). *Principles of Medical Statistics*. London: Lancet.
- Hoel, P. G. (1947). *Introduction to Mathematical Statistics*. New York: Wiley.
- Hogg, R. V. and Craig, A. T. (1959). *Introduction to Mathematical Statistics*. New York: Macmillan. [MR0137186](#)
- Horton, N. J. (2013). I hear, I forget. I do, I understand: A modified Moore-method mathematical statistics course. *American Statistician* **67**, 219–228. [MR3303815](https://doi.org/10.1080/00031305.2013.849207) <https://doi.org/10.1080/00031305.2013.849207>
- Hotelling, H. (1940). The teaching of statistics. *The Annals of Mathematical Statistics* **11**, 457–470. [MR0120695](#)
- Hotelling, H. (1951). The impact of R. A. Fisher on statistics. *Journal of the American Statistical Association* **46**, 35–46. [MR0359184](#)
- Jeffreys, H. (1939). *Theory of Probability*. Oxford: Clarendon Press. [MR0000924](#)
- Kempthorne, O. (1974). *International Statistical Review* **42**, 319–323. Snedecor, George W., 1881–1974. [MR0386955](#)
- Kendall, M. G. (1943). *The Advanced Theory of Statistics, Vol. 1*. London: Griffin. [MR0019869](#)
- Kendall, M. G. (1946). *The Advanced Theory of Statistics, Vol. 2*. London: Griffin. [MR0019869](#)
- Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics. Volume 1: Distribution Theory*. London: Griffin. [MR0019869](#)
- Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley. [MR0107933](#)
- Lehmann, E. L. (2008). *Reminiscences of a Statistician*. Berlin: Springer. [MR2367933](https://doi.org/10.1007/978-0-387-71597-1) <https://doi.org/10.1007/978-0-387-71597-1>

- Lindgren, B. W. (1962). *Statistical Theory*. New York: Macmillan. [MR0143273](#)
- Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint; Volume 1 Probability, Volume 2 Inference*. Cambridge: Cambridge University Press. [MR0168083](#)
- Magnello, M. E. (2009). Karl Pearson and the establishment of mathematical statistics. *International Statistical Review* **77**, 3–29. [MR0541179](#)
- McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. London: Chapman & Hall. [MR0727836](#)  
<https://doi.org/10.1007/978-1-4899-3244-0>
- Mendenhall, W. and Scheaffer, R. L. (1973). *Mathematical Statistics with Applications*. N. Scituate: Duxbury Press. [MR0353508](#)
- Mood, A. M. (1950). *An Introduction to the Theory of Statistics*. New York: McGraw-Hill. [MR0033470](#)
- Mood, A. M. (1990). Miscellaneous reminiscences. *Statistical Science* **5**, 38–43. [MR1054856](#)
- Mosteller, F. (1964). Samuel S. Wilks: Statesman of statistics. *American Statistician* **18**, 11–17. [MR0191011](#)
- Neyman, J. (1949). *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press. [MR1529649](#) <https://doi.org/10.2307/2310188>
- Noether, G. E. (1989). The teaching of statistics in the USA: The early years. In *Sesquicentennial Invited Paper Sessions, Proceedings of the American Statistical Association*, 279–285. Alexandria, VA: American Statistical Association.
- Nolan, D. and Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications*. <https://www.stat.berkeley.edu/users/statlabs>.
- Pearson, K. (1914). *Tables for Statisticians and Biometricians*. Cambridge: Cambridge University Press.
- Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard Business School Division of Research. (Later published by MIT Press and in 2000 in Wiley Classics Library.) [MR1789469](#)
- Randles, R. H. (2007). A conversation with Robert V. Hogg. *Statistical Science* **22**, 137–152. [MR2408664](#)  
<https://doi.org/10.1214/088342306000000637>
- Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: Wiley. [MR0050824](#)
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: Wiley. [MR0221616](#)
- Rice, J. A. (1988). *Mathematical Statistics and Data Analysis*. Brooks/Cole Pub. Co.
- Rietz, H. L., ed. (1923). On the subject matter of a course in mathematical statistics. *The American Mathematical Monthly* **30**, 155–166. [MR1520205](#) <https://doi.org/10.2307/2298448>
- Rietz, H. L., ed. (1924). *Handbook of Mathematical Statistics*. Boston: Houghton Mifflin.
- Rietz, H. L. (1927). *Mathematical Statistics*. Mathematical Association of America.
- Robert, C. P., Chopin, N. and Rousseau, J. (2009). Harold Jeffreys's theory of probability revisited. *Statistical Science* **24**, 141–172. [MR2655841](#) <https://doi.org/10.1214/09-STS284>
- Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. New York: Wiley. [MR0407916](#)
- Rossmann, A. and Chance, B. (2003). Notes from JSM 2003 panel “Is the math stat course obsolete?”. <https://nhorton.people.amherst.edu/MathStatObsolete.pdf>.
- Samuelson, P. A. (1950). Recent texts in mathematical statistics. *Review of Economics and Statistics* **32**, 139–143.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Dover. [MR0063582](#)
- Savage, L. J. (1961). The foundations of statistics reconsidered. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability; Volume 1: Contributions to the Theory of Statistics* (J. Neyman, ed.) 575–586. Berkeley: University of California. [MR0133898](#)
- Scheffé, H. (1967). *The Analysis of Variance*. New York: Wiley. [MR0116429](#)
- Schlaifer, R. (1959). *Probability and Statistics for Business Decisions*. New York: McGraw-Hill.
- Snedecor, G. W. (1937). *Statistical Methods, Subtitle Applied to Experiments in Agriculture and Biology*. Ames, Iowa: Collegiate Press.
- Statistics at a Crossroads (2019). [https://www.nsf.gov/mps/dms/documents/Statistics\\_at\\_a\\_Crossroads\\_Workshop\\_Report\\_2019.pdf](https://www.nsf.gov/mps/dms/documents/Statistics_at_a_Crossroads_Workshop_Report_2019.pdf).
- Stigler, S. M. (1978). Mathematical statistics in the early states. *The Annals of Statistics* **6**, 239–265. [MR0483118](#)
- Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press. [MR0852410](#)
- Stigler, S. M. (1999). *Statistics on the Table*. Cambridge: Harvard University Press. [MR1712969](#)
- Stuart, A. (1984). Sir Maurice Kendall, 1907–1983. *Journal of the Royal Statistical Society Series A* **147**, 120–122. [MR0748794](#)
- Tukey, J. (1977). *Exploratory Data Analysis*. Reading: Addison Wesley.
- Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. Cambridge: Cambridge University Press. [MR3967104](#) <https://doi.org/10.1017/9781108627771>
- Wald, A. (1942). *On the Principles of Statistical Inference. Notre Dame Mathematical Lectures* **1**. Notre Dame: University of Notre Dame Press. [MR0006681](#)

- Wald, A. (1947). *Sequential Analysis*. New York: Wiley. [MR0020764](#)
- Wald, A. (1949). *Statistical Decision Functions*. New York: Wiley. [MR0036976](#)
- Wasserman, L. (2003). *All of Statistics: A Concise Course in Statistical Inference*. Berlin: Springer. [MR2055670](#)  
<https://doi.org/10.1007/978-0-387-21736-9>
- Weatherburn, C. E. (1946). *A First Course in Mathematical Statistics*. Cambridge: Cambridge University Press. [MR0019262](#)
- West, C. J. (1918). *Introduction to Mathematical Statistics*. Columbus, Ohio: R. G. Adams and Company.
- Wilks, S. S. (1937). *The Theory of Statistical Inference*. Princeton: Princeton University.
- Wilks, S. S. (1947). *Mathematical Statistics*. Princeton: Princeton University Press. [MR0008657](#)
- Wilks, S. S. (1948). *Elementary Statistical Analysis*. Princeton: Princeton University Press.
- Wilks, S. S. (1962). *Mathematical Statistics*. New York: Wiley. [MR0144404](#)
- Wishart, J. (1939). Some aspects of the teaching of statistics (with discussion). *Journal of the Royal Statistical Society* **102**, 532–564.
- Wittstein, T. (1872). On mathematical statistics and its application to political economy and insurance. *Journal of the Institute of Actuaries and Assurance Magazine* **17**, 178–189. Translated by T. B. Sprague.
- Wolfenden, H. H. (1942). *The Fundamental Principles of Mathematical Statistics*. New York: Macmillan. [MR0006636](#)
- Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. London: Charles Griffin and Co. [MR1527705](#)  
<https://doi.org/10.2307/2308128>
- Yates, F. (1952). George Udny Yule, 1871–1951. *Biographical Memoirs of Fellows of the Royal Society* **8**, 309–323. [MR0051789](#)
- Yule, G. U. (1911). *An Introduction to the Theory of Statistics*. London: Charles Griffin and Co.
- Yule, G. U. and Kendall, M. G. (1950). *An Introduction to the Theory of Statistics*, 14th ed. London: Charles Griffin and Co. [MR0035938](#)



## The illusion of the illusion of sparsity: An exercise in prior sensitivity

Bruno Fava<sup>1</sup> and Hedibert F. Lopes<sup>2</sup>

<sup>1</sup>Northwestern University, USA

<sup>2</sup>Insper, Brazil, [hedibertfl@insper.edu.br](mailto:hedibertfl@insper.edu.br)

**Abstract.** The emergence of Big Data raises the question of how to model economic relations when there is a large number of possible explanatory variables. We revisit the issue by comparing the possibility of using dense or sparse models in a Bayesian approach, allowing for variable selection and shrinkage. More specifically, we discuss the results reached by Giannone, Lenza and Primiceri (2020) through a “Spike-and-Slab” prior, which suggest an “illusion of sparsity” in economic data, as no clear patterns of sparsity could be detected. We make a further revision of the posterior distributions of the model, and propose three experiments to evaluate the robustness of the adopted prior distribution. We find that the pattern of sparsity is sensitive to the prior distribution of the regression coefficients, and present evidence that the model indirectly induces variable selection and shrinkage, which suggests that the “illusion of sparsity” could be, itself, an illusion. Code is available on Github.\*

### References

- Barro, R. J. and Lee, J.-W. (1994). Sources of economic growth. In *Carnegie-Rochester Conference Series on Public Policy*, Vol. 40, 1–46.
- Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80**, 2369–2429. MR3001131 <https://doi.org/10.3982/ECTA9626>
- Belloni, A., Chernozhukov, V. and Hansen, C. (2011). Inference for high-dimensional sparse econometric models. Preprint. Available at [arXiv:1201.0220](https://arxiv.org/abs/1201.0220). MR2868201 [https://doi.org/10.1007/978-3-642-19989-9\\_3](https://doi.org/10.1007/978-3-642-19989-9_3)
- Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81**, 608–650. MR3207983 <https://doi.org/10.1093/restud/rdt044>
- Bhattacharya, A., Pati, S., Pillai, N. and Dunson, D. (2015). Dirichlet-Laplace priors for optimal shrinkage. *Journal of the American Statistical Association* **110**, 1479–1490. MR3449048 <https://doi.org/10.1080/01621459.2014.960967>
- Chen, D. L. and Yeh, S. (2012). Growth under the shadow of expropriation? The economic impacts of eminent domain. *Manuscript, Toulouse Inst. Advanced Study*.
- Donohue III, J. J. and Levitt, S. D. (2001). The impact of legalized abortion on crime. *The Quarterly Journal of Economics* **116**, 379–420.
- Freyberger, J., Neuhierl, A. and Weber, M. (2017). Dissecting characteristics nonparametrically. Technical report, National Bureau of Economic Research.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.
- Giannone, D., Lenza, M. and Primiceri, G. (2021). Economic Predictions with Big Data: The Illusion of Sparsity. *Econometrica* **89**, 2409–2437. <https://doi.org/10.3982/ECTA17842>
- Griffin, J. and Brown, P. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* **5**, 171–188. MR2596440 <https://doi.org/10.1214/10-BA507>
- Hahn, R., He, J. and Lopes, H. F. (2019). Efficient sampling for Gaussian linear regression with arbitrary priors. *Journal of Computational and Graphical Statistics* **28**, 142–154. MR3939378 <https://doi.org/10.1080/10618600.2018.1482762>

---

*Key words and phrases.* Sparsity, model selection, high dimensional data, shrinkage, Bayesian econometrics.

\*[github.com/bfava/IllusionOfIllusion](https://github.com/bfava/IllusionOfIllusion)

- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.
- Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics*, 730–773. MR2163158 <https://doi.org/10.1214/009053604000001147>
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**, 1023–1032. MR0997578 <https://doi.org/10.1214/009053604000001147>
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103**, 681–686.
- Rocková, V. and George, E. (2018). The Spike-and-Slab LASSO. *Journal of the American Statistical Association* **113**, 431–444. MR3803476 <https://doi.org/10.1080/01621459.2016.1260469>
- Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**, 1167–1179. MR1951271 <https://doi.org/10.1198/016214502388618960>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)* **58**, 267–288. MR1379242 <https://doi.org/10.1214/009053604000001147>
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* **21**, 1455–1508.

## Matrix-variate Lindley distributions and its applications

Mariem Tounsi<sup>1</sup> and Mouna Zitouni<sup>2</sup>

<sup>1</sup>*Computer Engineering and Applied Mathematics Department, Sfax National Engineering School, B.P. 1173, 3038, Tunisia, [tounsi.mariemfss@hotmail.fr](mailto:tounsi.mariemfss@hotmail.fr)*

<sup>2</sup>*Applied Mathematics Department, Sfax Faculty of Sciences, B.P. 802, 3038, Tunisia, [zitounimounaa@yahoo.fr](mailto:zitounimounaa@yahoo.fr)*

**Abstract.** Restricting on the fact that the definition of multivariate analogs of the real gamma distribution is replaced by the Wishart distribution on symmetric matrices, and based on the notion of mixture models which is a flexible and powerful tool for treating data taken from multiple subpopulations, we set forward a multivariate analog of the real Lindley distributions of the first and second kinds on the modern framework of symmetric cones which can be used to model waiting and survival times matrix data. Within this framework, we first construct a new probability distributions, named the matrix-variate Lindley distributions. Some fundamental properties of these new distributions are established. Their statistical properties including moments, the coefficient of variation, skewness and the kurtosis are discussed. We then create an iterative hybrid Expectation-Maximization Fisher-Scoring (EM-FS) algorithm to estimate the parameters of the new class of probability distributions. Through simulation as well as comparative studies with respect to the Wishart distribution, the effectiveness and reliability of the proposed distributions are proved. Finally, the usefulness and the applicability of the new models are elaborated and illustrated by means of two real data sets from biological sciences and medical image segmentation which is one of the most important and popular tasks in medical image analysis.

## References

- Aparicio, L., Bordyuh, M., Blumberg, A. J. and Rabadan, R. (2020). A random matrix theory approach to denoise single-cell data. *Patterns* **1** 100035. <https://doi.org/10.1016/j.patter.2020.100035>
- Arashi, M., Bekker, A. and van Niekerk, J. (2017). Weighted-type Wishart distributions with application. *REVSTAT Statistical Journal* **15**, 205–222. [MR3641252](https://doi.org/10.1016/j.patter.2020.100035)
- Bahri, Y., Kadmon, J., Pennington, J., Schoenholz, S. S., Sohl-Dickstein, J. and Ganguli, S. (2020). Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics* **11**, 501–528. <https://doi.org/10.1146/annurev-conmatphys-031119-050745>
- Bai, Z., Chen, Y. and Liang, Y.-C. (2009). *Random Matrix Theory and Its Applications: Multivariate Statistics and Wireless Communications, Vol. 18*. World Scientific. [MR2603191](https://doi.org/10.1142/9789814273121) <https://doi.org/10.1142/9789814273121>
- Bashaw, M., Bloomsmith, M., Maple, T. and Bercovitch, F. (2007). The structure of social relationships among captive female giraffe (*Giraffa camelopardalis*). *Journal of Comparative Psychology* **121**, 46–53.
- Casalis, M., Letac, G., et al (1996). The Lukacs–Olkin–Rubin characterization of Wishart distributions on symmetric cones. *The Annals of Statistics* **24**, 763–786. [MR1394987](https://doi.org/10.1214/aos/1032894464) <https://doi.org/10.1214/aos/1032894464>
- Chen, C.-F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society, Series B, Methodological* **41**, 235–248. [MR0547249](https://doi.org/10.1093/bjoms/41.2.235)
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, 1–38. [MR0501537](https://doi.org/10.1093/bjoms/41.2.235)
- Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika* **1**–26. [MR4092109](https://doi.org/10.1007/s11336-020-09697-3) <https://doi.org/10.1007/s11336-020-09697-3>
- Friedrich, O. and Eifler, T. (2018). Precision matrix expansion—efficient use of numerical simulations in estimating errors on cosmological parameters. *Monthly Notices of the Royal Astronomical Society* **473**, 4150–4163.
- Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation* **78**, 493–506. [MR2424558](https://doi.org/10.1016/j.matcom.2007.06.007) <https://doi.org/10.1016/j.matcom.2007.06.007>

- Gupta, A. and Nagar, D. (1999). *Matrix Variate Distributions, Vol. 104*. CRC Press. [MR1738933](#)
- Haff, L. R., Kim, P. T., Koo, J.-Y., Richards, D. S. P., et al (2011). Minimax estimation for mixtures of Wishart distributions. *The Annals of Statistics* **39**, 3417–3440. [MR3012414](#) <https://doi.org/10.1214/11-AOS951>
- Hassan, A., Wani, S. A., Shafi, S. and Ahmad, S. B. (2020). Lindley-quasi Xgamma: Properties and applications. *Pakistan Journal of Statistics Series A* **36**, 73–89. [MR4193492](#) <https://doi.org/10.2478/jamsi-2020-0009>
- Haykin, S., Principe, J., Sejnowski, T. and McWhirter, J. G. (2007). *New Directions in Statistical Signal Processing: From Systems to Brains*. MIT Press. [MR2590576](#)
- Hidot, S. and Saint-Jean, C. (2010). An expectation-maximization algorithm for the Wishart mixture model: Application to movement clustering. *Pattern Recognition Letters* **31**, 2318–2324. <https://doi.org/10.1016/j.patrec.2010.07.002>
- Hougaard, P. (2012). *Analysis of Multivariate Survival Data*. Springer. [MR1777022](#) <https://doi.org/10.1007/978-1-4612-1304-8>
- Jewell, N. P., Kimber, A. C., Lee, M.-L. T. and Whitmore, G. A. (1996). *Lifetime Data: Models in Reliability and Survival Analysis*. Springer. [MR2608285](#) <https://doi.org/10.1007/s10985-009-9138-0>
- Krishnamoorthy, K. and Mathew, T. (1999). Comparison of approximation methods for computing tolerance factors for a multivariate normal population. *Technometrics* **41**, 234–249. [MR1704363](#) <https://doi.org/10.2307/1270568>
- Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data, Vol. 362*. John Wiley & Sons. [MR0640866](#)
- Letac, G. and Massam, H. (2001). The normal quasi-Wishart distribution. *Contemporary Mathematics* **287**, 231–240. [MR1873678](#) <https://doi.org/10.1090/conm/287/04788>
- Letac, G. and Massam, H. (2004). All invariant moments of the Wishart distribution. *Scandinavian Journal of Statistics* **31**, 295–318. [MR2066255](#) <https://doi.org/10.1111/j.1467-9469.2004.01-043.x>
- Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series B, Methodological*, 102–107. [MR0095550](#)
- Mazucheli, J. and Achcar, J. A. (2011). The Lindley distribution applied to competing risks lifetime data. *Computer Methods and Programs in Biomedicine* **104**, 188–192.
- McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons. [MR1789474](#) <https://doi.org/10.1002/0471721182>
- McNicholas, P. D., Murphy, T. B., McDaid, A. F. and Frost, D. (2010). Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models. *Computational Statistics & Data Analysis* **54**, 711–723. [MR2744427](#) <https://doi.org/10.1016/j.csda.2009.02.011>
- Miettinen, J., Nordhausen, K. and Taskinen, S. (2017). Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. *Journal of Statistical Software* **76**, 1–31. <https://doi.org/10.18637/jss.v076.i02>
- Mühlbacher, A. and Guhr, T. (2018). Credit risk meets random matrices: Coping with non-stationary asset correlations. *Risks* **6**, 1–25.
- Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory, Vol. 197*. John Wiley & Sons. [MR0652932](#)
- Nordhausen, K. and Oja, H. (2018). Independent component analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics* **10**, e1440. [MR3850449](#) <https://doi.org/10.1002/wics.1440>
- Purba, S., Sutarman and Darnius, O. (2018). Maximum likelihood based on Newton Raphson, Fisher scoring and expectation maximization algorithm application on accident data. *International Journal of Advanced Research* **6**, 965–969. <https://doi.org/10.21474/ijar01/6306>
- Rajaratnam, B., Massam, H., Carvalho, C. M., et al (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of Statistics* **36**, 2818–2849. [MR2485014](#) <https://doi.org/10.1214/08-AOS619>
- Riecke, T. V., Sedinger, B. S., Williams, P. J., Leach, A. G. and Sedinger, J. S. (2019). Estimating correlations among demographic parameters in population models. *Ecology and Evolution* **9**, 13521–13531.
- Ripatti, S., Larsen, K. and Palmgren, J. (2002). Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Analysis* **8**, 349–360. [MR1951608](#) <https://doi.org/10.1023/A:1020566821163>
- Sankaran, M. (1970). 275. Note: The discrete Poisson–Lindley distribution. *Biometrics* 145–149.
- Shanker, R. (2013). A new quasi Lindley distribution. *International Journal of Statistics and Systems* **8**, 143–156.
- Shanker, R. and Hagos, F. (2015). On Poisson–Lindley distribution and its applications to biological sciences. *Biometrics & Biostatistics International Journal* **2**, 1–5.
- Shanker, R. and Mishra, A. (2013a). A quasi Lindley distribution. *African Journal of Mathematics and Computer Science Research* **6**, 64–71.
- Shanker, R. and Mishra, A. (2013b). A two-parameter Lindley distribution. *Statistics in Transition New Series* **1**, 45–56.
- Shanker, R. and Mishra, A. (2014). A two-parameter Poisson–Lindley distribution. *International Journal of Statistics and Systems* **9**, 79–85.

- Shanker, R. and Sharma, S. (2016). On two parameter Lindley distribution and its applications to model lifetime data. *Biometrics & Biostatistics International Journal* **3**, 1–8.
- Shanker, R., Sharma, S. and Shanker, R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. *Applied Mathematics* **4**, 363–368.
- Wang, W., Xiang, D., Ban, Y., Zhang, J. and Wan, J. (2019). Superpixel-based segmentation of polarimetric SAR images through two-stage merging. *Remote Sensing* **11**, 402.
- Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 32–52.
- Wong, C. S. and Wang, T. (1995). Laplace–Wishart distributions and Cochran theorems. *Sankhyā: The Indian Journal of Statistics, Series A*, 342–359. [MR1390503](#)
- Xue, X. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis* **2**, 277–289.

## Bayesian inference for zero-and/or-one augmented beta rectangular regression models

Ana R. S. Silva<sup>1</sup>, Caio L. N. Azevedo<sup>1,\*</sup>, Jorge L. Bazán<sup>2</sup> and Juvêncio S. Nobre<sup>3</sup>

<sup>1</sup>State University of Campinas, Campinas, Brazil, \*[cnaber@ime.unicamp.br](mailto:cnaber@ime.unicamp.br)

<sup>2</sup>University of São Paulo, São Paulo, Brazil

<sup>3</sup>Federal University of Ceará, Fortaleza, Brazil

**Abstract.** In this paper, we developed a full set of Bayesian inference tools, for zero-and/or-one augmented beta rectangular regression models to analyze limited-augmented data, under a new parameterization. This parameterization: facilitates the development of both regression models and inferential tools as well as make simplifies the respective computational implementations. The proposed Bayesian tools were parameter estimation, model fit assessment, model comparison (information criteria), residual analysis and case influence diagnostics, developed through MCMC algorithms. In addition, we adapted available methods of posterior predictive checking, using appropriate discrepancy measures. We conducted several simulation studies, considering some situations of practical interest, aiming to evaluate: prior sensitivity choice, parameter recovery of the proposed model and estimation method, the impact of transforming the observed zeros and ones, along with the use of non-augmented models, and the behavior of the proposed model fit assessment and model comparison tools. A psychometric real data set was analyzed to illustrate the performance of the developed tools, illustrating the advantages of the developed analysis framework.

## References

- Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* **94**, 443–458. MR2380571 <https://doi.org/10.1093/biomet/asm017>
- Atkinson, A. C. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford: Clarendon Press.
- Baghishani, H. and Mohammadzadeh, M. (2012). Asymptotic normality of posterior distributions for generalized linear mixed models. *Journal of Multivariate Analysis* **111**, 66–77. MR2944406 <https://doi.org/10.1016/j.jmva.2012.05.003>
- Bandyopadhyay, D., Galvis, D. M. and Lachos, V. H. (2017). Augmented mixed models for clustered proportion data. *Statistical Methods in Medical Research* **26**, 880–897. MR3635942 <https://doi.org/10.1177/0962280214561093>
- Bayarri, M. J. and Berger, J. O. (2000). P-values for composite null models. *Journal of the American Statistical Association* **95**, 1127–1142. MR1804239 <https://doi.org/10.2307/2669749>
- Bayes, C. L., Bazán, J. L. and García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis* **7**, 841–866. MR3000016 <https://doi.org/10.1214/12-BA728>
- Branscum, A. J., Johnson, W. O. and Thurmond, M. C. (2007). Bayesian beta regression: Application to household data and genetic distance between foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics* **49**, 287–301. MR2405396 <https://doi.org/10.1111/j.1467-842X.2007.00481.x>
- Buckley, J. (2003). Estimation of models with beta-distributed dependent variables: A replication and extension of Paolino's study. *Political Analysis* **11**, 204–205.
- Carlstrom, L., Woodward, J. and Palmer, C. (2000). Evaluating the simplified conjoint expected risk model: Comparing the use of objective and subjective information. *Risk Analysis* **20**, 385–392.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 1.

- Cho, H., Ibrahim, J. G., Sinha, D. and Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics* **65**, 116–124. MR2665851 <https://doi.org/10.1111/j.1541-0420.2008.01037.x>
- 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
- Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software* **34**, 1–24.
- Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 236–244.
- Ferrari, S. L. P. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31**, 799–815. MR2095753 <https://doi.org/10.1080/0266476042000214501>
- Figueroa-Zúñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. P. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis* **61**, 137–147. MR3063006 <https://doi.org/10.1016/j.csda.2012.12.002>
- Galvis, D. M., Bandyopadhyay, D. and Lachos, V. H. (2014). Augmented mixed beta regression models for periodontal proportion data. *Statistics in Medicine* **33**, 3759–3771. MR3260658 <https://doi.org/10.1002/sim.6179>
- Gamerman, D. and Lopes, H. (2006). *Stochastic Simulation for Bayesian Inference*, 2nd ed. New York, NY: Chapman & Hall/CRC.
- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. DTIC Document. MR1380275
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd ed. London: Taylor & Francis. MR2027492
- Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733–760. MR1422404
- Gonçalves, F. B., Gamerman, D. and Soares, T. M. (2013). Simultaneous multifactor DIF analysis and detection in item response theory. *Computational Statistics & Data Analysis* **59**, 144–160. MR3000048 <https://doi.org/10.1016/j.csda.2012.10.011>
- Hartigan, J. A. (2011). *Bayes Theory*. Berlin: Springer. MR0715782 <https://doi.org/10.1007/978-1-4613-8242-3>
- Homan, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623. MR3214779
- 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>
- Ma, Z. and Leijon, A. (2011). Bayesian estimation of beta mixture models with variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 2160–2173.
- Nogarroto, D. C., Azevedo, C. L. N. and Bazán, J. L. (2020). Bayesian modeling and prior sensitivity analysis for zero-one augmented beta regression models with an application to psychometric data. *Brazilian Journal of Probability and Statistics* **34**, 304–322. MR4093261 <https://doi.org/10.1214/18-BJPS423>
- Ospina, R. (2008). Inflated Beta Regression Model. Doctoral's thesis, IME-USP (in Portuguese).
- Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers* **51**, 111–126. MR2556590 <https://doi.org/10.1007/s00362-008-0125-4>
- Ospina, R. and Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis* **56**, 1609–1623. MR2892364 <https://doi.org/10.1016/j.csda.2011.10.005>
- Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis* **9**, 325–346.
- R Development Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* **12**, 1151–1172. MR0760681 <https://doi.org/10.1214/aos/1176346785>
- Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **71**, 319–392. MR2649602 <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Silva, A. R. S., Azevedo, C. L. N., Bazán, J. L. and Nobre, J. S. (2021). Supplement to “Bayesian inference for zero-and/or-one augmented beta rectangular regression models.” <https://doi.org/10.1214/21-BJPS505SUPP>
- Silva, A. R. S., Azevedo, C. L. N., Nobre, J. S. and Bazán, J. L. (2021). Augmented-limited regression models with an application to the study of the risk perceived using continuous scales. *Journal of Applied Statistics* **48**, 1998–2021. MR4296776 <https://doi.org/10.1080/02664763.2020.1783518>
- Silvapulle, M. J. and Sen, P. K. (2004). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. New York: Wiley. MR2099529
- Sinharay, S., Johnson, M. S. and Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement* **30**, 298–321. MR2227470 <https://doi.org/10.1177/0146621605285517>

- Sinharay, S. and Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference* **111**, 209–221. MR1955882 [https://doi.org/10.1016/S0378-3758\(02\)00303-8](https://doi.org/10.1016/S0378-3758(02)00303-8)
- Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* **11**, 54.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **64**, 583–639. MR1979380 <https://doi.org/10.1111/1467-9868.00353>
- Wang, J. and Luo, S. (2016). Augmented Beta rectangular regression models: A Bayesian perspective. *Biometrical Journal* **58**, 206–221. MR3457175 <https://doi.org/10.1002/bimj.201400232>
- Wang, J. and Luo, S. (2017). Bayesian multivariate augmented Beta rectangular regression models for patient-reported outcomes and survival data. *Statistical Methods in Medical Research* **26**, 1684–1699. MR3687172 <https://doi.org/10.1177/0962280215586010>



## Model-assisted SCAD calibration for non-probability samples

Zhan Liu, Chaofeng Tu and Yingli Pan\*

Hubei Key Laboratory of Applied Mathematics, School of Mathematics and Statistics, Hubei University, Wuhan 430062, China, \*[panyingli220@163.com](mailto:panyingli220@163.com)

**Abstract.** Increasing costs and non-response rates of probability samples have provoked the extensive use of non-probability samples. However, non-probability samples are subject to selection bias, resulting in difficulty for inference. Calibration is a popular method to reduce selection bias in non-probability samples. When rich covariate information is available, a key problem is how to select covariates and estimate parameters in calibration for non-probability samples. In this paper, the model-assisted SCAD calibration is proposed to make population inference from non-probability samples. A parametric model between the study variable and covariates is first established. SCAD is then used to estimate the model parameters based on non-probability samples. The modified forward Kullback–Leibler distance is lastly explored to conduct calibration for non-probability samples based on the estimated parametric model. The theoretical properties of the model-assisted SCAD calibration estimator are further derived. Results from simulation studies show that the model-assisted SCAD calibration estimator yields the smallest bias and mean square error compared with other estimators. Also, a real data from the 2017 Netizen Social Awareness Survey (NSAS) is used to demonstrate the proposed methodology.

## References

- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. and Tourangeau, R. (2013). Summary report of the AAPOR task force on nonprobability sampling. *Journal of Survey Statistics and Methodology* **1**, 90–143.
- Bethlehem, J. (2016). Solving the nonresponse problem with sample matching? *Social Science Computer Review* **34**, 59–77.
- Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95**, 555–571. MR2443175 <https://doi.org/10.1093/biomet/asn022>
- Chen, J. K. T., Valliant, R. and Elliott, M. R. (2018). Model-assisted calibration of non-probability sample survey data using adaptive LASSO. *Survey Methodology* **44**, 117–144.
- Chen, Y., Li, P. and Wu, C. (2020). Doubly robust inference with non-probability survey samples. *Journal of the American Statistical Association* **115**, 2011–2021. MR4189773 <https://doi.org/10.1080/01621459.2019.1677241>
- Dever, J. A. and Valliant, R. (2016). General regression estimation adjusted for undercoverage and estimated control totals. *Journal of Survey Statistics and Methodology* **4**, 289–318.
- Deville, J. C. (2000). Generalized calibration and application to weighting for non-response. In *Compstat—Proceedings in Computational Statistics: 14th Symposium Held in Utrecht, The Netherlands*, 65–76. New York: Springer.
- Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382. MR1173804
- Disogra, C., Curtiss, C., Elisa, C. and Dennis, J. (2011). Calibrating non-probability Internet samples with probability samples using early adopter characteristics. In *Joint Statistical Meetings (JSM) Proceedings (Survey Research Methods Section)*, 4501–4515. Alexandria, VA: American Statistical Association.
- Elliott, M. R. and Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science* **32**, 249–264. MR3648958 <https://doi.org/10.1214/16-STS598>

- Fan, J. Q. and Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360. MR1946581 <https://doi.org/10.1198/016214501753382273>
- Greenacre, Z. A. (2016). The importance of selection bias in Internet surveys. *Open Journal of Statistics* **6**, 397–404.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685. MR0053460
- Keiding, N. and Louis, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *Journal of the Royal Statistical Society: Series A* **179**, 319–376. MR3461587 <https://doi.org/10.1111/rssa.12136>
- Kott, P. S. (2014). Calibration weighting when model and calibration variables can differ. In *Contributions to Sampling Statistics: ITACOSM 2013 Selected Papers* (F. Mecatti, P. Conti and M. Ranalli, eds.) 1–18. Heidelberg, Germany: Springer.
- Kott, P. S. and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association* **105**, 1265–1275. MR2752620 <https://doi.org/10.1198/jasa.2010.tm09016>
- Lee, S. and Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research* **37**, 319–343. MR2649463 <https://doi.org/10.1177/0049124108329643>
- Mehranian, A., Rad, H. S., Ay, M. R., Rahmim, A., Ay, M. R. and Zaidi, H. (2013). Smoothly clipped absolute deviation (SCAD) regularization for compressed sensing MRI using an augmented Lagrangian scheme. *Magnetic Resonance Imaging* **31**, 1399–1411.
- Montanari, G. E. and Ranalli, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association* **100**, 1429–1442. MR2236453 <https://doi.org/10.1198/016214505000000141>
- Rivers, D. (2007). Sampling for web surveys. In *Joint Statistical Meetings. Salt Lake City, UT*.
- Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* **33**, 99–119.
- Singh, S. (2004). Golden and silver jubilee year-2003 of the linear regression estimators. In *Proceedings of the ASA Section on Survey Research Methods*, 4382–4389. Toronto. MR2230416 <https://doi.org/10.1007/s00184-005-0002-2>
- Singh, S. (2006). Survey statisticians celebrate Golden Jubilee Year-2003 of the linear regression estimator. *Metrika* **63**, 1–18. MR2230416 <https://doi.org/10.1007/s00184-005-0002-2>
- Singh, S. (2012). Calibration of design weights using adis placement function. *Metrika* **75**, 85–107. MR2878110 <https://doi.org/10.1007/s00184-010-0316-6>
- Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97**, 661–682. MR2672490 <https://doi.org/10.1093/biomet/asq035>
- Tan, Z. (2013). Simple design-efficient calibration estimators for rejective and high-entropy sampling. *Biometrika* **100**, 399–415. MR3068442 <https://doi.org/10.1093/biomet/ass090>
- Tan, Z. and Wu, C. (2015). Generalized pseudo empirical likelihood inferences for complex surveys. *Canadian Journal of Statistics* **43**, 1–17. MR3324425 <https://doi.org/10.1002/cjs.11237>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288. MR1379242
- Tourangeau, R., Conrad, F. G. and Couper, M. P. (2013). *The Science of Web Surveys*. New York: Oxford University Press.
- Valliant, R. and Dever, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research* **40**, 105–137. MR2758301 <https://doi.org/10.1177/0049124110392533>
- Wang, L., Wang, S. J. and Wang, G. N. (2014). Variable selection and estimation for longitudinal survey data. *Journal of Multivariate Analysis* **130**, 409–424. MR3229546 <https://doi.org/10.1016/j.jmva.2014.05.006>
- Wang, W., Rothschild, D., Goel, S. and Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting* **31**, 980–991.
- Wu, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika* **90**, 937–951. MR2024768 <https://doi.org/10.1093/biomet/90.4.937>
- Wu, C. and Lu, W. W. (2016). Calibration weighting methods for complex surveys. *International Statistical Review* **84**, 79–98. MR3491280 <https://doi.org/10.1111/insr.12097>
- Wu, C. and Sitter, R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association* **96**, 185–193. MR1952731 <https://doi.org/10.1198/016214501750333054>
- Yang, S., Kim, J. K. and Song, R. (2019). Doubly robust inference when combining probability and non-probability samples with high-dimensional data. arXiv preprint. Available at [arXiv:1903.05212v1](https://arxiv.org/abs/1903.05212v1). MR4084171

## A note on the Nielsen distribution

Diego I. Gallardo<sup>1</sup> and Marcelo Bourguignon<sup>2</sup>

<sup>1</sup>*Departamento de Matemática, Universidad de Atacama, Copiapó, Chile, [diego.gallardo@uda.cl](mailto:diego.gallardo@uda.cl)*

<sup>2</sup>*Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, Brazil, [m.p.bourguignon@gmail.com](mailto:m.p.bourguignon@gmail.com)*

**Abstract.** Castellares, Lemonte, and Santos [Brazilian Journal of Probability and Statistics **34**(1) (2020) 90–111] introduced a two-parameter discrete Nielsen distribution, derived its properties, and illustrated the advantages of the model in three data applications. In this note, we will present a corrected version for some results for the particular case  $\theta = 1$ .

## References

- Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters* **15**, 35–42. MR1649319 [https://doi.org/10.1016/S0167-7152\(98\)00012-1](https://doi.org/10.1016/S0167-7152(98)00012-1)
- Castellares, F., Lemonte, A. J. and Santos, M. (2020). On the Nielsen distribution. *Brazilian Journal of Probability and Statistics* **34**, 90–111. MR4058972 <https://doi.org/10.1214/18-BJPS414>
- Chen, M.-H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* **94**, 909–919. MR1723307 <https://doi.org/10.2307/2670006>
- Fisher, R., Corbet, A. and Williams, C. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**, 42–58.

# Absolute continuity of the super-Brownian motion with infinite mean

Rustam Mamin<sup>1</sup> and Leonid Mytnik<sup>2</sup>

<sup>1</sup>Faculty of Mathematics, Technion—Israel Institute of Technology, Haifa 32000, Israel, [rst@technion.ac.il](mailto:rst@technion.ac.il)

<sup>2</sup>Faculty of Industrial Engineering & Management, Technion—Israel Institute of Technology, Haifa 32000, Israel, [leonid@ie.technion.ac.il](mailto:leonid@ie.technion.ac.il)

**Abstract.** In this work, we prove that for any dimension  $d \geq 1$  and any  $\gamma \in (0, 1)$  super-Brownian motion corresponding to the log-Laplace equation

$$v(t, x) = (S_t f)(x) - \int_0^t (S_{t-s} v^\gamma(s, \cdot))(x) ds, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

is absolutely continuous with respect to Lebesgue measure at any fixed time  $t > 0$ .  $\{S_t\}_{t \geq 0}$  denotes a transition semigroup of a standard Brownian motion. Our proof is based on properties of solutions of the log-Laplace equation. We also prove that when initial datum  $v(0, \cdot)$  is a finite, non-zero measure, then the log-Laplace equation has a unique, continuous solution. Moreover this solution continuously depends on initial data.

## References

- Aguirre, J. and Escobedo, M. (1986/87). A Cauchy problem for  $u_t - \Delta u = u^p$  with  $0 < p < 1$ . Asymptotic behaviour of solutions. *Annales de la Faculté des Sciences de L'Université de Toulouse, Mathématique (5)* **8**, 175–203. [MR0928843](#)
- Dawson, D. A. (1992). Infinitely divisible random measures and superprocesses. In *Stochastic Analysis and Related Topics (Silivri, 1990)*. *Progr. Probab.* **31**, 1–129. Boston, MA: Birkhäuser Boston. [MR1203373](#)
- Dawson, D. A. (1993). Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*. *Lecture Notes in Math.* **1541**, 1–260. Berlin: Springer. [MR1242575](#) <https://doi.org/10.1007/BFb0084190>
- Ekeland, I. and Témam, R. (1999). *Convex Analysis and Variational Problems*, English ed. *Classics in Applied Mathematics* **28**. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Translated from the French. [MR1727362](#) <https://doi.org/10.1137/1.9781611971088>
- Fleischmann, K. (1988). Critical behavior of some measure-valued processes. *Mathematische Nachrichten* **135**, 131–147. [MR0944225](#) <https://doi.org/10.1002/mana.19881350114>
- Fleischmann, K., Mytnik, L. and Wachtel, V. (2010). Optimal local Hölder index for density states of superprocesses with  $(1 + \beta)$ -branching mechanism. *Annals of Probability* **38**, 1180–1220. [MR2674997](#) <https://doi.org/10.1214/09-AOP501>
- Fleischmann, K., Mytnik, L. and Wachtel, V. (2011). Hölder index at a given point for density states of super- $\alpha$ -stable motion of index  $1 + \beta$ . *Journal of Theoretical Probability* **24**, 66–92. [MR2782711](#) <https://doi.org/10.1007/s10959-010-0334-3>
- Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer-Verlag. [MR1876169](#) <https://doi.org/10.1007/978-1-4757-4015-8>
- Konno, N. and Shiga, T. (1988). Stochastic partial differential equations for some measure-valued diffusions. *Probability Theory and Related Fields* **79**, 201–225. [MR0958288](#) <https://doi.org/10.1007/BF00320919>
- Lang, S. (1997). *Undergraduate Analysis*, 2nd ed. *Undergraduate Texts in Mathematics*. New York: Springer-Verlag. [MR1476913](#) <https://doi.org/10.1007/978-1-4757-2698-5>
- Li, Z. (2011). *Measure-Valued Branching Markov Processes*. *Probability and Its Applications (New York)*. Heidelberg: Springer. [MR2760602](#) <https://doi.org/10.1007/978-3-642-15004-3>
- Mytnik, L. and Perkins, E. (2003). Regularity and irregularity of  $(1 + \beta)$ -stable super-Brownian motion. *Annals of Probability* **31**, 1413–1440. [MR1989438](#) <https://doi.org/10.1214/aop/1055425785>
- Mytnik, L. and Wachtel, V. (2015). Multifractal analysis of superprocesses with stable branching in dimension one. *Annals of Probability* **43**, 2763–2809. [MR3395474](#) <https://doi.org/10.1214/14-AOP951>

- Mytnik, L. and Wachtel, V. (2016). *Regularity and Irregularity of Superprocesses with  $(1 + \beta)$ -Stable Branching Mechanism*. *SpringerBriefs in Probability and Mathematical Statistics*. Cham: Springer. MR3616202 <https://doi.org/10.1007/978-3-319-50085-0>
- Reimers, M. (1989). One-dimensional stochastic partial differential equations and the branching measure diffusion. *Probability Theory and Related Fields* **81**, 319–340. MR0983088 <https://doi.org/10.1007/BF00340057>
- Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill Book Co. MR0924157 <https://doi.org/10.1007/BFb0084190>
- Watanabe, S. (1968). A limit theorem of branching processes and continuous state branching processes. *Journal of Mathematics of Kyoto University* **8**, 141–167. MR0237008 <https://doi.org/10.1215/kjm/1250524180>

# A change-point approach for the identification of financial extreme regimes

Chiara Lattanzi<sup>1</sup> and Manuele Leonelli<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italy, [chiara.lattanzi3@studio.unibo.it](mailto:chiara.lattanzi3@studio.unibo.it)

<sup>2</sup>School of Human Sciences and Technology, IE University, Madrid, Spain, [manuele.leonelli@ie.edu](mailto:manuele.leonelli@ie.edu)

**Abstract.** Inference over tails is usually performed by fitting an appropriate limiting distribution over observations that exceed a fixed threshold. However, the choice of such threshold is critical and can affect the inferential results. Extreme value mixture models have been defined to estimate the threshold using the full dataset and to give accurate tail estimates. Such models assume that the tail behavior is constant for all observations. However, the extreme behavior of financial returns often changes considerably in time and such changes occur by sudden shocks of the market. Here the extreme value mixture model class is extended to formally take into account distributional extreme change-points, by allowing for the presence of regime-dependent parameters modelling the tail of the distribution. This extension formally uses the full dataset to both estimate the thresholds and the extreme change-point locations, giving uncertainty measures for both quantities. Estimation of functions of interest in extreme value analyses is performed via MCMC algorithms. Our approach is evaluated through a series of simulations, applied to real data sets and assessed against competing approaches. Evidence demonstrates that the inclusion of different extreme regimes outperforms both static and dynamic competing approaches in financial applications.

## References

- Albert, J. H. and Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* **11**, 1–15.
- Behrens, C. N., Lopes, H. F. and Gamberman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* **4**, 227–244. MR2062102 <https://doi.org/10.1191/1471082X04st075oa>
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**, 542–547.
- Caldara, D., Fuentes-Albero, C., Gilchrist, S. and Zakrajšek, E. (2016). The macroeconomic impact of financial and uncertainty shocks. *European Economic Review* **88**, 185–207.
- Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society Series C* **41**, 389–405. MR1278223
- Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes* **12**, 53–76. MR2480723 <https://doi.org/10.1007/s10687-008-0068-0>
- 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>
- Catania, L. and Bernardi, M. (2017). MCS: Model confidence set procedure. R package version 0.1.3. Available at <https://CRAN.R-project.org/package=MCS>.
- Chang, C., Jiménez-Martín, J., McAleer, M. and Pérez-Amaral, T. (2011). Risk management of risk under the basel accord: Forecasting value-at-risk of VIX futures. *Managerial Finance* **37**, 1088–1106.
- 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](https://doi.org/10.1016/S0304-4076(97)00115-2)
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B* **52**, 393–442. MR1086795

- de Carvalho, M., Pereira, S., Pereira, P. and de Zea Bermudez, Z. (2020). An extreme value Bayesian lasso for the conditional bulk and tail. Available at [arXiv:201007164](https://arxiv.org/abs/201007164).
- Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *EnvironMetrics* **21**, 661–686. [MR2838438 https://doi.org/10.1002/env.1041](https://doi.org/10.1002/env.1041)
- DuMouchel, W. H. (1983). Estimating the stable index  $\alpha$  in order to measure tail thickness: A critique. *The Annals of Statistics* **11**, 1019–1031. [MR0865342 https://doi.org/10.1214/aos/1176346318](https://doi.org/10.1214/aos/1176346318)
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall. [MR2260716](https://doi.org/10.1002/9781118131163)
- Gilli, M. and K ellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics* **27**, 207–228.
- Gonz alez-Rivera, G., Lee, T. and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* **20**, 629–645.
- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**, 27–62.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732. [MR1380810 https://doi.org/10.1093/biomet/82.4.711](https://doi.org/10.1093/biomet/82.4.711)
- Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set. *Econometrica* **79**, 453–497. [MR2809377 https://doi.org/10.3982/ECTA5771](https://doi.org/10.3982/ECTA5771)
- Horv ath, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test* **23**, 219–255. [MR3210268 https://doi.org/10.1007/s11749-014-0368-4](https://doi.org/10.1007/s11749-014-0368-4)
- Huerta, G. and Sans o, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**, 285–299. [MR2405331 https://doi.org/10.1007/s10651-007-0014-3](https://doi.org/10.1007/s10651-007-0014-3)
- Jaru skov a, D. and Rencov a, 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](https://doi.org/10.1002/env.865)
- Killick, R. and Eckley, I. (2014). `changeoint`: An R package for changepoint analysis. *Journal of Statistical Software* **58**, 1–19.
- Ko, S. I. M., Chong, T. T. L. 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](https://doi.org/10.1214/14-BA910)
- Leonelli, M. and Gamerman, D. (2020). Semiparametric bivariate modelling with flexible extremal dependence. *Statistics and Computing* **30**, 221–236. [MR4064619 https://doi.org/10.1007/s11222-019-09878-w](https://doi.org/10.1007/s11222-019-09878-w)
- Lima, S. R., Nascimento, F. F. and Ferraz, V. R. S. (2018). Regression models for time-varying extremes. *Journal of Statistical Computation and Simulation* **88**, 235–249. [MR3740727 https://doi.org/10.1080/00949655.2017.1385788](https://doi.org/10.1080/00949655.2017.1385788)
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis* **55**, 2137–2157. [MR2785120 https://doi.org/10.1016/j.csda.2011.01.005](https://doi.org/10.1016/j.csda.2011.01.005)
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* **7**, 271–300.
- Nascimento, F. F., Gamerman, D. and Davis, R. (2016). A Bayesian semi-parametric approach to extreme regime identification. *Brazilian Journal of Probability and Statistics* **30**, 540–561. [MR3582389 https://doi.org/10.1214/15-BJPS293](https://doi.org/10.1214/15-BJPS293)
- 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](https://doi.org/10.1007/s10651-010-0148-6)
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* **22**, 661–675. [MR2865043 https://doi.org/10.1007/s11222-011-9270-z](https://doi.org/10.1007/s11222-011-9270-z)
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2016). Time-varying extreme pattern with dynamic models. *Test* **25**, 131–149. [MR3463806 https://doi.org/10.1007/s11749-015-0444-4](https://doi.org/10.1007/s11749-015-0444-4)
- Nascimento, F. F. and Moura e Silva, W. V. (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](https://doi.org/10.1080/02664763.2016.1254733)
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131. [MR0423667](https://doi.org/10.1214/aos/1176346318)
- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367. [MR2749836 https://doi.org/10.1198/jcgs.2009.06134](https://doi.org/10.1198/jcgs.2009.06134)
- Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys* **28**, 82–108.
- Sampid, M. G., Hasim, H. M. and Dai, H. (2018). Refining value-at-risk estimates using a Bayesian Markov-switching GJR-GARCH copula-EVT model. *PLoS ONE* **13**, e0198753.
- Samuel, Y. M. Z. (2008). Value at risk and conditional extreme value theory via Markov regime switching models. *Journal of Futures Markets: Futures, Options, and Other Derivative Products* **28**, 155–181.

- Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal* **10**, 33–60. [MR2912370](#)
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. [MR0468014](#)
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639. [MR1979380](#) <https://doi.org/10.1111/1467-9868.00353>
- Stephens, D. A. (1994). Bayesian retrospective multiple-changepoint identification. *Journal of the Royal Statistical Society Series C* **43**, 159–178.
- Tancredi, A., Anderson, C. and O’Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**, 87–106. [MR2329799](#) <https://doi.org/10.1007/s10687-006-0009-8>
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594. [MR2756194](#)
- Wiper, M., Rious Insua, D. and Ruggeri, F. (2001). Mixtures of Gamma distributions with applications. *Journal of Computational and Graphical Statistics* **10**, 440–454. [MR1939034](#) <https://doi.org/10.1198/106186001317115054>
- Zhao, X., Scarrott, C. J., Oxley, L. and Reale, M. (2011). GARCH dependence in extreme value models with Bayesian inference. *Mathematics and Computers in Simulation* **81**, 1430–1440. [MR2781673](#) <https://doi.org/10.1016/j.matcom.2010.08.002>



## Some preservation properties of shifted stochastic orders

Sameen Naqvi<sup>1</sup>, Neeraj Misra<sup>2</sup> and P. S. Chan<sup>3</sup>

<sup>1</sup>Indian Institute of Technology Hyderabad, Kandi, India, [sameen@math.iith.ac.in](mailto:sameen@math.iith.ac.in)

<sup>2</sup>Indian Institute of Technology Kanpur, Kanpur, India, [neeraj@iitk.ac.in](mailto:neeraj@iitk.ac.in)

<sup>3</sup>The Chinese University of Hong Kong, New Territories, Hong Kong, [benchan@cuhk.edu.hk](mailto:benchan@cuhk.edu.hk)

**Abstract.** In this paper, we provide results on preservation of various shifted stochastic orders under the operations of mixtures. These results have been established with respect to up (down) shifted hazard rate order, up (down) shifted reversed hazard rate order and up (down) shifted likelihood ratio order. In addition, a discussion on preservation of shifted stochastic orders under the operations of convolutions is also presented. Some examples have been obtained to illustrate applications of these results.

## References

- Aboukalam, F. and Kayid, M. (2007). Some new results about shifted hazard and shifted likelihood ratio orders. In *International Mathematical Forum*, Vol. 31, 1525–1536. MR2333566 <https://doi.org/10.12988/imf.2007.07138>
- Artin, E. (1931). *Einführung in die Theorie der Gamma-funktion*. Leipzig: Teubner.
- Belzunce, F., Martínez-Riquelme, C. and Muler, J. (2016). *An Introduction to Stochastic Orders*. Amsterdam: Elsevier. MR3430823 <https://doi.org/10.1016/B978-0-12-803768-3.00001-6>
- Belzunce, F., Ruiz, J. M. and Ruiz, M. C. (2002). On preservation of some shifted and proportional orders by systems. *Statistics & Probability Letters* **60**, 141–154. MR1945437 [https://doi.org/10.1016/S0167-7152\(02\)00302-4](https://doi.org/10.1016/S0167-7152(02)00302-4)
- Brown, M. and Shanthikumar, J. G. (1998). Comparing the variability of random variables and point processes. *Probability in the Engineering and Informational Sciences* **12**, 425–444. MR1643827 <https://doi.org/10.1017/S0269964800005301>
- Di Crescenzo, A. and Longobardi, M. (2001). The up reversed hazard rate stochastic order. *Scientiae Mathematicae Japonicae* **54**, 575–581. MR1874182
- Hu, T., Nanda, A. K., Xie, H. and Zhu, Z. (2004). Properties of some stochastic orders: A unified study. *Naval Research Logistics* **51**, 193–216. MR2036498 <https://doi.org/10.1002/nav.10111>
- Hu, T. and Zhu, Z. (2001). An analytic proof of the preservation of the up-shifted likelihood ratio order under convolutions. *Stochastic Processes and Their Applications* **95**, 55–61. MR1847091 [https://doi.org/10.1016/S0304-4149\(01\)00089-8](https://doi.org/10.1016/S0304-4149(01)00089-8)
- Karlin, S. (1968). *Total Positivity (Vol. 1)*. Stanford: Stanford University Press. MR0230102
- Lai, C. and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. Berlin: Springer. MR2223811
- Li, H. and Li, X. (2013). Stochastic orders in reliability and risk. In *Honor of Professor Moshe Shaked*. Berlin: Springer. MR3157470 <https://doi.org/10.1007/978-1-4614-6892-9>
- Lillo, R. E., Nanda, A. K. and Shaked, M. (2000). Some shifted stochastic orders. In *Recent Advances in Reliability Theory*, 85–103. Boston, MA: Birkhauser. MR1783476
- Lillo, R. E., Nanda, A. K. and Shaked, M. (2001). Preservation of some likelihood ratio stochastic orders by order statistics. *Statistics & Probability Letters* **51**, 111–119. MR1808033 [https://doi.org/10.1016/S0167-7152\(00\)00137-1](https://doi.org/10.1016/S0167-7152(00)00137-1)
- Lynch, J. D. (1999). On conditions for mixtures of increasing failure rate distributions to have an increasing failure rate. *Probability in the Engineering and Informational Sciences* **13**, 33–36. MR1666372 <https://doi.org/10.1017/S0269964899131036>
- Marshall, A. W., Olkin, I. and Arnold, B. C. (1979). *Inequalities: Theory of Majorization and Its Applications*, Vol. 143. New York: Academic press. MR0552278
- Misra, N. and Naqvi, S. (2018). Some unified results on stochastic properties of residual lifetimes at random times. *Brazilian Journal of Probability and Statistics* **32**, 422–436. MR3787760 <https://doi.org/10.1214/16-BJPS348>

---

*Key words and phrases.*  $RR_2$ ,  $TP_2$ , up (down) shifted hazard rate order, up (down) shifted reversed hazard rate order, up (down) shifted likelihood ratio order.

- Mosler, K. and Scarsini, M. (2012). *Stochastic Orders and Applications: A Classified Bibliography, Vol. 401*. Berlin: Springer. MR1245033 <https://doi.org/10.1007/978-3-642-49972-2>
- Nakai, T. (1995). A partially observable decision problem under a shifted likelihood ratio ordering. *Mathematical and Computer Modelling* **22**, 237–246. MR1366696 [https://doi.org/10.1016/0895-7177\(95\)00201-C](https://doi.org/10.1016/0895-7177(95)00201-C)
- Nanda, A. K., Bhattacharjee, S. and Alam, S. S. (2006). On upshifted reversed mean residual life order. *Communications in Statistics Theory and Methods* **35**, 1513–1523. MR2328492 <https://doi.org/10.1080/03610920600637271>
- Pečarić, J. E., Proschan, F. and Tong, Y. L. (1992) *Convex Functions, Partial Orderings, and Statistical Applications*. *Math. Sci. Engrg.* Boston, MA: Academic Press. MR1162312
- Prékopa, A. (1971). Logarithmic concave measures with application to stochastic programming. *Acta Scientiarum Mathematicarum* **32**, 301–316. MR0315079
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. New York: Springer. MR2265633 <https://doi.org/10.1007/978-0-387-34675-5>
- Shanthikumar, J. G. and Yao, D. D. (1986). The preservation of likelihood ratio ordering under convolutions. *Stochastic Processes and Their Applications* **23**, 259–267. MR0876048 [https://doi.org/10.1016/0304-4149\(86\)90039-6](https://doi.org/10.1016/0304-4149(86)90039-6)

## Beta rectangular regression models to longitudinal data

Vinícius S. O. Ribeiro<sup>1,\*</sup>, Juvêncio S. Nobre<sup>1,†</sup>, José Roberto S. dos Santos<sup>1,‡</sup> and Caio L. N. Azevedo<sup>2</sup>

<sup>1</sup>*Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Brazil, [vinicius@osterne.com](mailto:vinicius@osterne.com); [juvencio@ufc.br](mailto:juvencio@ufc.br); [silvadosjr@dema.ufc.br](mailto:silvadosjr@dema.ufc.br)*

<sup>2</sup>*Departamento de Estatística, Universidade Estadual de Campinas, Brazil, [cnaber@ime.unicamp.br](mailto:cnaber@ime.unicamp.br)*

**Abstract.** In this paper, we propose a new approach for modeling repeated measures restricted to the interval  $(0, 1)$  (or  $(a, b)$ , with  $a < b$  known real numbers). More specifically, we developed Generalized Estimating Equations, considering beta rectangular marginal distributions, which is more robust than the usual beta distribution, against the presence of extreme observations. Diagnostic tools, including local influence analysis, were also developed. A simulation study was performed, indicating that our estimation algorithm is efficient, in terms of parameter recovery. The results also suggest that our methodology outperformed the usual beta model, for the most of the considered scenarios, indicating an interesting advantage on the estimation of the dispersion parameter. Furthermore, the developed methodology was illustrated through a real data analysis, concerning to an ophthalmic study.

## References

- Artes, R. (1997). Extensions of the theory of generalized estimation equations to circular data and dispersion models. Unpublished thesis, Instituto de Matemática e Estatística, Universidade de São Paulo. Brazil. (in Portuguese). MR1794059 <https://doi.org/10.1111/1467-842X.00131>
- Artes, R., Paula, G. A. and Ranvaud, R. (2000). Analysis of circular longitudinal data based on generalized estimating equations. *Australian & New Zealand Journal of Statistics* **42**, 347–358. MR1128675 [https://doi.org/10.1016/0047-259X\(91\)90008-P](https://doi.org/10.1016/0047-259X(91)90008-P)
- Barndorff-Nielsen, O. E. and Jørgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis* **39**, 106–116. MR3000016 <https://doi.org/10.1214/12-BA728>
- Bayes, C. L., Bazán, J. L., García, C., et al (2012). A new robust regression model for proportions. *Bayesian Analysis* **7**, 841–866. MR3886127
- Bonat, W. H., Lopes, J. E., Shimakura, S. E. and Ribeiro, P. J. Jr. (2018). Likelihood analysis for a class of simplex mixed models. *Chilean Journal of Statistics* **9**, 3–17. MR3276943 <https://doi.org/10.1080/02664763.2014.947248>
- Bonat, W. H., Ribeiro, P. J. Jr. and Zeviani, W. M. (2015). Likelihood analysis for a class of beta mixed models. *Journal of Applied Statistics* **42**, 252–266. MR1394064 <https://doi.org/10.2307/2291379>
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25. MR1974073 <https://doi.org/10.1111/1467-9876.00281>
- Cadigan, N. and Farrell, P. (2002). Generalized local influence with applications to fish stock cohort analysis. *Journal of the Royal Statistical Society Series C Applied Statistics* **51**, 469–483. MR0867994
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B (Methodological)* **48**, 133–155. MR0909363 <https://doi.org/10.1093/biomet/74.3.591>
- Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74**, 591–597.
- Cui, J. (2007). QIC program and model selection in GEE analyses. *Stata Journal* **7**, 209–220. MR1478673 <https://doi.org/10.1017/CBO9780511802843>
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application. Number 1*. Cambridge: Cambridge University Press. MR0501537
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38. MR2095753 <https://doi.org/10.1080/0266476042000214501>

- Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31**, 799–815. MR3063006 <https://doi.org/10.1016/j.cstda.2012.12.002>
- Figuroa-Zúñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis* **61**, 137–147. MR3260658 <https://doi.org/10.1002/sim.6179>
- Galvis, D. M., Bandyopadhyay, D. and Lachos, V. H. (2014). Augmented mixed beta regression models for periodontal proportion data. *Statistics in Medicine* **33**, 3759–3771. MR1837793 <https://doi.org/10.1214/lnms/1215455034>
- Godambe, V. (1997). Estimating functions: A synthesis of least squares and maximum likelihood methods. *Lecture Notes-Monograph Series* **32**, 5–15.
- Hahn, E. D. (2008). Mixture densities for project management activity times: A robust approach to pert. *European Journal of Operational Research* **188**, 450–459. MR2000388
- Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. London: Chapman and Hall/CRC. MR0055650 <https://doi.org/10.2307/3001853>
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics* **9**, 226–252. MR1461808 <https://doi.org/10.1007/b98823>
- Heyde, C. C. (2008). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Berlin: Springer.
- Hunger, M., Döring, A. and Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Medical Research Methodology* **12**, 144–144. MR1749538 <https://doi.org/10.1111/1467-9868.00230>
- Jamshidian, M. and Jennrich, R. I. (2000). Standard errors for em estimation. *Journal of the Royal Statistical Society. Series B* **62**, 257–270. MR3328438
- Joe, H. (2015). *Dependence Modeling with Copulas*. London: Chapman and Hall/CRC. MR2001447 <https://doi.org/10.1007/978-1-4757-3803-2>
- Lahiri, S. N. (2013). *Resampling Methods for Dependent Data*. Berlin: Springer.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974. MR3527412 <https://doi.org/10.1002/bimj.201500030>
- Lemonte, A. J. and Bazán, J. L. (2016). New class of Johnson distributions and its associated regression model for rates and proportions. *Biometrical Journal* **58**, 727–746. MR0836430 <https://doi.org/10.1093/biomet/73.1.13>
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- Markatou, M. (2000). Mixture models, robustness, and the weighted likelihood methodology. *Biometrics* **56**, 483–486. MR2988457 <https://doi.org/10.1214/12-EJS721>
- Masarotto, G., Varin, C., et al (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics* **6**, 1517–1549.
- Meyers, S. M., Ambler, J. S., Tan, M., Werner, J. C. and Huang, S. S. (1992). Variation of perfluoropropane disappearance after vitrectomy. *Retina* **12**, 359–363. MR3016961 <https://doi.org/10.1007/s00362-011-0417-y>
- Mitnik, P. A. and Baek, S. (2013). The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers* **54**, 177–192.
- Mousa, A. M., El-Sheikh, A. A. and Abdel-Fattah, M. A. (2016). A gamma regression for bounded continuous variables. *Advances and Applications in Statistics* **49**, 305–305. MR1500116
- Muthén, B., Asparouhov, T., Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008). *Longitudinal Data Analysis*. Boca Raton: CRC Press. MR1833297 <https://doi.org/10.1111/j.0006-341X.2001.00120.x>
- Pan, W. (2001). Akaike’s information criterion in generalized estimating equations. *Biometrics* **57**, 120–125. MR2468863 <https://doi.org/10.1111/j.1467-9469.2008.00603.x>
- Qiu, Z., Song, P. X.-K. and Tan, M. (2008). Simplex mixed-effects models for longitudinal proportional data. *Scandinavian Journal of Statistics* **35**, 577–596.
- Ribeiro, V. S. O., Nobre, J., dos Santos, J. R. S. and Azevedo, C. L. N. (2021). Supplement to “Beta rectangular regression models to longitudinal data.” <https://doi.org/10.1214/21-BJPS511SUPP>
- Rodrigues-Motta, M., Galvis Soto, D. M., Lachos, V. H., Vilca, F., Baltar, V. T., Junior, E. V., Fisberg, R. M. and Lobo Marchioni, D. M. (2015). A mixed-effect model for positive responses augmented by zeros. *Statistics in Medicine* **34**, 1761–1778.
- Santos, A. R., Azevedo, C. L., Bazan, J. L. and Nobre, J. S. (2017). Zero-one augmented heteroscedastic rectangular beta regression models. Unpublished thesis, Instituto de Matemática, Estatística e Computação Científica, Universidade de Campinas, São Paulo. Brazil. (in Portuguese). MR4296776 <https://doi.org/10.1080/02664763.2020.1783518>
- Santos, A. R., Azevedo, C. L., Bazan, J. L. and Nobre, J. S. (2020). Augmented-limited regression models with an application to the study of the risk perceived using continuous scales. *Journal of Applied Statistics*. MR3686569 <https://doi.org/10.1111/insr.12178>

- Singer, J. M., Rocha, F. M. and Nobre, J. S. (2017). Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* **85**, 290–324.
- Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* **11**, 54–54. MR2101142 <https://doi.org/10.1002/bimj.200110052>
- Song, P. X.-K., Qiu, Z. and Tan, M. (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **46**, 540–553.
- Song, P. X.-K. and Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics* **56**, 496–502.
- Tan, M., Qu, Y. and Kutner, M. H. (1997). Model diagnostics for marginal regression analysis of correlated binary data. *Communications in Statistics Simulation and Computation* **26**, 539–558. MR4140778 <https://doi.org/10.1007/s11749-019-00675-1>
- Tsuyuguchi, A. B., Paula, G. A. and Barros, M. (2020). Analysis of correlated birnbaum-saunders data based on estimating equations. *TEST*, 29:661–681.
- Usuga, O. (2013). *Beta Regression Models with Normal and Non-normal Random Effects for Longitudinal Data*. Brazil: Instituto de Matemática e Estatística, Universidade de São Paulo. Unpublished thesis (in Portuguese).
- Venezuela, M. K. (2008). Generalized estimation equation and local influence for beta regression models with repeated measures. Unpublished thesis, Instituto de Matemática e Estatística, Universidade de São Paulo. Brazil. (in Portuguese). MR2409950 <https://doi.org/10.1080/10629360600780488>
- Venezuela, M. K., Botter, D. A. and Sandoval, M. C. (2007). Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation* **77**, 879–888. MR2748685 <https://doi.org/10.1016/j.csda.2010.10.020>
- Venezuela, M. K., Sandoval, M. C. and Botter, D. A. (2011). Local influence in estimating equations. *Computational Statistics & Data Analysis* **55**, 1867–1883.
- Verkuilen, J. and Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics* **37**, 82–113. MR0012401 <https://doi.org/10.2307/1990256>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482. MR3780987 <https://doi.org/10.4310/SII.2018.v11.n2.a7>
- Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: A review and recent developments. *Advances in Statistics* **2014**, 1–11. MR2727758 <https://doi.org/10.1016/j.csda.2009.11.006>
- Wang, Y.-G. and Hin, L.-Y. (2010). Modeling strategies in longitudinal data analysis: Covariate, variance function and correlation structure selection. *Computational Statistics & Data Analysis* **54**, 3359–3370.
- Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *American Statistician* **39**, 95–101. MR0980999 <https://doi.org/10.2307/2531734>
- Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049–1060. MR3760463 <https://doi.org/10.1002/sim.7573>
- Zhao, W., Lian, H. and Bandyopadhyay, D. (2018). A partially linear additive model for clustered proportion data. *Statistics in Medicine* **37**, 1009–1030. MR3648627 <https://doi.org/10.1002/sim.7240>
- Zheng, X., Qin, G. and Tu, D. (2017). A generalized partially linear mean-covariance regression model for longitudinal proportional data, with applications to the analysis of quality of life data from cancer clinical trials. *Statistics in Medicine* **36**, 1884–1894.

## On variable selection in joint modeling of mean and dispersion

Edmilson Rodrigues Pinto\* and Leandro Alves Pereira†

Federal University of Uberlândia, Uberlândia, Brazil, \*[edmilson.pinto@ufu.br](mailto:edmilson.pinto@ufu.br); †[leandro.ap@ufu.br](mailto:leandro.ap@ufu.br)

**Abstract.** The joint modeling of mean and dispersion (JMMD) provides an efficient method to obtain useful models for the mean and dispersion, especially in problems of robust design experiments. However, in the literature on JMMD there are few works dedicated to variable selection and this theme is still a challenge. In this article, we propose a procedure for selecting variables in JMMD, based on hypothesis testing and the quality of the model's fit. A criterion for checking the goodness of fit is used, in each iteration of the selection process, as a filter for choosing the terms that will be evaluated by a hypothesis test. Three types of criteria were considered for checking the quality of the model fit in our variable selection procedure. The criteria used were: the extended Akaike information criterion, the corrected Akaike information criterion and a specific criterion for the JMMD, proposed by us, a type of extended adjusted coefficient of determination. Simulation studies were carried out to verify the efficiency of our variable selection procedure. In all situations considered, the proposed procedure proved to be effective and quite satisfactory. The variable selection process was applied to a real example from an industrial experiment.

## References

- Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Journal of the Royal Statistical Society Series C Applied Statistics* **36**, 332–339.
- Antoniadis, A., Gijbels, I., Lacroix, S. L. and Poggi, J. M. (2016). Joint estimation and variable selection for mean and dispersion in proper dispersion models. *Electronic Journal of Statistics* **10**, 1630–1676. [MR3522656](https://doi.org/10.1214/16-EJS1152) <https://doi.org/10.1214/16-EJS1152>
- Bayer, F. M. and Cribari Neto, F. (2017). Model selection criteria in beta regression with varying dispersion. *Communications in Statistics Simulation and Computation* **46**, 729–746. [MR3563524](https://doi.org/10.1080/03610918.2014.977918) <https://doi.org/10.1080/03610918.2014.977918>
- Bonat, W. H. and Jorgensen, 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) <https://doi.org/10.1111/rssc.12145>
- Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall. [MR1014890](https://doi.org/10.1007/978-1-4899-2873-3) <https://doi.org/10.1007/978-1-4899-2873-3>
- Charalambous, C., Pan, J. and Tranmer, M. (2014). Variable selection in joint mean and dispersion models via double penalized likelihood. *Sankhya The Indian Journal of Statistics* **76**, 276–304. [MR3302274](https://doi.org/10.1007/s13571-014-0079-6) <https://doi.org/10.1007/s13571-014-0079-6>
- Charalambous, C., Pan, J. and Tranmer, M. (2015). Variable selection in joint modelling of the mean and variance for hierarchical data. *Statistical Modelling* **15**, 24–50. [MR3306576](https://doi.org/10.1177/1471082X13520424) <https://doi.org/10.1177/1471082X13520424>
- Cuervo, E. and Gamerman, D. (2001). Bayesian modeling of variance heterogeneity in normal regression models. *Brazilian Journal of Probability and Statistics* **14**, 207–221.
- Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* **82**, 1079–1091. [MR0922172](https://doi.org/10.1080/01621459.1987.10477172)
- Davidian, M. and Carroll, R. J. (1988). A note on extended quasi-likelihood. *Journal of the Royal Statistical Society, Series B* **50**, 74–82. [MR0954733](https://doi.org/10.1093/bjbs/50.1.74)
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709–721. [MR0860505](https://doi.org/10.1080/01621459.1986.10477172)
- Engel, J. (1992). Modeling variation in industrial experiments. *Applied Statistics* **41**, 579–593.

- Engel, J. and Huele, A. F. (1996). A generalized linear modeling approach to robust design. *Technometrics* **38**, 365–373.
- Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference* **22**, 137–152. MR1004344 [https://doi.org/10.1016/0378-3758\(89\)90106-7](https://doi.org/10.1016/0378-3758(89)90106-7)
- Hu, B. and Shao, J. (2008). Generalized linear model selection using  $R^2$ . *Journal of Statistical Planning and Inference* **138**, 3705–3712. MR2455959 <https://doi.org/10.1016/j.jspi.2007.12.009>
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307. MR1016020 <https://doi.org/10.1093/biomet/76.2.297>
- Jorgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B* **49**, 127–162. MR0905186
- Jorgensen, B. (1992). *The Theory of Exponential Dispersion Models and Analysis of Deviance*. Rio de Janeiro: IMPA. MR1182990
- Jorgensen, B. (1997). *The Theory of Dispersion Models*. New York: Chapman and Hall. MR1462891
- Lee, K. I. and Koval, J. J. (1997). Determination of the best significance level in forward stepwise logistic regression. *Communications in Statistics Simulation and Computation* **26**, 559–575.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B* **58**, 619–678. MR1410182
- Lee, Y. and Nelder, J. A. (1998). Generalized linear models for analysis of quality improvement experiments. *Canadian Journal of Statistics* **26**, 95–105. MR1624385 <https://doi.org/10.2307/3315676>
- Lee, Y. and Nelder, J. A. (2000). The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling Geissler’s human sex ratio data. *Applied Statistics* **49**, 413–419. MR1824549 <https://doi.org/10.1111/1467-9876.00201>
- Lee, Y. and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987–1006. MR1872215 <https://doi.org/10.1093/biomet/88.4.987>
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood*. Boca Raton: Chapman and Hall. MR2259540 <https://doi.org/10.1201/9781420011340>
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall. MR3223057 <https://doi.org/10.1007/978-1-4899-3242-6>
- Nair, V. N., ed. (1992). Taguchi’s parameter design: A panel discussion. *Technometrics* **34**, 127–161. MR2439873 <https://doi.org/10.1198/004017008000000136>
- Nelder, J. A. (2000). Quasi-likelihood and pseudo-likelihood are not the same thing. *Journal of Applied Statistics* **27**, 1007–1011.
- Nelder, J. A. and Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis* **7**, 107–120. MR1151920 <https://doi.org/10.1002/asm.3150070405>
- Nelder, J. A. and Lee, Y. (1992). Likelihood, quasi-likelihood and pseudolikelihood: Some comparisons. *Journal of the Royal Statistical Society, Series B* **54**, 273–284. MR1157725
- Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221–232. MR0903123 <https://doi.org/10.1093/biomet/74.2.221>
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A* **135**, 370–384. MR0375592 <https://doi.org/10.1093/biomet/61.3.439>
- Phadke, M. S. (1989). *Quality Engineering Using Robust Design*. New York: Prentice-Hall.
- R Development Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austrian: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society, Series B* **51**, 47–60. MR0984992
- Smyth, G. K. and Verbyla, A. P. (1999a). Double generalized linear models: Approximate REML and diagnostics. In *Statistical Modelling: Proceedings of the 14th International Workshop on Statistical Modelling (IWSM14)* (H. Friedl, A. Berghold and G. Kauermann, eds.) 66–88. Graz: Austrian.
- Smyth, G. K. and Verbyla, A. P. (1999b). Adjusted likelihood methods for modeling dispersion in generalized linear models. *EnvironMetrics* **10**, 696–709.
- Sugiura, N. (1978). Further analysis of the data by Akaike’s information criterion and the finite corrections. *Communications in Statistics Theory and Methods* **7**, 13–26.
- Taguchi, G. (1986). *Introduction to Quality Engineering*. New York: Unipub/Kraus International Publications.
- Verbyla, A. P. (1993). Modelling variance heterogeneity: Residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society, Series B* **55**, 493–508. MR1224412
- Wang, D. and Zhang, Z. (2009). Variable selection in joint generalized linear models. *Chinese Journal of Applied Probability and Statistics* **25**, 245–256. MR2641680
- Wu, L. and Li, H. (2012). Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. *Metrika* **75**, 795–808. MR2956276 <https://doi.org/10.1007/s00184-011-0352-x>

- Wu, L., Li, S. and Tao, Y. (2020). Estimation and variable selection for mixture of joint mean and variance models. *Communications in Statistics—Theory and Methods*, 1–18. [MR4071582](#)
- Wu, L., Zhang, Z. and Xu, D. (2012). Variable selection for joint mean and dispersion models of the lognormal distribution. *Hacettepe Journal of Mathematics and Statistics* **41**, 307–320. [MR3012185](#)
- Zhang, D. (2017). A coefficient of determination for generalized linear models. *American Statistician* **71**, 310–316. [MR3750938](#) <https://doi.org/10.1080/00031305.2016.1256839>



## 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/>

**Permissions Policy.** Authorization to photocopy items for internal or personal use is granted by the Institute of Mathematical Statistics. For multiple copies or reprint permission, contact The Copyright Clearance Center, 222 Rosewood Drive, Danvers, Massachusetts 01923. Telephone (978) 750-8400. <http://www.copyright.com>. If the permission is not found at the Copyright Clearance Center, please contact the IMS Business Office: [ims@imstat.org](mailto:ims@imstat.org).

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