Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 35 • Number 2 • May 2021

Brazilian Journal of Probability and Statistics

Volume 35 • Number 2 • May 2021

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 35, Number 2, May 2021. Published quarterly by the Brazilian Statistical Association.

POSTMASTER:

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

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

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 205–223
https://doi.org/10.1214/20-BJPS472

The local principle of large deviations for compound Poisson process with catastrophes

Artem Logachov[1,2,3,*], Olga Logachova[3,4,†] and Anatoly Yambartsev[5]

[1]*Laboratory of Probability Theory and Mathematical Statistics, Sobolev Institute of Mathematics, Siberian Branch of the RAS, Koptyuga str. 4, Novosibirsk 630090, Russian Federation*
[2]*Novosibirsk State University, Pirogova str. 1, Novosibirsk 630090, Russian Federation*
[3]*Novosibirsk State University of Economics and Management, Kamenskaya str. 56, Novosibirsk 630099, Russian Federation,* [*]*a.v.logachev@sgugit.ru;* [†]*o.m.logacheva@sgugit.ru*
[4]*Siberian State University of Geo-systems and Technologies, Plakhotnogo str. 10, Novosibirsk 630108, Russian Federation*
[5]*Institute of Mathematics and Statistics, University of São Paulo, 1010 Rua do Matão, CEP 05508–090, São Paulo SP, Brazil,* *yambar@ime.usp.br*

Abstract. The continuous time Markov process considered in this paper belongs to a class of population models with linear growth and catastrophes. There, the catastrophes happen at the arrival times of a Poisson process, and at each catastrophe time, a randomly selected portion of the population is eliminated. For this population process, we derive an asymptotic upper bound for the maximum value and prove the local large deviation principle.

References

Ben-Ari, I., Roitershtein, A. and Schinazi, R. B. (2019). A random walk with catastrophes. *Electronic Journal of Probability* **24**, Art. ID 28. MR3933207 https://doi.org/10.1214/19-EJP282

Bhat, U., De Bacco, C. and Redner, S. (2016). Stochastic search with Poisson and deterministic resetting. *Journal of Statistical Mechanics: Theory and Experiment* **2016**, Art. ID 083401. MR3548032 https://doi.org/10.1088/1742-5468/2016/08/083401

Borovkov, A. A. and Mogul'skii, A. A. (2010). On large deviation principles in metric spaces. *Siberian Mathematical Journal* **51**, 989–1003. MR2797595 https://doi.org/10.1007/s11202-010-0098-0

Borovkov, A. A. and Mogulskii, A. A. (2012). Large deviation principles for random walk trajectories. I. *Theory of Probability and Its Applications* **56**, 538–561. MR3267282 https://doi.org/10.1137/S0040585X97986370

Borovkov, A. A. and Mogul'skii, A. A. (2013). Inequalities and principles of large deviations for the trajectories of processes with independent increments. *Siberian Mathematical Journal* **54**, 217–226. MR3088596 https://doi.org/10.1134/S0037446613020055

Brockwell, P. J., Gani, J. and Resnick, S. I. (1982). Birth, immigration and catastrophe processes. *Advances in Applied Probability* **14**, 709–731. MR0677553 https://doi.org/10.2307/1427020

Budhiraja, A., Dupuis, P. and Ganguly, A. (2016). Moderate deviation principles for stochastic differential equations with jumps. *The Annals of Probability* **44**, 1723–1775. MR3502593 https://doi.org/10.1214/15-AOP1007

Chao, X. (1995). A queueing network model with catastrophes and product form solution. *Operations Research Letters* **18**, 75–79. MR1361864 https://doi.org/10.1016/0167-6377(95)00029-0

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. New York: Springer. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

Den Hollander, F., Majumdar, S. N., Meylahn, J. M. and Touchette, H. (2019). Properties of additive functionals of Brownian motion with resetting. *Journal of Physics A: Mathematical and Theoretical* **52**, Art. ID 175001. MR3938125 https://doi.org/10.1088/1751-8121/ab0efd

Di Crescenzo, A., Giorno, V., Nobile, A. G. and Ricciardi, L. M. (2008). A note on birth–death processes with catastrophes. *Statistics & Probability Letters* **78**, 2248–2257. MR2458034 https://doi.org/10.1016/j.spl.2008.01.093

Dimou, S. and Economou, A. (2013). The single server queue with catastrophes and geometric reneging. *Methodology and Computing in Applied Probability* **15**, 595–621. MR3085882 https://doi.org/10.1007/s11009-011-9271-6

Key words and phrases. Compound Poisson processes, processes with resettings, processes with catastrophes, large deviation principle, local large deviation principle.

Doyne Farmer, J., Gillemot, L., Lillo, F., Mike, S. and Sen, A. (2004). What really causes large price changes? *Quantitative Finance* **4**, 383–397.

Eule, S. and Metzger, J. J. (2016). Non-equilibrium steady states of stochastic processes with intermittent resetting. *New Journal of Physics* **18**, Art. ID 033006.

Evans, M. R. and Majumdar, S. N. (2011). Diffusion with stochastic resetting. *Physical Review Letters* **106**, Art. ID 160601.

Feng, J. and Kurtz, T. (2006). *Large Deviations for Stochastic Processes*. *Mathematical Surveys and Monographs* **131**. Providence, RI: American Mathematical Society. MR2260560 https://doi.org/10.1090/surv/131

Gupta, S., Majumdar, S. N. and Schehr, G. (2014). Fluctuating interfaces subject to stochastic resetting. *Physical Review Letters* **112**, Art. ID 220601.

Kaplan, N., Sudbury, A. and Nilsen, T. S. (1975). A branching process with disasters. *Journal of Applied Probability* **12**, 47–59. MR0362531 https://doi.org/10.1017/s0021900200033088

Kyriakidis, E. G. (1993). A Markov decision algorithm for optimal pest control through uniform catastrophes. *European Journal of Operational Research* **64**, 38–44.

Kyriakidis, E. G. and Abakuks, A. (1989). Optimal pest control through catastrophes. *Journal of Applied Probability* **26**, 873–879. MR1025404 https://doi.org/10.2307/3214392

Li, Q. L., Ma, J. Y. and Chang, Y. X. (2018). Blockchain queue theory. In *International Conference on Computational Social Networks* 25–40. Cham: Springer.

Logachov, A. (2012). Large deviation principle for processes with Poisson noise term. *Theory of Stochastic Processes* **18**, 59–76. MR3124775

Logachov, A., Logachova, O. and Yambartsev, A. (2019). Large deviations in a population dynamics with catastrophes. *Statistics & Probability Letters* **149**, 29–37. MR3906785 https://doi.org/10.1016/j.spl.2019.01.029

Ma, X. and Xi, F. (2017). Moderate deviations for neutral stochastic differential delay equations with jumps. *Statistics & Probability Letters* **126**, 97–107. MR3634585 https://doi.org/10.1016/j.spl.2017.02.034

Meylahn, J. M., Sabhapandit, S. and Touchette, H. (2015). Large deviations for Markov processes with resetting. *Physical Review E* **92**, Art. ID 062148.

Mogulskii, A. A. (1993). Large deviations for processes with independent increments. *The Annals of Probability* **21**, 202–215. MR1207223

Montero, M. and Villarroel, J. (2016). Directed random walk with random restarts: The Sisyphus random walk. *Physical Review E* **94**, Art. ID 032132. MR3731822 https://doi.org/10.1103/physreve.94.032132

Pakes, A. G., Trajstman, A. C. and Brockwell, P. J. (1979). A stochastic model for a replicating population subjected to mass emigration due to population pressure. *Mathematical Biosciences* **45**, 137–157. MR0535957 https://doi.org/10.1016/0025-5564(79)90099-3

Puhalskii, A. (2004). On some degenerate large deviation problems. *Electronic Journal of Probability* **9**, 862–886. MR2110021 https://doi.org/10.1214/EJP.v9-232

Riesz, F. and Szőkefalvi-Nagy, B. (1990). *Functional Analysis*. New York: Dover. MR1068530

Röckner, M. and Zhang, T. (2007). Stochastic evolution equations of jump type: Existence, uniqueness and large deviation principles. *Potential Analysis* **26**, 255–279. MR2286037 https://doi.org/10.1007/s11118-006-9035-z

Trajstman, A. C. (1981). A bounded growth population subjected to emigrations due to population pressure. *Journal of Applied Probability* **18**, 571–582. MR0621221 https://doi.org/10.2307/3213312

Visco, P., Allen, R. J., Majumdar, S. N. and Evans, M. R. (2010). Switching and growth for microbial populations in catastrophic responsive environments. *Biophysical Journal* **98**, 1099–1108.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 224–241
https://doi.org/10.1214/20-BJPS474

A switch convergence for a small perturbation of a linear recurrence equation

Gerardo Barrera* **and Shuo Liu†**

Department of Mathematical and Statistical Sciences, University of Alberta, Central Academic Building, 116 Street and 85 Avenue, Edmonton, Alberta T6G-2G1, Canada, *barrerav@ualberta.ca; †sliu5@ualberta.ca

Abstract. In this article we study a small random perturbation of a linear recurrence equation. If all the roots of its corresponding characteristic equation have modulus strictly less than one, the random linear recurrence goes exponentially fast to its limiting distribution in the total variation distance as time increases. By assuming that all the roots of its corresponding characteristic equation have modulus strictly less than one and rather mild conditions, we prove that this convergence happens as a switch-type, i.e., there is a sharp transition in the convergence to its limiting distribution. This fact is known as a cut-off phenomenon in the context of stochastic processes.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**, 243–247. MR0246476

Allen, B. and Nowak, M. (2015). Games among relatives revisited. *Journal of Theoretical Biology* **378**, 103–116. MR3350852

Barrera, G. (2018). Abrupt convergence for a family of Ornstein–Uhlenbeck processes. *Brazilian Journal of Probability and Statistics* **32**, 188–199. MR3770869

Barrera, G. and Jara, M. (2016). Abrupt convergence of stochastic small perturbations of one dimensional dynamical systems. *Journal of Statistical Physics* **163**, 113–138. MR3472096

Barrera, G. and Jara, M. (2020). Thermalisation for small random perturbation of dynamical systems. *The Annals of Applied Probability* **30** (3), 1164–1208.

Barrera, J. and Ycart, B. (2014). Bounds for left and right window cutoffs. *ALEA. Latin American Journal of Probability and Mathematical Statistics* **11**, 445–458. MR3265085

Coddington, E. and Levinson, N. (1987). *Theory of Ordinary Differential Equations*. McGraw-Hill Book Company, Inc. Ninth Reprint. MR0069338

Cormen, T., Leiserson, C., Rivest, R. and Stein, C. (2009). *Introduction to Algorithms*, 3rd ed. MIT Press. MR2572804

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37. MR1429916

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer. MR2664452

Diaconis, P. (1996). The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1659–1664. MR1374011

Dubeau, F. (1993). The rabbit problem revisited. *The Fibonacci Quarterly* **31**, 268–273. MR1228566

Elaydi, S. (2005). *An Introduction to Difference Equations*, 3rd ed. Springer. MR2128146

Everest, G., van der Poorten, A., Shparlinski, I. and Ward, T. (2003). *Recurrence Sequences*. *Mathematical Surveys and Monographs* **104**. American Mathematical Society. MR1990179

Ferguson, C. (1960). Learning, expectations, and the cobweb model. *Zeitschrift für Nationalökonomie* **20**, 297–315.

Flannery, M. and James, C. (1984). The effect of interest rate changes on the common stock returns of financial institutions. *The Journal of Finance* **39**, 1141–1153.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press. MR1278033

Klee, H. (1986). Some novel applications of a continuous system simulation language. *Computers & Industrial Engineering* **11**, 385–389.

Key words and phrases. Cut-off phenomenon, linear recurrences, Gaussian distribution, total variation distance.

Levin, D. and Peres, Y. (2017). *Markov Chains and Mixing Times*, 2nd ed. American Mathematical Society. MR3726904

Lueker, G. (1980). Some techniques for solving recurrences. *Computing Surveys* **12**, 419–436. MR0600177

Mao, X. (2008). *Stochastic Differential Equations and Applications*, 2nd ed. Horwood Publishing Limited. MR2380366

Martínez, S. and Ycart, B. (2001). Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space. *Advances in Applied Probability* **33**, 188–205. MR1825322

Matiyasevich, Y. (1993). *Hilbert's Tenth Problem*. MIT Press. MR1244324

Ouaknine, J. and Worrell, J. (2014). On the positivity problem for simple linear recurrence sequences. In *Automata, Languages, and Programming*. *Lecture Notes in Computer Science*, Vol. **8573**, 318–329. Berlin, Heidelberg: Springer. MR3238382

Reiss, R. (2012). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. *Springer Series in Statistics*. MR0988164

Smale, S. and Williams, R. (1976). The qualitative analysis of a difference equation of population growth. *Journal of Mathematical Biology* **3**, 1–4. MR0414147

Spigler, R. and Vianello, M. (1992). Liouville–Green approximations for a class of linear oscillatory difference equations of the second order. *Journal of Computational and Applied Mathematics* **41**, 105–116. MR1181712

Viana, M. and Oliveira, K. (2016). *Foundations of Ergodic Theory*. Cambridge University Press. MR3558990

Yule, G. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character* **226**, 267–298.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 242–262
https://doi.org/10.1214/20-BJPS475

Finite-memory elephant random walk and the central limit theorem for additive functionals

Iddo Ben-Ari[1,*], Jonah Green[2], Taylor Meredith[3], Hugo Panzo[4] and Xioran Tan[1,†]

[1]*University of Connecticut, Storrs, Connecticut 06269, USA, [*]iddo.ben-ari@uconn.edu; [†]xiaoran.tan@uconn.edu*
[2]*Lehman College, City University of New York, New York, New York 10468, USA, jonahgreen1129@yahoo.com*
[3]*New York University, New York, New York 10012, USA, taylor.meredith@nyu.edu*
[4]*Technion—Israel Institute of Technology, Haifa 32000, Israel, panzo@campus.technion.ac.il*

Abstract. The Central Limit Theorem (CLT) for additive functionals of Markov chains is a well-known result with a long history. In this paper, we present applications to two finite-memory versions of the Elephant Random Walk, solving a problem from Gut and Stadtmüeller (2018). We also present a derivation of the CLT for additive functionals of finite state Markov chains, which is based on positive recurrence, the CLT for IID sequences and some elementary linear algebra, and which focuses on characterization of the variance.

References

Bercu, B. (2018). A martingale approach for the elephant random walk. *J. Phys. A* **51** (1), 015201, 16. MR3741953 https://doi.org/10.1088/1751-8121/aa95a6

Coletti, C. F., Gava, R. and Schütz, G. M. (2017). Central limit theorem and related results for the elephant random walk. *J. Math. Phys.* **58**, 053303. 8. MR3652225 https://doi.org/10.1063/1.4983566

Cuny, C. and Peligrad, M. (2012). Central limit theorem started at a point for stationary processes and additive functionals of reversible Markov chains. *J. Theoret. Probab.* **25**, 171–188. MR2886384 https://doi.org/10.1007/s10959-010-0321-8

Gut, A. (2013). *Probability: A Graduate Course*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR2977961 https://doi.org/10.1007/978-1-4614-4708-5

Gut, A. and Stadtmüeller, U. (2018). Variations of the elephant random walk. https://arxiv.org/abs/1812.01915.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104**, 1–19. MR0834478

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28**, 713–724. MR1782272 https://doi.org/10.1214/aop/1019160258

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. London: Springer-Verlag London, Ltd. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

Schütz, G. M. and Trimper, S. (2004). Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk. *American Physical Society: Physical Review E* **70**.

Seppäläinen, T. (2006). Notes on the martingale approach to central limit theorems for Markov chains. Available at http://www.math.wisc.edu/~seppalai/courses/833/cltnotes.pdf.

Varadhan, S. R. S. (2001). *Probability Theory*. *Courant Lecture Notes in Mathematics* **7**. New York: New York University, Courant Institute of Mathematical Sciences; Providence, RI: American Mathematical Society. MR1852999 https://doi.org/10.1090/cln/007

Varga, R. S. (2000). *Matrix Iterative Analysis*, expanded ed. *Springer Series in Computational Mathematics* **27**. Berlin: Springer. MR1753713 https://doi.org/10.1007/978-3-642-05156-2

Key words and phrases. Central limit theorem, additive functional, finite-state Markov chain, elephant random walk.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 263–280
https://doi.org/10.1214/20-BJPS479

Angle distribution of two random chords in the disc: A sine law

Jesus Igor Heberto Barahona-Torres, Paulo Cesar Manrique-Mirón and Erick Treviño-Aguilar*

Unidad Cuernavaca del Instituto de Matemáticas, Universidad Nacional Autónoma de México, Av. Universidad s/n. CP 62210, Cuernavaca, Morelos, México, *erick.trevino@im.unam.mx

Abstract. Motivated by models in engineering and also biology we determine in closed form the probability density function of the angle shaped by two random chords in a fixed disc. Our main result focus on the event in which the intersection locates inside the fixed disc and establishes a sine law.

References

Barton, D. E., David, F. N. and Fix, E. (1963). Random points in a circle and the analysis of chromosome patterns. *Biometrika* **50**, 23–29.

Brian Carleton, J., et al. (2015). Microscale modeling of layered fibrous networks with applications to biomaterials for tissue engineering. Ph.D. thesis.

Cai, T., Fan, J. and Jiang, T. (2013). Distributions of angles in random packing on spheres. *Journal of Machine Learning Research* **14**, 1837–1864. MR3104497

Christian, B., Hannes, H. and Perez-Costa, X. (2004). Stochastic properties of the random waypoint mobility model. *Wireless Networks* **10**, 555–567.

Cloherty, S. L., Hughes, N. J., Hietanen, M. A., Bhagavatula, P. S., Goodhill, G. J. and Ibbotson, M. R. (2016). Sensory experience modifies feature map relationships in visual cortex. *Elife* **5**, e13911.

Coon, J. P., Badiu, M. and Gunduz, D. (2018). On the conditional entropy of wireless networks, 1–6.

Coon, J. P. and Smith, P. J. (2017). Topological entropy in wireless networks subject to composite fading. In *2017 IEEE International Conference on Communications (ICC)*, 1–7.

Klain, D. A., Rota, G. C. and di Brozolo, L. A. R. (1997). *Introduction to Geometric Probability. Lezioni Lincee*. Cambridge: Cambridge University Press. MR1608265

Laurent-Gengoux, P., Petit, V. and Larue, L. (2019). Modeling and analysis of melanoblast motion. *Journal of Mathematical Biology*. MR4031807 https://doi.org/10.1007/s00285-019-01422-8

Morton, R. R. A. (1966). The expected number and angle of intersections between random curves in a plane. *Journal of Applied Probability* **3**, 559–562. MR0203756 https://doi.org/10.1017/s0021900200114342

Orozco-Barbosa, L. (2017). Performance study of the IEEE 802.15.6 slotted aloha mechanism with power control in a multiuser environment. In *Ad Hoc Networks*, 27–37. Berlin: Springer.

Seneta, E., Hunger Parshall, K. and Jongmans, F. (2001). Nineteenth-century developments in geometric probability: J. J. Sylvester, M. W. Crofton, J.-É. Barbier, and J. Bertrand. *Archive for History of Exact Sciences* **55**, 501–524. MR1836759 https://doi.org/10.1007/s004070100038

Solomon, H. (1978). *Geometric Probability*. *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.

Key words and phrases. Geometric probability, random angle, sine law.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 281–292
https://doi.org/10.1214/20-BJPS480

The minimal observable clade size of exchangeable coalescents

Fabian Freund[1] **and Arno Siri-Jégousse**[2]

[1]*Crop Plant Biodiversity and Breeding Informatics Group (350b), Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany, fabian.freund@uni-hohenheim.de*

[2]*Departamento de Probabilidad y Estadística, IIMAS, Universidad Nacional Autónoma de México, Mexico City, Mexico, arno@sigma.iimas.unam.mx*

Abstract. For Λ-n-coalescents with mutation, we analyse the minimal observable clade size O_n. For a given individual $i \in \{1, \dots, n\}$, the minimal observable clade size $O_n(i)$ is 1 plus the number of other individuals j such that every non-singleton mutation inherited by i is also inherited by j. We provide asymptotics of O_n as $n \to \infty$ and a recursion for all moments of O_n for finite n. This statistic gives an upper bound for the minimal clade size (*Advances in Applied Probability* **37** (2005) 647–662) which is not observable in real data (in contrast to O_n). In applications to genetics, it has been shown to be useful to lower classification errors in genealogical model selection (Freund and Siri-Jégousse, 2020).

References

Arratia, R., Barbour, A. D. and Tavaré, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. Zürich: European Mathematical Society (EMS). MR2032426 https://doi.org/10.4171/000

Blum, M. G. B. and François, O. (2005). Minimal clade size and external branch length under the neutral coalescent. *Advances in Applied Probability* **37**, 647–662. MR2156553 https://doi.org/10.1239/aap/1127483740

Caliebe, A., Neininger, R., Krawczak, M. and Rösler, U. (2007). On the length distribution of external branches in coalescence trees: Genetic diversity within species. *Theoretical Population Biology* **72**, 245–252.

Desai, M. M., Walczak, A. M. and Fisher, D. S. (2013). Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* **193**, 565–585.

Dhersin, J.-S., Freund, F., Siri-Jégousse, A. and Yuan, L. (2013). On the length of an external branch in the Beta-coalescent. *Stochastic Processes and Their Applications* **123**, 1691–1715. MR3027896 https://doi.org/10.1016/j.spa.2012.12.010

Eldon, B. and Wakeley, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172**, 2621–2633.

Freund, F. (2020). Cannings models, populations size changes and multiple-merger coalescents. *Journal of Mathematical Biology* **80**, 1497–1521. MR4071423 https://doi.org/10.1007/s00285-020-01470-5

Freund, F. and Möhle, M. (2009). On the time back to the most recent common ancestor and the external branch length of the Bolthausen–Sznitman coalescent. *Markov Processes and Related Fields* **15**, 387–416. MR2554368

Freund, F. and Möhle, M. (2017). On the size of the block of 1 for Ξ-coalescents with dust. *Modern Stochastics: Theory and Applications* **4**, 407–425. MR3739016 https://doi.org/10.15559/17-vmsta92

Freund, F. and Siri-Jégousse, A. (2014). Minimal clade size in the Bolthausen–Sznitman coalescent. *Journal of Applied Probability* **51**, 657–668. MR3256218 https://doi.org/10.1239/jap/1409932665

Freund, F. and Siri-Jégousse, A. (2020). The impact of genetic diversity statistics on model selection between coalescents. *Computational Statistics & Data Analysis*, 107055. https://doi.org/10.1016/j.csda.2020.107055

Gnedin, A., Iksanov, A. and Marynych, A. (2014). Λ-coalescents: A survey. *Journal of Applied Probability* **51A**, 23–40. MR3317347 https://doi.org/10.1239/jap/1417528464

Goldschmidt, C. and Martin, J. B. (2005). Random recursive trees and the Bolthausen–Sznitman coalescent. *Electronic Journal of Probability* **10**, 718–745. MR2164028 https://doi.org/10.1214/EJP.v10-265

Griffiths, R. C. and Tavaré, S. (1994). Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **344**, 403–410.

Key words and phrases. Clade size, Λ-coalescent, recursion.

Kingman, J. F. C. (1975). Random discrete distributions. *Journal of the Royal Statistical Society, Series B* **37**, 1–15. MR0368264

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and Their Applications* **13**, 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

Kingman, J. F. C. (1993). *Poisson Processes*. New York: Wiley Online Library. MR1207584

Matuszewski, S., Hildebrandt, M. E., Achaz, G. and Jensen, J. D. (2018). Coalescent processes with skewed offspring distributions and non-equilibrium demography. *Genetics* **208**, 323–338.

Neher, R. A. and Hallatschek, O. (2013). Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 437–442.

Pitman, J. (1999). Coalescents with multiple collisions. *Annals of Probability* **27**, 1870–1902. MR1742892 https://doi.org/10.1214/aop/1022677552

Rauch, E. and Bar-Yam, Y. (2004). Theory predicts the uneven distribution of genetic diversity within species. *Nature* **431**, 449–452.

Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability* **36**, 1116–1125. MR1742154 https://doi.org/10.1017/s0021900200017903

Schweinsberg, J. (2003). Coalescent processes obtained from supercritical Galton–Watson processes. *Stochastic Processes and Their Applications* **106**, 107–139. MR1983046

Schweinsberg, J. (2017). Rigorous results for a population model with selection II: Genealogy of the population. *Electron. J. Probab.*, **22**. MR3646064 https://doi.org/10.1214/17-ejp58

Siri-Jégousse, A. and Yuan, L. (2016). Asymptotics of the minimal clade size and related functionals of certain beta-coalescents. *Acta Applicandae Mathematicae* **142**, 127–148. MR3466919 https://doi.org/10.1007/s10440-015-0020-7

Siri-Jégousse, A. and Yuan, L. (2018). A note on the small-time behaviour of the largest block size of beta n-coalescents. In *XII Symposium of Probability and Stochastic Processes*. Cham: Springer. MR3835795 https://doi.org/10.1007/978-3-319-77643-9_8

Spence, J. P., Kamm, J. A. and Song, Y. S. (2016). The site frequency spectrum for general coalescents. *Genetics* **202**, 1549–1561.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 293–314
https://doi.org/10.1214/20-BJPS473

Modeling $\mathbb{Z}$-valued time series based on new versions of the Skellam INGARCH model

Yan Cui[1,*], **Qi Li**[2] **and Fukang Zhu**[1,†]

[1]*School of Mathematics, Jilin University, 2699 Qianjin, Changchun 130012, China,* [*]*ycui0311@163.com;* [†]*zfk8010@163.com*

[2]*College of Mathematics, Changchun Normal University, Changchun 130032, China, 46968158@qq.com*

Abstract. Recently, there has been a growing interest in integer-valued time series models, including integer-valued autoregressive (INAR) models and integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models, but only a few of them can deal with data on the full set of integers, that is, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. Although some attempts have been made to deal with $\mathbb{Z}$-valued time series, these models do not provide enough flexibility in modeling some specific integers (e.g., 0, ± 1). A symmetric Skellam INGARCH(1, 1) model was proposed in the literature, but it only considered zero-mean processes, which limits its application. We first extend the symmetric Skellam INGARCH model to an asymmetric version, which can deal with non-zero-mean processes. Then we propose a modified Skellam model which adopts a careful treatment on integers 0 and ± 1 to satisfy a special feature of the data. Our models are easy-to-use and flexible. The maximum likelihood method is used to estimate unknown parameters and the log-likelihood ratio test statistic is provided for testing the asymmetric model against the modified one. Simulation studies are given to evaluate performances of the parametric estimation and log-likelihood ratio test. A real data example is also presented to demonstrate good performances of newly proposed models.

References

Alomani, G., Alzaid, A. and Omair, M. (2018). A Skellam INGARCH model. *Brazilian Journal of Probability and Statistics* **32**, 200–214. MR3770870 https://doi.org/10.1214/16-BJPS338

Alzaid, A. and Omair, M. (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society* **33**, 17–45. MR2603338

Alzaid, A. and Omair, M. (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society* **37**, 465–485. MR3188051

Andersson, J. and Karlis, D. (2014). A parametric time series model with covariates for integers in $\mathbb{Z}$. *Statistical Modelling* **14**, 135–156. MR3192556 https://doi.org/10.1177/1471082X13504719

Barreto-Souza, W. and Bourguignon, M. (2015). A skew INAR(1) process on $\mathbb{Z}$. *Advances in Statistical Analysis* **99**, 189–208. MR3328160 https://doi.org/10.1007/s10182-014-0236-2

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer. MR2839251

Chesneau, C. and Kachour, M. (2012). A parametric study for the first-order signed integer-valued autoregressive process. *Journal of Statistical Theory and Practice* **6**, 760–782. MR3196577 https://doi.org/10.1080/15598608.2012.719816

Christou, V. and Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* **35**, 55–78. MR3148248 https://doi.org/10.1111/jtsa.12050

Christou, V. and Fokianos, K. (2015). Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electronic Journal of Statistics* **9**, 1357–1377. MR3360730 https://doi.org/10.1214/15-EJS1044

Key words and phrases. 0 and ± 1 inflations, INGARCH processes, modified Skellam, Skellam, $\mathbb{Z}$-valued time series.

Davis, R. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* **26**, 1673–1707. MR3586234

Doukhan, P., Fokianos, K. and Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statistics & Probability Letters* **82**, 942–948. MR2910041 https://doi.org/10.1016/j.spl.2012.01.015

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* **27**, 923–942. MR2328548 https://doi.org/10.1111/j.1467-9892.2006.00496.x

Fisz, M. (1953). The limiting distribution of the difference of two Poisson random variables. *Applicationes Mathematicae* **1**, 41–45 (in Polish). MR0056863 https://doi.org/10.4064/am-1-1-41-45

Fisz, M. (1955). Refinement of a probability limit theorem and its application to Bessel functions. *Acta Mathematica Hungarica* **6**, 199–202. MR0086433 https://doi.org/10.1007/bf02021276

Fokianos, K. (2016). Statistical analysis of count time series models: A GLM perspective. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. H. Holan, R. Lund and N. Ravishanker, eds.) 3–27. Boca Raton, FL: Chapman and Hall/CRC. MR3699398

Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104**, 1430–1439. MR2596998 https://doi.org/10.1198/jasa.2009.tm08270

Freeland, R. (2010). True integer value time series. *Advances in Statistical Analysis* **94**, 217–229. MR2733172 https://doi.org/10.1007/s10182-010-0135-0

Gonçalves, E. and Mendes-Lopes, N. (2020). Signed compound Poisson integer-valued GARCH processes. *Communications in Statistics—Theory and Methods* **49**, 5468–5492. https://doi.org/10.1080/03610926.2019.1619767

Kachour, M. and Truquet, L. (2011). A p-order signed integer-valued autoregressive (SINAR(p)) model. *Journal of Time Series Analysis* **32**, 223–236. MR2808251 https://doi.org/10.1111/j.1467-9892.2010.00694.x

Kachour, M. and Yao, J. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis* **30**, 417–448. MR2536061 https://doi.org/10.1111/j.1467-9892.2009.00620.x

Karlis, D. and Ntzoufras, I. (2009). Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics* **20**, 133–145.

Kim, H. and Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers* **49**, 485–502. MR2399216 https://doi.org/10.1007/s00362-006-0028-1

Koopman, S., Lit, R. and Lucas, A. (2017). Intraday stochastic volatility in discrete price changes: The dynamic Skellam model. *Journal of the American Statistical Association* **112**, 1490–1503. MR3750871 https://doi.org/10.1080/01621459.2017.1302878

Liu, X. and Luger, R. (2015). Unfolded GARCH models. *Journal of Economic Dynamics and Control* **58**, 186–217. MR3378428 https://doi.org/10.1016/j.jedc.2015.06.007

Milhøj, A. (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics* **12**, 281–292. MR0841449

Münnix, M. C., Schäfer, R. and Guhr, T. (2010). Impact of the tick-size on financial returns and correlations. *Physica A: Statistical Mechanics and Its Applications* **389**, 4828–4843.

Nastic, A., Ristic, M. and Djordjevic, M. (2016). An INAR model with discrete Laplace marginal distributions. *Brazilian Journal of Probability and Statistics* **30**, 107–126. MR3453517 https://doi.org/10.1214/14-BJPS262

Neumann, M. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17**, 1268–1284. MR2854772 https://doi.org/10.3150/10-BEJ313

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

Silva, R. B. and Barreto-Souza, W. (2019). Flexible and robust mixed Poisson INGARCH models. *Journal of Time Series Analysis* **40**, 788–814. MR3995668 https://doi.org/10.1111/jtsa.12459

Skellam, J. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society Series A* **109**, 296. MR0020750

Strackee, J. and van der Gon, J. J. D. (1962). The frequency distribution of the difference between two Poisson variates. *Statistica Neerlandica* **16**, 17–23.

Tjøstheim, D. (2016). Count time series with observation-driven autoregressive parameter dynamics. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. H. Holan, R. Lund and N. Ravishanker, eds.) 77–100. Boca Raton, FL: Chapman and Hall/CRC. MR3699401

Weiß, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. Hoboken, NJ: Wiley.

Zhang, H., Wang, D. and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference* **140**, 667–683. MR2558396 https://doi.org/10.1016/j.jspi.2009.08.012

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* **32**, 54–67. MR2790672 https://doi.org/10.1111/j.1467-9892.2010.00684.x

Zhu, F. (2012). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference* **142**, 826–839. MR2863870 https://doi.org/10.1016/j.jspi.2011.10.002

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 315–334
https://doi.org/10.1214/20-BJPS476

Estimation of semiparametric models with errors following a scale mixture of Gaussian distributions

Marcelo M. Taddeo[1] **and Pedro A. Morettin**[2]

[1]*Departamento de Estatística, Universidade Federal da Bahia, Salvador-Ba, Brasil, marcelo.magalhaes@ufba.br*
[2]*Departamento de Estatística, Universidade de São Paulo, São Paulo-SP, Brasil, pam@ime.usp.br*

Abstract. In this paper, we consider a semiparametric regression model where the error follows a scale mixture of Gaussian distributions. The purpose is to estimate the target function which is assumed to belong to some class of functions using the EM algorithm and approximations via P-splines and B-splines. We illustrate the proposed methodology through several simulation studies. Other forms of function approximation are also studied, namely Fourier and wavelet expansions.

References

Adler, R. (1990). An introduction to continuity, extrema and related topics for general Gaussian processes. *Institute of Mathematical Statistics* **12**, 1–155. MR1088478

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

Cai, T. T. and Brown, L. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics* **26**, 1783–1799. MR1673278 https://doi.org/10.1214/aos/1024691357

Cai, T. T. and Brown, L. (1999). Wavelet estimation for samples with random uniform design. *Statistics & Probability Letters* **42**, 313–321. MR1688134 https://doi.org/10.1016/S0167-7152(98)00223-5

Casella, G. and Robert, C. P. (2004). *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer. MR2080278 https://doi.org/10.1007/978-1-4757-4145-2

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**, 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–102. MR1435485 https://doi.org/10.1214/ss/1038425655

Fang, K., Kotz, S. and Ng, K. W. (1989). *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall CRC. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models—A Roughness Penalty Approach*. London: Chapman & Hall. MR1270012 https://doi.org/10.1007/978-1-4899-4473-3

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* **5**, 81–102.

Hastie, T. J. and Tibshirani, R. J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

Johnstone, I. M. and Silverman, B. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B* **59**, 319–351. MR1440585 https://doi.org/10.1111/1467-9868.00071

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Springer. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

Porto, R. F., Morettin, P. A. and Aubin, E. C. (2008). Wavelet regression with correlated errors on a piecewise Hölder class. *Statistics & Probability Letters* **78**, 2739–2743. MR2465116 https://doi.org/10.1016/j.spl.2008.03.015

Porto, R. F., Morettin, P. A., Percival, D. B. and Aubin, E. C. (2016). Wavelet shrinkage for regression models with random design and correlated errors. *Brazilian Journal of Probability and Statistics* **30**, 614–652. MR3582392 https://doi.org/10.1214/15-BJPS296

Taddeo, M. M. and Morettin, P. A. (2017). Analysis of single-index models with scale mixture of normals errors by using Bayesian P-splines. *Chilean Journal of Statistics* **8**, 3–23. MR3711737

Key words and phrases. Semiparametric models, splines, scale mixture of Normals.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: Wiley. MR1681904 https://doi.org/10.1002/9780470317020

West, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646–648. MR0909372 https://doi.org/10.1093/biomet/74.3.646

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 335–350
https://doi.org/10.1214/20-BJPS477

On a transform for modeling skewness

Li Kang[1], Paul Damien[2] and Stephen Walker[3]

[1]*Department of Statistics and Data Sciences, University of Texas at Austin, Austin, Texas, 78712, USA,*
likang@utexas.edu
[2]*McCombs Schools of Business, University of Texas at Austin, Austin, Texas, 78712, USA*
[3]*Department of Mathematics, University of Texas at Austin, Austin, Texas, 78712, USA,*
s.g.walker@math.utexas.edu

Abstract. In many applications, data exhibit skewness and in this paper we present a new family of density functions modeling skewness based on a transformation, analogous to those of location and scale. Here we note that location will always refer to mode. Hence, in order to model data to include shape, we need only to find a family of densities exhibiting a variety of shapes, since we can obtain the other three properties via the transformations. The chosen class of densities with the variety of shape is, we argue, the simplest available. Illustrations including regression and time series models are given.

References

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

Arellano-Valle, R. B., Gomez, H. W. and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* **128**, 427–443. MR2102768 https://doi.org/10.1016/j.jspi.2003.11.014

Ayebo, A. and Kozubowski, T. J. (2003). An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science* **1**, 187–210.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica Sinica* **46**, 199–208. MR0877720

Azzalini, A. and Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726. MR1440039 https://doi.org/10.1093/biomet/83.4.715

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. and Tiao, G. C. (2011). *Bayesian Inference in Statistical Analysis*. New York: Wiley.

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

Delicado, P. and Goria, M. N. (2008). A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. *Computational Statistics & Data Analysis* **52**, 1661–1673. MR2422762 https://doi.org/10.1016/j.csda.2007.05.021

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications: Vol. 2* 157–158. New York: Wiley. MR0270403

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

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

Fung, T. and Seneta, E. (2008). A characterization of scale mixtures of the uniform distribution. *Statistics & Probability Letters* **78**, 2883–2888. MR2474377 https://doi.org/10.1016/j.spl.2008.04.015

Furman, E. (2008). On a multivariate gamma distribution. *Statistics & Probability Letters* **78**, 2353–2360. MR2462670 https://doi.org/10.1016/j.spl.2008.02.012

Hinkley, D. V. and Revankar, N. S. (1977). Estimation of the Pareto law from underreported data: A further analysis. *Journal of Econometrics* **5**, 1–11. MR0652697 https://doi.org/10.1016/0304-4076(77)90031-8

Key words and phrases. Auxiliary variable, Bayesian inference, Markov chain Monte Carlo, mixtures of uniforms.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. *Springer Texts in Statistics*. New York: Springer. MR2648134 https://doi.org/10.1007/978-0-387-92407-6

Jones, M. C. (2006). A note on rescalings, reparameterizations and classes of distributions. *Journal of Statistical Planning and Inference* **136**, 3730–3733. MR2284672 https://doi.org/10.1016/j.jspi.2005.01.013

Jones, M. C. (2014). Generating distributions by transformation of scale. *Statistica Sinica* **24**, 749–771. MR3235397

Jones, M. C. (2015). On families of distributions with shape parameters (with discussion). *International Statistical Review* **83**, 175–192. MR3377075 https://doi.org/10.1111/insr.12071

Khintchine, A. Y. (1938). On unimodal distributions. *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki* **2**, 1–7.

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics* **22**, 891–921. MR2395888 https://doi.org/10.1002/jae.961

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew normal distribution for analyzing near normal data. *Journal of Statistical Planning and Inference* **83**, 291–309. MR1748020 https://doi.org/10.1016/S0378-3758(99)00096-8

Naranjo, L., Pérez, C. J. and Martín, J. (2015). Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing* **25**, 497–514. MR3334413 https://doi.org/10.1007/s11222-014-9449-1

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics* **31**, 705–767.

Rodríguez, C. E. and Walker, S. G. (2014). Univariate Bayesian nonparametric mixture modeling with unimodal kernels. *Statistics and Computing* **24**, 35–49. MR3147696 https://doi.org/10.1007/s11222-012-9351-7

Rubio, F. J. and Steel, M. F. (2014). Inference in two-piece location-scale models with Jeffrey's priors. *Bayesian Analysis* **9**, 1–22. MR3188293 https://doi.org/10.1214/13-BA849

Rubio, F. J. and Steel, M. F. (2015). Bayesian modeling of skewness and kurtosis with two-piece scale and shape distributions. *Electronic Journal of Statistics* **9**, 1884–1912. MR3391123 https://doi.org/10.1214/15-EJS1060

Subbotin, M. T. (1923). On the law of frequency of error. *Matematičeskij Sbornik* **31**, 296–301.

Theodossiou, P. (2015). Skewed generalized error distribution of financial assets and option pricing. *Multinational Finance Journal* **19**, 223–266.

Villa, C., Leisen, F. and Rossini, L. (2020). Loss-based approach to two-piece location-scale distributions with applications to dependent data. *Statistical Methods and Applications* **29**, 309–333. MR4106803 https://doi.org/10.1007/s10260-019-00481-x

Wallis, K. F. (2014). The two piece normal, binormal, or double Gaussian distributions: Its origin and rediscoveries. *Statistical Science* **29**, 106–112. MR3201857 https://doi.org/10.1214/13-STS417

West, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646–648. MR0909372 https://doi.org/10.1093/biomet/74.3.646

West, M. and Harrison, J. (1999). *Bayesian Forecasting & Dynamic Models*. New York: Springer. MR1482232

Wichitaksorn, N., Choy, S. T. B. and Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics* **42**, 579–596. MR3281462 https://doi.org/10.1002/cjs.11228

Wichitaksorn, N., Wang, J. J. J., Choy, S. T. and Gerlach, R. (2015). Analyzing return asymmetry and quantiles through stochastic volatility models using asymmetric Laplace error via uniform scale mixtures. *Applied Stochastic Models in Business and Industry* **31**, 584–608. MR3407888 https://doi.org/10.1002/asmb.2062

Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM Review* **37**, 531–551. MR1368386 https://doi.org/10.1137/1037125

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics—Theory and Methods* **34**, 1867–1879. MR2205240 https://doi.org/10.1080/03610920500199018

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* **148**, 86–99. MR2494820 https://doi.org/10.1016/j.jeconom.2008.09.038

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 351–374
https://doi.org/10.1214/20-BJPS478

On multivariate selection scale-mixtures of normal distributions

Roohollah Roozegar[1], Narayanaswamy Balakrishnan[2], Andriette Bekker[3] and Ahad Jamalizadeh[4]

[1]*Department of Mathematics, College of Sciences, Yasouj University, Yasouj, Iran, roozegar@yu.ac.ir*
[2]*McMaster University, Hamilton, Canada, bala@mcmaster.ca*
[3]*Department of Statistics, University of Pretoria, Pretoria, South Africa, andriette.bekker@up.ac.za*
[4]*Department of Statistics, Shahid Bahonar University, Kerman, Iran, A.Jamalizadeh@mail.uk.ac.ir*

Abstract. In this paper, we establish some results for multivariate selection scale-mixtures of normal distributions with arbitrary mixing variable. First, we discuss their stochastic representation in terms of multivariate selection normal distributions. Next, the conditional distributions as well as the first two moments of multivariate selection scale-mixtures of normal distributions are obtained when the selection set is an arbitrary rectangle in the q-dimensional Euclidean space of $\mathbb{R}^q$. The unified skew-scale mixture of normal (SUSMN) distributions are subsequently discussed as a special case. As a subclass of SUSMN distributions, the class of unified skew-symmetric generalized hyperbolic (SUSGH) distributions are studied in detail. Finally, we show that our results can be used to obtain moments of L-statistics and of multivariate concomitants from multivariate scale-mixtures of normal distributions.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**, 561–574. MR2298065 https://doi.org/10.1111/j.1467-9469.2006.00503.x

Arellano-Valle, R. B., Branco, M. D. and Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics* **34**, 581–601. MR2347047 https://doi.org/10.1002/cjs.5550340403

Arellano-Valle, R. B. and Genton, M. G. (2010a). An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators. *Annals of the Institute of Statistical Mathematics* **62**, 363–381. MR2592103 https://doi.org/10.1007/s10463-008-0175-3

Arellano-Valle, R. B. and Genton, M. G. (2010b). Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics* **1**, 17–33. Special Issue on "Tribute to Pilar Loreto Iglesias Zuazola". MR2756082

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2009). Multivariate order statistics via multivariate concomitants. *Journal of Multivariate Analysis* **100**, 946–951. MR2498725 https://doi.org/10.1016/j.jmva.2008.09.011

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**, 159–188. MR2188669 https://doi.org/10.1111/j.1467-9469.2005.00426.x

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B* **61**, 579–602. MR1707862 https://doi.org/10.1111/1467-9868.00194

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

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **74**, 715–729. MR1440039 https://doi.org/10.1093/biomet/83.4.715

Balakrishnan, N. (1993). Multivariate normal distribution and multivariate order statistics induced by ordering linear combinations. *Statistics & Probability Letters* **23**, 239–242. MR1237779 https://doi.org/10.1016/0167-7152(93)90253-F

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* **11**, 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8

Key words and phrases. Concomitants of order statistics, inducedorder statistics, L-statistics, multivariate scale-mixtures of normal distributions, selection distribution, unified skew-scale mixtures of normal distributions.

Cook, R. D. and Johnson, M. E. (1986). Generalized Burr Pareto logistic distributions with applications to a uranium exploration data set. *Technometrics* **28**, 123–131. MR0841485 https://doi.org/10.2307/1270448

David, H. A. (1973). Concomitants of order statistics. *Bulletin of the International Statistical Institute* **45**, 295–300. MR0373149

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

Ho, H. J., Lin, T. I., Chen, H. Y. and Wang, W. L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference* **142**, 25–40. MR2827127 https://doi.org/10.1016/j.jspi.2011.06.006

Jamalizadeh, A. and Balakrishnan, N. (2010). Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of unified skew-elliptical distributions. *Journal of Multivariate Analysis* **101**, 1412–1427. MR2609502 https://doi.org/10.1016/j.jmva.2009.12.012

Jamalizadeh, A. and Balakrishnan, N. (2019). Conditional distributions of multivariate normal mean-variance mixtures. *Statistics & Probability Letters* **145**, 312–316. MR3873920 https://doi.org/10.1016/j.spl.2018.10.005

Kelker, D. (1970). Distribution theory of spherical distributions and a location scale parameter generalization. *Sankhya Series A* **32**, 419–430. MR0287628

McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, New Jersey: Princeton University Press. MR2175089

Roozegar, R., Jamalizadeh, A. and Nematollahi, A. (2016). Concomitants of multivariate order statistics from multivariate elliptical distributions. *Communications in Statistics Theory and Methods* **45**, 722–738. MR3453009 https://doi.org/10.1080/03610926.2013.835416

Song, R. and Deddens, J. A. (1993). A note on moments of variables summing to normal order statistics. *Statistics & Probability Letters* **17**, 337–341. MR1237778 https://doi.org/10.1016/0167-7152(93)90252-E

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 375–391
https://doi.org/10.1214/20-BJPS481

Efficient estimation of the odds using judgment post stratification

Mozhgan Alirezaei Dizicheh*, Ehsan Zamanzade†,‡ and Nasrollah Iranpanah§

*Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran, *mozhganalirezaei@gmail.com; †e.zamanzade@sci.ui.ac.ir; ‡ehsanzamanzadeh@yahoo.com; §iranpanah@stat.ui.ac.ir*

Abstract. This work deals with problem of estimating the odds using judgment post stratification (JPS) sampling design. Several estimators of the odds are described and the asymptotic normality of each of them is established. Monte Carlo simulation study is then used to compare different estimators of the odds in the JPS with the standard estimator in simple random sampling (SRS) with replacement for both perfect/imperfect ranking and for both JPS data with/without empty strata. The comparison results indicate that the estimators developed here can be highly more efficient than their SRS counterpart in some certain circumstances. Finally, a real dataset from the third National Health and Nutrition Examination Survey (NHANES III) is employed for illustration purposes.

References

Al-Saleh, M. F. and Samawi, H. (2010). On estimating the odds using moving extreme ranked set sampling. *Statistical Methodology* **7**, 133–140. MR2591716 https://doi.org/10.1016/j.stamet.2009.11.004

Alirezaei Dizicheh, M., Zamanzade, E. and Iranpanah, N. (2020). Supplement to "Efficient estimation of the odds using judgment post stratification." https://doi.org/10.1214/20-BJPS481SUPP

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley. MR0474575

Das, A. K. and Tripathi, T. P. (1987). Use of auxiliary information in estimating the finite population variance. *Sankhya* **40**, 139–148.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. New York: Springer. MR2664452

Dastbaravarde, A., Arghami, N. R. and Sarmad, M. (2016). Some theoretical results concerning non parametric estimation by using a judgment poststratification sample. *Communications in Statistics Theory and Methods* **45**, 2181–2203. MR3480644 https://doi.org/10.1080/03610926.2013.878355

Dastbaravarde, A. and Zamanzade, E. (2020). On estimation of $P(X > Y)$ based on judgment post stratification. *Statistical Papers* **61**, 767–785. MR4079255 https://doi.org/10.1007/s00362-017-0962-0

Dell, T. R. and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics* **28**, 545–555.

Duembgen, L. and Zamanzade, E. (2020). Inference on a distribution function from ranked set samples. *Annals of the Institute of Statistical Mathematics* **72**, 157–185. MR4052654 https://doi.org/10.1007/s10463-018-0680-y

Frey, J. (2016). A more efficient mean estimator for judgment post-stratification. *Journal of Statistical Computation and Simulation* **86**, 1404–1414. MR3455820 https://doi.org/10.1080/00949655.2015.1064928

Frey, J. and Feeman, T. G. (2012). An improved mean estimator for judgment post-stratification. *Computational Statistics & Data Analysis* **56**, 418–426. MR2842349 https://doi.org/10.1016/j.csda.2011.08.006

Frey, J. and Feeman, T. G. (2013). Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics* **65**, 551–569. MR3067438 https://doi.org/10.1007/s10463-012-0386-5

Frey, J. and Ozturk, O. (2011). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics* **63**, 769–789. MR2805907 https://doi.org/10.1007/s10463-009-0255-z

Goga, C. and Ruiz-Gazen, A. (2018). Improving the estimation of the odds ratio using auxiliary information. Available at arXiv:1406.7691v1 [stat.ME]. MR0891703

Halls, L. K. and Dell, T. R. (1966). Trail of ranker set sampling for forage yields. *Forest Science* **12**, 22–26.

Huang, J. (1997). Asymptotic properties of the npmle of a distribution function based on ranked set samples. *The Annals of Statistics* **25**, 1036–1049. MR1447740 https://doi.org/10.1214/aos/1069362737

Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association* **78**, 117–123. MR0696855

Key words and phrases. Odds, prevalence of osteoporosis, judgment post stratified sample, imperfect ranking.

Johnell, O., Kanis, J. A., Oden, A., Sernbo, I., Redlund-Johnell, I., Petterson, C., De Laet, C. and Jnsson, B. (2004). Mortality after osteoporotic fractures. *Osteoporosis International* **15**, 38–42.

Kvam, P. H. and Samaniego, F. J. (1994). Nonparametric maximum likelihood estimation based on ranked set samples. *Journal of the American Statistical Association* **89**, 526–537. MR1294079

Leboime, A., Confavreux, C. B., Mehsen, N., Paccou, J., David, C. and Roux, C. (2010). Osteoporosis and mortality. *Joint Bone Spine* **77**, S107–S112.

Lips, P. and van Schoor, N. M. (2005). Quality of life in patients with osteoporosis. *Osteoporosis International* **16**, 447–455.

MacEachern, S. N., Stasny, E. A. and Wolfe, D. A. (2004). Judgment post-stratification with imprecise rankings. *Biometrics* **60**, 207–215. MR2044117 https://doi.org/10.1111/j.0006-341X.2004.00144.x

Mahdizadeh, M. and Zamanzade, E. (2019). Efficient body fat estimation using multistage pair ranked set sampling. *Statistical Methods in Medical Research* **28**, 223–234. MR3894524 https://doi.org/10.1177/0962280217720473

Omidvar, S., Jafari Jozani, M. and Nematollahi, N. (2018). Judgment post-stratification in finite mixture modeling: An example in estimating the prevalence of osteoporosis. *Statistics in Medicine* **37**, 4823–4836. MR3883442 https://doi.org/10.1002/sim.7984

Ozturk, O. (2014). Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics & Data Analysis* **77**, 188–205. MR3210057 https://doi.org/10.1016/j.csda.2014.02.021

Ozturk, O. (2015). Distribution free two-sample methods for judgment post-stratified data. *Statistica Sinica* **25**, 1691–1712. MR3409088

Ozturk, O. (2016). Statistical inference based on judgment post-stratified samples in finite population. *Survey Methodology* **42**, 239–262.

Robertson, T. and Waltman, P. (1968). On estimating monotone parameters. *Annals of the Institute of Statistical Mathematics* **39**, 1030–1039. MR0230408 https://doi.org/10.1214/aoms/1177698335

Srivastava, S. K. and Jhajj, H. S. (1981). A class on estimators of the population mean in survey sampling using auxiliary information. *Biometrika* **68**, 341–343. MR0614978 https://doi.org/10.1093/biomet/68.1.341

Terpstra, J. F. (2004). On estimating a population proportion via ranked set sampling. *Biometrical Journal* **46**, 264–272. MR2061924 https://doi.org/10.1002/bimj.200310022

Upadhyaya, L. N. and Singh, H. P. (1999). Use of transform auxiliary variable in estimating the finite population mean. *Biometrical Journal* **41**, 627–636. MR1720232 https://doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.3.CO;2-N

Wang, X., Lim, J. and Stokes, L. (2008). A nonparametric mean estimator for judgment poststratified data. *Biometrics* **64**, 355–363. MR2432441 https://doi.org/10.1111/j.1541-0420.2007.00900.x

Wang, X., Lim, J. and Stokes, S. L. (2016). Using ranked set sampling with cluster randomized designs for improved inference on treatment effects. *Journal of the American Statistical Association* **111**, 1576–1590. MR3601719 https://doi.org/10.1080/01621459.2015.1093946

Wang, X., Wang, K. and Lim, J. (2012). Isotonized CDF estimation from judgment poststratification data with empty strata. *Biometrics* **68**, 194–202. MR2909868 https://doi.org/10.1111/j.1541-0420.2011.01655.x

Yadav, S. K. and Kadilar, C. (2014). A two parameter variance estimator using auxiliary information. *Applied Mathematics and Computation* **226**, 117–122. MR3144295 https://doi.org/10.1016/j.amc.2013.10.044

Zamanzade, E. (2016). An isotonized variance estimator for judgment post stratified data. *Journal of the Korean Statistical Society* **45**, 25–32. MR3456319 https://doi.org/10.1016/j.jkss.2015.06.001

Zamanzade, E. and Mahdizadeh, M. (2020). Using ranked set sampling with extreme ranks in estimating the population proportion. *Statistical Methods in Medical Research* **29**, 165–177. MR4055129 https://doi.org/10.1177/0962280218823793

Zamanzade, E. and Vock, M. (2015). Variance estimation in ranked set sampling using a concomitant variable. *Statistics & Probability Letters* **105**, 1–5. MR3371973 https://doi.org/10.1016/j.spl.2015.04.034

Zamanzade, E. and Vock, M. (2018). Some nonparametric tests of perfect judgment ranking for judgment post stratification. *Statistical Papers* **59**, 1085–1100. MR3844500 https://doi.org/10.1007/s00362-016-0805-4

Zamanzade, E. and Wang, X. (2017). Estimation of population proportion for judgment post-stratification. *Computational Statistics & Data Analysis* **112**, 257–269. MR3645604 https://doi.org/10.1016/j.csda.2017.03.016

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 392–420
https://doi.org/10.1214/20-BJPS482

Bayesian estimation of a decreasing density

Geurt Jongbloed[*], **Frank van der Meulen**[†] and **Lixue Pang**[‡]

Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands, [*]*g.jongbloed@tudelft.nl;* [†]*f.h.vandermeulen@tudelft.nl;* [‡]*l.pang@tudelft.nl*

Abstract. Suppose $X_1, \dots, X_n$ is a random sample from a bounded and decreasing density f_0 on $[0, \infty)$. We are interested in estimating such f_0, with special interest in $f_0(0)$. This problem is encountered in various statistical applications and has gained quite some attention in the statistical literature. It is well known that the maximum likelihood estimator is inconsistent at zero. This has led several authors to propose alternative estimators which are consistent. As any decreasing density can be represented as a scale mixture of uniform densities, a Bayesian estimator is obtained by endowing the mixture distribution with the Dirichlet process prior. Assuming this prior, we derive contraction rates of the posterior density at zero by carefully revising arguments presented in Salomond (*Electronic Journal of Statistics* **8** (2014) 1380–1404). Several choices of base measure are numerically evaluated and compared. In a simulation various frequentist methods and a Bayesian estimator are compared. Finally, the Bayesian procedure is applied to current durations data described in Slama et al. (*Human Reproduction* **27** (2012) 1489–1498).

References

Bezanson, J., Edelman, A., Karpinski, S. and Shah, V. (2017). Julia: A fresh approach to numerical computing. *SIAM Review* **59**, 65–98. MR3605826 https://doi.org/10.1137/141000671

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problem. *The Annals of Statistics* **1**, 209–230. MR0350949

Ghosal, S., Ghosh, J. K. and Van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics* **28**, 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

Grenander, U. (1956). On the theory of mortality measurement. II. *Skandinavisk Aktuarietidskrift* **39**, 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints*. Cambridge: Cambridge University Press. MR3445293 https://doi.org/10.1017/CBO9781139020893

Groeneboom, P. and Jongbloed, G. (2015). Nonparametric confidence intervals for monotone functions. *The Annals of Statistics* **43**, 2019–2054. MR3375875 https://doi.org/10.1214/15-AOS1335

Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *The Annals of Statistics* **34**, 742–768. MR2283391 https://doi.org/10.1214/009053606000000100

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265. MR1823804 https://doi.org/10.2307/1390653

Salomond, J. B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electronic Journal of Statistics* **8**, 1380–1404. MR3263126 https://doi.org/10.1214/14-EJS929

Slama, R., Højbjerg Hansen, O. K., Ducot, B., Bohet, A., Sorensen, D., Allemand, L., Eijkemans, M. J., Rosetta, L., Thalabard, J. C., Keiding, N., et al (2012). Estimation of the frequency of involuntary infertility on a nation-wide basis. *Human Reproduction* **27**, 1489–1498.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

Van der Vaart, A. W. and Ghosal, S. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics*. MR3587782 https://doi.org/10.1017/9781139029834

Key words and phrases. Bayesian nonparametrics, posterior consistency, contraction rate.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751–761. MR1041420 https://doi.org/10.1093/biomet/76.4.751

Watson, G. S. (1971). Estimating functionals of particle size distributions. *Biometrika* **58**, 483–490. MR0312639 https://doi.org/10.1093/biomet/58.3.483

Wehrhahn, C., Jara, A. and Barrientos, A. F. (2019). On the small sample behavior of dirichlet process mixture models for data supported on compact intervals. *Communications in Statistics Simulation and Computation* **50** (3), 786–810. https://doi.org/10.1080/03610918.2019.1568470

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. New York: Springer.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal* **23**, 189–207. MR0077581

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* **3**, 501–515. MR1243398

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 2, 421–434
https://doi.org/10.1214/20-BJPS483

A Bayesian nonparametric estimation to entropy

Luai Al-Labadi[*], Vishakh Patel[†], Kasra Vakiloroayaei[‡] and Clement Wan[§]

Department of Mathematical and Computational Sciences, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada, [*]luai.allabadi@utoronto.ca; [†]vishakh.patel@mail.utoronto.ca; [‡]k.vakiloroayaei@mail.utoronto.ca; [§]cmclement.wan@mail.utoronto.ca

Abstract. A Bayesian nonparametric estimator to entropy is proposed. The derivation of the new estimator relies on using the Dirichlet process and adapting the well-known frequentist estimators of Vasicek (*Journal of Royal Statistical Society B* **38** (1976) 54–59) and Ebrahimi, Pflughoeft and Soofi (*Statistics & Probability Letters* **20** (1994) 225–234). Several theoretical properties, such as consistency, of the proposed estimator are obtained. The quality of the proposed estimator has been investigated through several examples, in which it exhibits excellent performance.

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover. MR0415962

Alizadeh Noughabi, H. (2010). A new estimator of entropy and its application in testing normality. *Journal of Statistical Computation and Simulation* **80**, 1151–1162. MR2742527 https://doi.org/10.1080/00949650903005656

Alizadeh Noughabi, H. and Arghami, N. R. (2010). A new estimator of entropy. *Journal of the Iranian Statistical Society* **9**, 53–64. MR2771854

Al-Labadi, L. and Abdelrazeq, I. (2017). On functional central limit theorems of Bayesian nonparametric priors. *Statistical Methods & Applications* **26**, 215–229. MR3652494 https://doi.org/10.1007/s10260-016-0365-8

Al-Labadi, L. and Evans, M. (2018). Prior-based model checking. *Canadian Journal of Statistics* **46**, 380–398. MR3884427 https://doi.org/10.1002/cjs.11457

Al-Labadi, L. and Zarepour, M. (2014). On simulations from the two-parameter Poisson–Dirichlet process and the normalized inverse-Gaussian process. *Sankhyā A* **76**, 158–176. MR3167777 https://doi.org/10.1007/s13171-013-0033-0

Al-Omari, A. I. (2016). A new measure of entropy of continuous random variable. *Journal of Statistical Theory and Practice* **10**, 721–735. MR3558398 https://doi.org/10.1080/15598608.2016.1217444

Atkinson, C. M. (2017). Weak and strong laws of large numbers of negatively associated random variables. Ph.D. thesis, Univ. Regina. Available at https://ourspace.uregina.ca/bitstream/handle/10294/7844/Atkinson_Christopher_200324544_PHD_MATH_Fall2017.pdf. MR3799773 https://doi.org/10.1186/s13660-018-1699-6

Beirlant, J., Dudewicz, E. J., Györia, L. and van der Meulen, E. C. (1997). Nonparametric entropy estimation: An overview. *International Journal of Mathematics and Statistics* **6**, 17–39. MR1471870

Berrett, T. B., Grose, D. J. and Samworth, R. J. (2018). Package 'IndepTest': Nonparametric independence tests based on entropy estimation. Available at https://cran.r-project.org/web/packages/IndepTest/IndepTest.pdf.

Berrett, T. B., Samworth, R. J. and Yuan, M. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics* **47**, 288–318. MR3909934 https://doi.org/10.1214/18-AOS1688

Correa, J. C. (1995). A new estimator of entropy. *Communications in Statistics—Theory and Methods* **24**, 2439–2449. MR1354465 https://doi.org/10.1080/03610929508831626

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd ed. New York: John Wiley & Sons, Inc. MR1122806 https://doi.org/10.1002/0471200611

Ebrahimi, N., Pflughoeft, K. and Soofi, E. (1994). Two measures of sample entropy. *Statistics & Probability Letters* **20**, 225–234. MR1294108 https://doi.org/10.1016/0167-7152(94)90046-9

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230. MR0350949

Key words and phrases. Dirichlet process, Kullback–Leibler divergence, model checking.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. New York: Chapman & Hall/CRC. MR1699953 https://doi.org/10.1007/978-1-4899-4549-5

Grzegorzewski, P. and Wieczorkowski, R. (1999). Entropy-based goodness-of-fittest for exponentiality. *Communications in Statistics—Theory and Methods* **28**, 1183–1202. MR1704358 https://doi.org/10.1080/03610929908832351

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173. MR1952729 https://doi.org/10.1198/016214501750332758

Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics* **30**, 269–283. MR1926065 https://doi.org/10.2307/3315951

Kozachenko, L. F. and Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Problems of Information Transmission* **23**, 95–101. MR0908626

Mazzuchi, T. A., Soofi, E. S. and Soyer, R. (2008). Bayes estimate and inference for entropy and information index of fit. *Econometric Reviews* **27**, 428–456. MR2439863 https://doi.org/10.1080/07474930801960311

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423, 623–656. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society B* **38**, 54–59. MR0420958

van Es, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings. *Scandinavian Journal of Statistics* **19**, 61–72. MR1172967

Wieczorkowski, R. and Grzegorzewski, P. (1999). Entropy estimators—Improvements and comparisons. *Communications in Statistics—Simulation and Computation* **28**, 541–567. MR1704417 https://doi.org/10.1080/03610919908813564

Zarepour, M. and Al-Labadi, L. (2012). On a rapid simulation of the Dirichlet process. *Statistics & Probability Letters* **82**, 916–924. MR2910038 https://doi.org/10.1016/j.spl.2012.01.020

AIMS AND SCOPE

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

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

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

GENERAL INFORMATION

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

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

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

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

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

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

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

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