Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 35 • Number 3 • August 2021

Brazilian Journal of Probability and Statistics

Volume 35 • Number 3 • August 2021

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 35, Number 3, August 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).

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 435–441
https://doi.org/10.1214/20-BJPS484

Dirac distributions related to sums of independent nonidentically uniform random variables

Youssef Lazar[1,2] **and Bander Almutairi**[3]

[1]*University of East Anglia, Norwich, UK*
[2]*Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, KSA, ylazar77@gmail.com*
[3]*King Saud University, KSU, Riyadh, KSA, baalmutairi@ksu.edu.sa*

Abstract. The aim of this note is to give an elegant proof of a result due to E. G. Olds which concerns the density distribution of the sum of independent uniform random variables non-identically distributed. The proof uses both analytical and combinatorial properties of Dirac distributions and their convolutions. The method is new and can apply to other situations.

References

Bradley, D. M. and Gupta, R. C. (2002). On the distribution of the sum of n non-identically distributed uniform random variables. *Annals of the Institute of Statistical Mathematics* **54**, 689–700. MR1932412 https://doi.org/10.1023/A:1022483715767

Buonocore, A., Pirozzi, E. and Caputo, L. (2009). A note on the sum of uniform random variables. *Statistics & Probability Letters* **79**, 2092–2097. MR2571773 https://doi.org/10.1016/j.spl.2009.06.020

Duistermaat, J. J. and Kolk, J. A. C. (2010). *Distributions, Theory and Applications*. *Cornerstones*. New York: Springer. MR2680692 https://doi.org/10.1007/978-0-8176-4675-2

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

Hall, P. (1927). The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika* **19**, 240–245.

Hörmander, L. (1983). *The Analysis of Linear Partial Differential Operators I. Grundl. Math. Wissenschaft.* **256**. Berlin: Springer. MR0717035 https://doi.org/10.1007/978-3-642-96750-4

Irwin, J. O. (1927). On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's type II. *Biometrika* **19**, 225–239.

Mitra, S. K. (1971). On the probability distribution of the sums of uniformly distributed random variables. *SIAM Journal on Applied Mathematics* **20**, 195–198. MR0307326 https://doi.org/10.1137/0120026

Olds, E. G. (1952). A note on the convolution of uniform distributions. *The Annals of Mathematical Statistics* **23**, 282–285. MR0048750 https://doi.org/10.1214/aoms/1177729446

Roach, S. A. (1963). The frequency distribution of the sample mean where each member of the sample is drawn from a different rectangular distribution. *Biometrika* **50**, 508–513. MR0163381 https://doi.org/10.1093/biomet/50.3-4.508

Sadooghi-Alvandi, S. M., Nematollahi, A. R. and Habibi, R. (2009). On the distribution of the sums of independent uniform variables. *Statistical Papers* **50**, 171–175. MR2476176 https://doi.org/10.1007/s00362-007-0049-4

Schwartz, L. (1966). *Théorie des Distributions*. Paris: Hermann. MR0209834

Ueda, S., Okada, S., Sato, H. and Shimuzu, K. (1994). Distribution of the sum of uniform random variables with different ranges. *SUT Journal of Mathematics* **30**, 65–73. MR1280701

Key words and phrases. Sums of uniform distributions, convolutions, Dirac distributions.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 442–465
https://doi.org/10.1214/20-BJPS485

Geometric generated family of distributions: A review

Teena Goyal[1], **Sandeep Kumar Maurya**[2] **and Saralees Nadarajah**[3]

[1]*Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India, teenagoyalnewai@gmail.com*
[2]*Department of Statistics, Central University of South Bihar, Bihar, India, sandeepmaurya.maurya48@gmail.com*
[3]*School of Mathematics, University of Manchester, Manchester, UK, Saralees.Nadarajah@manchester.ac.uk*

Abstract. The present article represents a review of the geometric generated family of distributions. Based on this family of distribution, several distributions are proposed. The family can be proposed by using the compounding concept of zero truncated geometric distribution with any other model or family of distributions. Here, we provide a complete survey on this family of distributions and also listed the contributory related research work, their submodels, hazard rates, and utilized real datasets. We also address 10 power series distributions, 60 distributions based on the geometric family of distribution. These numbers show the importance of the geometric family of distribution.

References

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability* **36**, 106–108.

Abouammoh, A., Abdulghani, S. and Qamber, I. (1994). On partial orderings and testing of new better than renewal used classes. *Reliability Engineering and Systems Safety* **43**, 37–41. MR1428520

Adamidis, K., Dimitrakopoulou, T. and Loukas, S. (2005). On an extension of the exponential-geometric distribution. *Statistics and Probability Letters* **73**, 259–269. MR2179285 https://doi.org/10.1016/j.spl.2005.03.013

Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters* **39**, 35–42. MR1649319 https://doi.org/10.1016/S0167-7152(98)00012-1

Afify, A. Z., Nofal, Z. M. and Butt, N. S. (2014). Transmuted complementary Weibull geometric distribution. *Pakistan Journal of Statistics and Operation Research* **10**, 435–454. MR3302433 https://doi.org/10.18187/pjsor.v10i4.836

Alizadeh, M., Fazel Bagheri, S., Alizadeh, M. and Nadarajah, S. (2017). A new four-parameter lifetime distribution. *Journal of Applied Statistics* **44**, 767–797. MR3627598 https://doi.org/10.1080/02664763.2016.1182137

Alkarni, S. H. (2013). A compound class of geometric and lifetimes distributions. *The Open Statistics and Probability Journal* **5**, 1–5. MR3067639 https://doi.org/10.2174/1876527001305010001

Andrews, D. F. and Herzberg, A. M. (2012). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Berlin: Springer.

Ashour, S. K. and Wahed, M. L. A. (2014). Kummer beta-Weibull geometric distribution: A new generalizations of beta-Weibull geometric distribution. *International Journal of Sciences Basic and Applied Research* **16**, 258–273.

Barlow, R. E. and Campo, R. A. (1975). Total time on test processes and applications to failure data analysis. Technical report, California University Berkeley Operations Research Centre. MR0448771

Barreto-Souza, W., de Morais, A. L. and Cordeiro, G. M. (2011). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation* **81**, 645–657. MR2788571 https://doi.org/10.1080/00949650903436554

Bhaumik, D. K., Kapur, K. and Gibbons, R. D. (2009). Testing parameters of a gamma distribution for small samples. *Technometrics* **51**, 326–334. MR2751075 https://doi.org/10.1198/tech.2009.07038

Bidram, H. (2012). The beta exponential-geometric distribution. *Communications in Statistics Simulation and Computation* **41**, 1606–1622. MR2924006 https://doi.org/10.1080/03610918.2011.611309

Bidram, H., Behboodian, J. and Towhidi, M. (2013a). The beta Weibull-geometric distribution. *Journal of Statistical Computation and Simulation* **83**, 52–67. MR3005334 https://doi.org/10.1080/00949655.2011.603089

Key words and phrases. Lifetime models, geometric distribution, power series distribution, compounding technique, real data application.

Bidram, H., Behboodian, J. and Towhidi, M. (2013b). A new generalized exponential geometric distribution. *Communications in Statistics Theory and Methods* **42**, 528–542. MR3005794 https://doi.org/10.1080/03610926.2011.579703

Bidram, H. and Nadarajah, S. (2016). A new lifetime model with decreasing, increasing, bathtub-shaped, and upside-down bathtub-shaped hazard rate function. *Statistics* **50**, 139–156. MR3435173 https://doi.org/10.1080/02331888.2015.1006224

Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability* **6**, 328–347. MR0251807 https://doi.org/10.2307/3212004

Bjerkedal, T. (1960). Acquisition of resistance in Guinea pies infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene* **72**, 130–148.

Blundell, R., Duncan, A. and Pendakur, K. (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics* **13**, 435–461.

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

Bordbar, F. and Nematollahi, A. (2016). The modified exponential-geometric distribution. *Communications in Statistics Theory and Methods* **45**, 173–181. MR3440378 https://doi.org/10.1080/03610926.2013.827721

Bortolini, J., Pascoa, M. A., de Lima, R. R. and Oliveira, A. C. (2017). The extended generalized gamma geometric distribution. *International Journal of Statistics and Probability* **6**, 1–22.

Castellares, F. and Lemonte, A. J. (2016). On the Marshall–Olkin extended distributions. *Communications in Statistics Theory and Methods* **45**, 4537–4555. MR3509539 https://doi.org/10.1080/03610926.2014.922986

Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis* **53**, 4433–4440. MR2744336 https://doi.org/10.1016/j.csda.2009.06.016

Chung, Y. and Kang, Y. (2014). The exponentiated Weibull-geometric distribution: Properties and estimations. *Communications for Statistical Applications and Methods* **21**, 147–160.

Consul, P. C. and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics* **15**, 791–799. MR0391332 https://doi.org/10.1080/00401706.1973.10489112

Cordeiro, G. M., Silva, G. O. and Ortega, E. M. (2013). The beta- Weibull geometric distribution. *Statistics* **47**, 817–834. MR3175718 https://doi.org/10.1080/02331888.2011.577897

Cox, D. R. (1966). *The Statistical Analysis of Series of Events*. *Monographs on Applied Probability and Statistics*.

Dahl, A., Askim, T., Stock, R., Langørgen, E., Lydersen, S. and Indredavik, B. (2008). Short-and long-term outcome of constraint-induced movement therapy after stroke: A randomized controlled feasibility trial. *Clinical Rehabilitation* **22**, 436–447.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Berlin: Springer. MR1883764

Davis, D. (1952). An analysis of some failure data. *Journal of the American Statistical Association* **47**, 113–150.

Elbatal, I., Altun, E., Afify, A. Z. and Ozel, G. (2019). The generalized Burr xii power series distributions with properties and applications. *Annals of Data Science* **6**, 571–597. MR3827506 https://doi.org/10.31801/cfsuasmas.443699

Elbatal, I., El Gebaly, Y. M. and Amin, E. A. (2017). The beta generalized inverse Weibull geometric distribution. *Pakistan Journal of Statistics and Operation Research* **13**, 75–90. MR3637151 https://doi.org/10.18187/pjsor.v13i1.1791

Elbatal, I., Mansour, M. and Ahsanullah, M. (2016). The additive Weibull-geometric distribution: Theory and applications. *Journal of Statistical Theory and Applications* **15**, 125–141. MR3519344 https://doi.org/10.2991/jsta.2016.15.2.3

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics Theory and Methods* **31**, 497–512. MR1902307 https://doi.org/10.1081/STA-120003130

Fattah, A. A., Nadarajah, S. and Ahmed, A.-H. N. (2017). The exponentiated transmuted Weibull geometric distribution with application in survival analysis. *Communications in Statistics Simulation and Computation* **46**, 4244–4263. MR3672543 https://doi.org/10.1080/03610918.2015.1112909

Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. and Harrington, D. P. (1980). Modified Kolmogorov Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics* **36**, 607–625. MR0672339 https://doi.org/10.2307/2556114

Flores, J., Borges, P., Cancho, V. G. and Louzada, F. (2013). The complementary exponential power series distribution. *Brazilian Journal of Probability and Statistics* **27**, 565–584. MR3105044 https://doi.org/10.1214/11-bjps182

Fonseca, M. and Franca, M. (2007). A influência da fertilidade do solo e caracterizaçao da fixaçao biológica de n2 para o crescimento de dimorphandra wilsonii rizz. Master's Thesis, Universidade Federal de Minas Gerais.

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

Gitifar, N., Rezaei, S. and Nadarajah, S. (2016). Compound distributions motivated by linear failure rate. *Statistics and Operations Research Transactions* **40**, 177–200. MR3530371

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test* **19**, 399–415. MR2677735 https://doi.org/10.1007/s11749-009-0169-3

Gross, A. J. and Clark, V. A. (1975). Survival distributions: Reliability applications in the biomedical sciences. University Microfilm International.

Harandi, S. S. and Alamatsaz, M. (2017). A complementary generalized linear failure rate-geometric distribution. *Communications in Statistics Simulation and Computation* **46**, 679–703. MR3563522 https://doi.org/10.1080/03610918.2014.950744

Hassan, A. S. and Assar, S. M. (2019). A new class of power function distribution: Properties and applications. *Annals of Data Science*. https://doi.org/10.1007/s40745-019-00195-7

Hirose, H. (1993). Estimation of threshold stress in accelerated life testing. *IEEE Transactions on Reliability* **42**, 650–657.

Hoel, D. G. (1972). A representation of mortality data by competing risks. *Biometrics* **28**, 475–488.

Kemp, A. W. (2008). The discrete half-normal distribution. In *Advances in Mathematical and Statistical Modeling*, 353–360. Berlin: Springer. MR2790107 https://doi.org/10.1007/978-0-8176-4626-4_27

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

Kuş, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis* **51**, 4497–4509. MR2364461 https://doi.org/10.1016/j.csda.2006.07.017

Kwan, K., Breault, G., Umbenhauer, E., McMahon, F. and Duggan, D. (1976). Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man. *Journal of Pharmacokinetics and Biopharmaceutics* **4**, 255–280.

Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*, Vol. **362**. New York: Wiley. MR0640866

Lee, E. T. and Wang, J. (2003). *Statistical Methods for Survival Data Analysis*, Vol. **476**. New York: Wiley. MR1968483 https://doi.org/10.1002/0471458546

Louzada, F., Cancho, V. G., Roman, M. and Leite, J. G. (2012). A new long-term lifetime distribution induced by a latent complementary risk framework. *Journal of Applied Statistics* **39**, 2209–2222. MR2968019 https://doi.org/10.1080/02664763.2012.706264

Louzada, F., Marchi, V. and Carpenter, J. (2013). The complementary exponentiated exponential geometric lifetime distribution. *Journal of Probability and Statistics* **2013**, 1–13. MR3037955 https://doi.org/10.1155/2013/502159

Louzada, F., Marchi, V. and Roman, M. (2014). The exponentiated exponential geometric distribution: A distribution with decreasing, increasing and unimodal failure rate. *Statistics* **48**, 167–181. MR3175765 https://doi.org/10.1080/02331888.2012.667103

Louzada, F., Roman, M. and Cancho, V. G. (2011). The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart. *Computational Statistics and Data Analysis* **55**, 2516–2524. MR2787009 https://doi.org/10.1016/j.csda.2011.02.018

Maguire, B. A., Pearson, E. and Wynn, A. (1952). The time intervals between industrial accidents. *Biometrika* **39**, 168–180. MR0047978 https://doi.org/10.1093/biomet/39.1-2.130

Mahmoudi, E. and Jafari, A. A. (2012). Generalized exponential–power series distributions. *Computational Statistics and Data Analysis* **56**, 4047–4066. MR2957854 https://doi.org/10.1016/j.csda.2012.04.009

Mahmoudi, E. and Jafari, A. A. (2017). The compound class of linear failure rate-power series distributions: Model, properties, and applications. *Communications in Statistics Simulation and Computation* **46**, 1414–1440. MR3606890 https://doi.org/10.1080/03610918.2015.1005232

Mahmoudi, E. and Shiran, M. (2012). Exponentiated Weibull-geometric distribution and its applications. Preprint. Available at arXiv:1206.4008.

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

Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika* **84**, 641–652. MR1603936 https://doi.org/10.1093/biomet/84.3.641

Meeker, W. Q. and Escobar, L. A. (2014). *Statistical Methods for Reliability Data*. New York: Wiley.

Mendoza, N. V., Ortega, E. M. and Cordeiro, G. M. (2016). The exponentiated-log-logistic geometric distribution: Dual activation. *Communications in Statistics Theory and Methods* **45**, 3838–3859. MR3502183 https://doi.org/10.1080/03610926.2014.909937

Merovci, F. and Elbatal, I. (2014a). Transmuted Lindley-geometric distribution and its applications. *Journal of Statistics Applications and Probability* **3**, 77–91.

Merovci, F. and Elbatal, I. (2014b). Transmuted Weibull-geometric distribution and its applications. *Scientia Magna* **10**, 68–82.

Morais, A. L. and Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions. *Computational Statistics and Data Analysis* **55**, 1410–1425. MR2741424 https://doi.org/10.1016/j.csda.2010.09.030

Murthy, D. P., Xie, M. and Jiang, R. (2004). *Weibull Models*, Vol. **505**. New York: Wiley. MR2013269

Nadarajah, S., Cancho, V. G. and Ortega, E. M. (2013). The geometric exponential Poisson distribution. *Statistical Methods and Applications* **22**, 355–380. MR3082803 https://doi.org/10.1007/s10260-013-0230-y

Nadarajah, S., Cordeiro, G. M. and Ortega, E. M. (2015). The exponentiated g geometric family of distributions. *Journal of Statistical Computation and Simulation* **85**, 1634–1650. MR3315761 https://doi.org/10.1080/00949655.2014.885977

Nadarajah, S., Shahsanaei, F. and Rezaei, S. (2014). A new four-parameter lifetime distribution. *Journal of Statistical Computation and Simulation* **84**, 248–263. MR3169324 https://doi.org/10.1080/00949655.2012.705284

Nassar, M. and Nada, N. (2012). A new generalization of the exponential-geometric distribution. *Journal of Statistics: Advances in Theory and Applications* **7**, 25–48. MR3073053 https://doi.org/10.1016/j.joems.2013.01.003

Nichols, M. D. and Padgett, W. (2006). A bootstrap control chart for Weibull percentiles. *Quality and reliability engineering international* **22**, 141–151.

Nofal, Z. M., Afify, A. Z., Yousof, H. M. and Cordeiro, G. M. (2017). The generalized transmuted-g family of distributions. *Communications in Statistics Theory and Methods* **46**, 4119–4136. MR3590859 https://doi.org/10.1080/03610926.2015.1078478

Ortega, E., Cordeiro, G. and Pascoa, M. (2011). The generalized gamma geometric distribution. *Journal of Statistical Theory and Applications* **10**, 433–454. MR2868280

Perdona, G. d. S. C. and Louzada, N. F. (2006). Modelos de riscos aplicados à análise de sobrevivência.

Picciotto, R. and Hersh, S. (1972). The tensile fatigue behavior of a warp yarn and its influence on weaving performance. *Textile Research Journal* **42**, 512–522.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* **5**, 375–383.

Pulsoni, A., Pagano, L., Latagliata, R., Casini, M., Cerri, R., Crugnola, M., De Paoli, L., Di Bona, E., Invernizzi, R. and Marmont, F. (2004). Survival of elderly patients with acute myeloid leukemia. *Haematologica* **89**, 296–302.

Quesenberry, C. and Kent, J. (1982). Selecting among probability distributions used in reliability. *Technometrics* **24**, 59–65. MR0653113 https://doi.org/10.2307/1267579

Rahmouni, M. and Orabi, A. (2018). The exponential-generalized truncated geometric distribution: A new lifetime distribution. *International Journal of Statistics and Probability* **7**, 1–20.

Rezaei, S., Hashami, S. and Najjar, L. (2014). Extended exponential geometric proportional hazard model. *Annals of Data Science* **1**, 173–189.

Ristić, M. M. and Kundu, D. (2016). Generalized exponential geometric extreme distribution. *Journal of Statistical Theory and Practice* **10**, 179–201. MR3453035 https://doi.org/10.1080/15598608.2015.1101628

Saboor, A., Elbatal, I. and Cordeiro, G. M. (2016). The transmuted exponentiated Weibull geometric distribution: Theory and applications. *Hacettepe Journal of Mathematics and Statistics* **45**, 973–987. MR3588236 https://doi.org/10.15672/hjms.20156410830

Shahsanaei, F., Rezaei, S. and Pak, A. (2012). A new two-parameter lifetime distribution with increasing failure rate. *Stochastics and Quality Control* **27**, 1–17.

Silva, R. B., Barreto-Souza, W. and Cordeiro, G. M. (2010). A new distribution with decreasing, increasing and upside-down bathtub failure rate. *Computational Statistics and Data Analysis* **54**, 935–944. MR2580928 https://doi.org/10.1016/j.csda.2009.10.006

Silva, R. B., Bourguignon, M., Dias, C. R. and Cordeiro, G. M. (2013). The compound class of extended Weibull power series distributions. *Computational Statistics and Data Analysis* **58**, 352–367. MR2997948 https://doi.org/10.1016/j.csda.2012.09.009

Silva, R. B. and Cordeiro, G. M. (2015). The Burr XII power series distributions: A new compounding family. *Brazilian Journal of Probability and Statistics* **29**, 565–589. MR3355748 https://doi.org/10.1214/13-BJPS234

Smith, R. L. and Naylor, J. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society Series C* **36**, 358–369. MR0918854 https://doi.org/10.2307/2347795

Tahir, M. H. and Cordeiro, G. M. (2016). Compounding of distributions: A survey and new generalized classes. *Journal of Statistical Distributions and Applications* **3**, 1–35.

Tahir, M. H. and Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established g families. *Anais Da Academia Brasileira de Ciências* **87**, 539–568. MR3408221 https://doi.org/10.1590/0001-3765201520140299

Tojeiro, C., Louzada, F., Roman, M. and Borges, P. (2014). The complementary Weibull geometric distribution. *Journal of Statistical Computation and Simulation* **84**, 1345–1362. MR3169397 https://doi.org/10.1080/00949655.2012.744406

Wang, M. (2013). A new three-parameter lifetime distribution and associated inference. Preprint. Available at arXiv:1308.4128.

Wang, M. and Elbatal, I. (2015). The modified Weibull geometric distribution. *Metron* **73**, 303–315. MR3429801 https://doi.org/10.1007/s40300-014-0052-1

Zakerzadeh, H. and Mahmoudi, E. (2012). A new two parameter lifetime distribution: Model and properties. Preprint. Available at arXiv:1204.4248.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 466–483
https://doi.org/10.1214/20-BJPS486

A Monte Carlo integration approach to estimating drift and minorization coefficients for Metropolis–Hastings samplers

David A. Spade

Department of Mathematical Sciences, University of Wisconsin–Milwaukee, 3200 N. Cramer Street, EMS E403, Milwaukee, Wisconsin 53211, USA, spade@uwm.edu

Abstract. Bayesian statistical methodology has become highly popular in a myriad of applications over the past several decades. In Bayesian statistics, it is often required to draw samples from intractable probability distributions. Markov chain Monte Carlo (MCMC) algorithms are common methods of obtaining samples from these distributions. When an MCMC algorithm is used, it is important to be able to obtain an answer to the question of how many iterations the chain must run before it is "close enough" to its target distribution to allow approximate sampling from this distribution. Several methods of approaching this question exist in the literature. Some rely on the output of the chain, and some are based on Markov chain theory. These techniques suffer from major practical limitations. This work provides a computational method of bounding the mixing time of a Metropolis–Hastings algorithm. This approach extends the work of Spade (*Statistics and Computing* **26** (2016) 761–781) and Spade (*Markov Processes and Related Fields* **26** (2020) 487–516) to general versions of the Metropolis–Hastings algorithm, while examining the convergence behavior of such samplers under symmetric and asymmetric proposal densities.

References

Chen, R. (2002). Convergence analysis and comparisons of Markov chain Monte Carlo algorithms in digital communications. *IEEE Transactions on Signal Processing* **50**, 255–270.

Chib, S., Nardari, F. and Shephard, N. (1998). Markov chain Monte Carlo methods for generalized stochastic volatility models. *Journal of Econometrics* **108**, 281–316. MR1894758 https://doi.org/10.1016/S0304-4076(01)00137-3

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 883–904. MR1395755 https://doi.org/10.2307/2291683

Cowles, M. K. and Rosenthal, J. S. (1998). A simulation-based approach to convergence rates for Markov chain Monte Carlo algorithms. *Statistics and Computing* **8**, 115–124.

Eraker, B. (2001). MCMC analysis of diffusion models with applications to finance. *Journal of Business & Economic Statistics* **19**, 177–191. MR1939708 https://doi.org/10.1198/073500101316970403

Erkkila, T., Lehmusvaara, S., Ruusuvuori, P., Visakorpi, T., Shmulevich, I. and Lahdemakki, H. (2010). Probabilistic analysis of gene expression from heterogeneous tissues. *Bioinformatics* **26**, 2571–2577.

Fort, G., Moulines, E., Roberts, G. O. and Rosenthal, J. S. (2003). On the geometric ergodicity of hybrid samplers. *Journal of Applied Probability* **40**, 123–146. MR1953771 https://doi.org/10.1239/jap/1044476831

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B, Methodological* **56**, 501–514. MR1278223

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

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith, eds.) 169–193. Oxford, U.K.: Oxford University Press. MR1380276

Key words and phrases. Markov chain Monte Carlo, mixing time, geometric ergodicity, minorization.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operational Research* **31**, 1109–1144.

Ishwaran, H., James, L. F. and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association* **96**, 1316–1332. MR1946579 https://doi.org/10.1198/016214501753382255

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications* **35**, 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

Kamary, K., Mengersen, K. L., Robert, C. and Rousseau, J. (2014). Testing hypotheses as a mixture estimation model. Available at arXiv:1214.2044.

Koike, T. and Minami, M. (2019). Estimation of risk contributions with MCMC. *Quantitative Finance* **19**, 1579–1597. MR3991080 https://doi.org/10.1080/14697688.2019.1588469

Li, S., Pearl, D. K. and Doss, H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. *Journal of the American Statistical Association* **95**, 493–508.

Liang, F. (2007). Continuous contour Monte Carlo for marginal density estimation with an application to a spatial statistical model. *Journal of Computational and Graphical Statistics* **16**, 608–632. MR2351082 https://doi.org/10.1198/106186007X238459

Llorente, F., Martino, L., Delgado, D. and Lopez-Santiago, J. (2020). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. Available at arXiv:2005.08334v1.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics* **24**, 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201

Meyn, S. P. and Tweedie, R. L. (2005). *Markov Chains and Stochastic Stability*, 2nd ed. London: Springer. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

Neal, R. M. (1998). Annealed importance sampling. Technical Report, University of Toronto Department of Statistics.

Oh, M. and Berger, J. O. (1989). Adaptive importance sampling in Monte Carlo integration. Technical Report, Purdue University Department of Statistics. MR1117055 https://doi.org/10.1090/conm/115/1117055

Raftery, A. E. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith, eds.) 763–773. Oxford, U.K.: Oxford University Press.

Roberts, G. O. (1992). Convergence diagnostics of the Gibbs sampler. In *Bayesian Statistics 4* (J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith, eds.) 775–782. Oxford, U.K.: Oxford University Press. MR1380305

Roberts, G. O. (1996). Methods for estimating L^2 convergence of Markov chain Monte Carlo. In *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner* (D. Berry, I. Chaloner and J. Geweke, eds.) 373–384. Amsterdam: North-Holland. MR1367306

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability* **2**. Paper 2. MR1448322 https://doi.org/10.1214/ECP.v2-981

Roberts, G. O. and Rosenthal, J. S. (1998). Two convergence properties of hybrid samplers. *The Annals of Applied Probability* **8**, 397–407. MR1624941 https://doi.org/10.1214/aoap/1028903533

Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society, Series B, Methodological* **61**, 643–660. MR1707866 https://doi.org/10.1111/1467-9868.00198

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558–566. MR1340509

Rosenthal, J. S. (2001). A review of asymptotic convergence for general state space Markov chains. *Far East Journal of Theoretical Statistics* **5**, 37–50. MR1848443

Spade, D. A. (2016). A computational procedure for efficient estimation of the mixing time of a random-scan Metropolis algorithm. *Statistics and Computing* **26**, 761–781. MR3515020 https://doi.org/10.1007/s11222-015-9568-3

Spade, D. A. (2020). A computational approach to bounding the mixing time of a Metropolis–Hastings sampler. *Markov Processes and Related Fields* **26**, 487–516.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.

Yang, Z. and Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9264–9269.

Yu, B. (1994). Monitoring the convergence of Markov samplers based on estimated L^1 error. Technical Report 409, University of California at Berkeley Department of Statistics.

Yu, B. and Mykland, P. (1994). Looking at Markov samplers through CUSUM path plots: A simple diagnostic idea. Technical Report 413, University of California at Berkeley Department of Statistics.

Zellner, A. and Min, C. K. (1995). Gibbs sampler convergence criteria. *Journal of the American Statistical Association* **90**, 921–927.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 484–498
https://doi.org/10.1214/20-BJPS491

Inferring association from reliability functions: An approach based on copulas

B. Vineshkumar[1] **and N. Unnikrishnan Nair**[2]

[1]*Department of Statistics, Government Arts College, Thiruvananthapuram-695014, India, Vinesh910@gmail.com*
[2]*Department of Statistics, Cochin University of Science and Technology, Cochin-682022, India, unnikrishnannair4@gmail.com*

Abstract. Nair, Sankaran and John (*Metron* **76** (2018) 133–153) have defined and studied the properties of reliability functions in terms of copulas. In the present paper, we investigate the utility of such functions in inferring the time-dependent association of bivariate distributions. We consider the Clayton measure of association for the study. A general expression for this measure in terms of the generator of Archimedean copulas is given, and a method of finding nature of association using the generators is provided. We derive the relationship of the association measure with the ageing property of the distribution, associated with the generator. We analyze how the hazard rate of survival copulas can be utilized in studying the association between two random variables. Applications of the results in real life situations are discussed.

References

Anderson, J. E., Louis, T. A., Holm, N. V. and Harvald, B. (1992). Time dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association* **87**, 641–650. MR1185189

Averous, J. and Dortet-Bernadet, J. (2004). Dependence for Archimedean copulas and aging properties of their generating functions. *Sankhya* **66**, 607–620. MR2205812

Bairamov, I., Kotz, S. and Kozubowski, T. (2003). A new measure of linear local dependence. *Statistics* **37**, 243–258. MR1986179 https://doi.org/10.1080/0233188021000011839

Bjerve, S. and Doksum, K. A. (1993). Correlation curves: Measures of association as functions of covariate values. *The Annals of Statistics* **21**, 890–902. MR1232524 https://doi.org/10.1214/aos/1176349156

Chen, S. X. and Huang, T. (2007). Estimation of copula functions for dependence modelling. *Canadian Journal of Statistics* **35**, 1–18. MR2393609 https://doi.org/10.1002/cjs.5550350205

Clayton, D. G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151. MR0501698 https://doi.org/10.1093/biomet/65.1.141

Csorga, S. and Welsh, A. (1989). Testing for exponential and Marshall–Olkin distributions. *Journal of Statistical Planning and Inference* **23**, 287–300. MR1032686 https://doi.org/10.1016/0378-3758(89)90073-6

Cuevas, J. R. T., Yela, J. and Achcar, J. A. (2019). A method to select bivariate copula functions. *Revista Colombiana de Estadística* **42**, 61–80. MR3919838 https://doi.org/10.15446/rce.v42n1.71780

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* **88**, 1034–1043. MR1242947

Gupta, R. C. (2003). On some association measures in bivariate distributions and their relationships. *Journal of Statistical Planning and Inference* **117**, 83–98. MR2001143 https://doi.org/10.1016/S0378-3758(02)00367-1

John, P., Sankaran, P. G. and Nair, N. U. (2017). Some families of bivariate distributions and their applications. Doctoral dissertation. Available at http://dyuthi.cusat.ac.in/purl/5246.

Kochar, S. C. and Wiens, D. P. (1987). Partial orderings of life distributions with respect to their aging properties. *Naval Research Logistics* **34**, 823–829. MR0913467 https://doi.org/10.1002/1520-6750(198712)34:6<823::AID-NAV3220340607>3.0.CO;2-R

Nair, N. U. and Sankaran, P. G. (2010). A new measure of association for bivariate survival data. *Journal of Statistical Planning and Inference* **140**, 2569–2581. MR2644078 https://doi.org/10.1016/j.jspi.2010.03.019

Nair, N. U. and Sankaran, P. G. (2014). A family of bivariate exponential distributions and their copulas. *Sankhya Series B* **76**, 1–18. MR3194305 https://doi.org/10.1007/s13571-013-0067-2

Key words and phrases. Archimedean copula, Clayton measure, hazard rate of survival copulas.

Nair, N. U., Sankaran, P. G. and John, P. (2018). Modelling bivariate lifetime data using copula. *Metron* **76**, 133–153. MR3828945 https://doi.org/10.1007/s40300-018-0135-5

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed. New York: Springer. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487–493. MR1010337

Shih, J. and Emura, T. (2019). Bivariate dependence measures and bivariate competing risks models under the generalized FGM copula. *Statistical Papers* **60**, 1101–1118. MR4008679 https://doi.org/10.1007/s00362-016-0865-5

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 499–522
https://doi.org/10.1214/20-BJPS492

A Central Limit Theorem for incomplete U-statistics over triangular arrays

Matthias Löwe* and Sara Terveer†

Fachbereich Mathematik und Informatik, Universität Münster, Einsteinstraße 62, 48149 Münster, Germany,
**maloewe@uni-muenster.de;* †*sara.terveer@uni-muenster.de*

Abstract. We analyze the fluctuations of incomplete U-statistics over a triangular array of independent random variables. We give criteria for a Central Limit Theorem (CLT, for short) to hold in the sense that we prove that an appropriately scaled and centered version of the U-statistic converges to a normal random variable. Our method of proof relies on a martingale CLT. An application, a CLT for the hitting time for random walks on random graphs, will be presented in Löwe and Terveer (2020).

References

Bentkus, V. and Götze, F. (1995). On minimal moment assumptions in Berry–Esséen theorems for U-statistics. *Teoriâ Veroâtnostej I Ee Primeneniâ* **40**, 596–614. MR1401988 https://doi.org/10.1137/1140048

Bickel, P. J., Götze, F. and van Zwet, W. R. (1986). The Edgeworth expansion for U-statistics of degree two. *The Annals of Statistics* **14**, 1463–1484. MR0868312 https://doi.org/10.1214/aos/1176350170

Christofides, T. C. (1992). A strong law of large numbers for U-statistics. *Journal of Statistical Planning and Inference* **31**, 133–145. MR1173541 https://doi.org/10.1016/0378-3758(92)90025-N

Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*. *Advanced Lectures in Mathematics*. Braunschweig: Friedr. Vieweg & Sohn. MR0889896 https://doi.org/10.1007/978-3-663-14229-4

Eichelsbacher, P. (1998). Moderate and large deviations for U-processes. *Stochastic Processes and Their Applications* **74**, 273–296. MR1624408 https://doi.org/10.1016/S0304-4149(97)00113-0

Eichelsbacher, P. and Löwe, M. (1995). A large deviation principle for m-variate von Mises-statistics and U-statistics. *Journal of Theoretical Probability* **8**, 807–824. MR1353555 https://doi.org/10.1007/BF02410113

Eichelsbacher, P. and Löwe, M. (1998). Large deviations principle for partial sums U-processes. *Teoriâ Veroâtnostej I Ee Primeneniâ* **43**, 97–115. MR1669988 https://doi.org/10.1137/S0040585X97976647

Girko, V. L. (1990). *Theory of Random Determinants*. *Mathematics and Its Applications* **45**. Dordrecht: Kluwer Academic Publishers. MR1080966 https://doi.org/10.1007/978-94-009-1858-0

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

Jammalamadaka, S. R. and Janson, S. (1986). Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. *Annals of Probability* **14**, 1347–1358. MR0866355

Janson, S. (1984). The asymptotic distributions of incomplete U-statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **66**, 495–505. MR0753810 https://doi.org/10.1007/BF00531887

Janson, S. (1990). A functional limit theorem for random graphs with applications to subgraph count statistics. *Random Structures & Algorithms* **1**, 15–37. MR1068489 https://doi.org/10.1002/rsa.3240010103

Lee, A. J. (1990). *U-Statistics: Theory and Practice*. *Statistics: Textbooks and Monographs* **110**. New York: Marcel Dekker, Inc. MR1075417

Löwe, M. and Terveer, S. (2020). A central limit theorem for the mean starting hitting time for a random walk on a random graph. Preprint. Available at arXiv:2003.13011.

Malevich, T. L. and Abdurakhmanov, G. R. (1987). On conditions for the asymptotic normality of U-statistics in the case of triangular arrays. *Izvestiya Akademii Nauk UzSSR. Seriya Fiziko-Matematicheskikh Nauk. UzSSR Fanlar Akademiyasining Akhboroti. Fizika-Matematika Fanlari Seriyasi* **2**, 78.

Tan, Z. (2013). Precise asymptotics for products of sums and U-statistics. *Brazilian Journal of Probability and Statistics* **27**, 20–30. MR2991776 https://doi.org/10.1214/11-BJPS146

Key words and phrases. Central Limit Theorem, U-statistics, incomplete U-statistics, triangular arrays.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 523–543
https://doi.org/10.1214/20-BJPS493

Testing of Poisson mean with under-reported counts

Debjit Sengupta[1,*], **Surupa Roy**[1] **and Tathagata Banerjee**[2]

[1]*Department of Statistics, St. Xavier's College, 30, Mother Teresa Sarani, Kolkata-700016, India,*
[*]*debjits10@gmail.com*

[2]*Indian Institute of Management, Vastrapur, Ahmedabad 380015, India*

Abstract. For modelling unbounded count data, Poisson distribution is a natural choice. However, count data arising in various fields of scientific research are often under-reported. In such situations, inference carried out on the basis of Poisson model will result in biased parameter estimates and suboptimal tests. A modified Poisson model is developed to accommodate the possible undercount. For model-identifiability a double sampling scheme of data collection has been adopted. The focus of this paper is to develop asymptotically optimal tests for the Poisson mean in presence of undercount. Simulation study is conducted to compare the performance of the tests with respect to level and power and also to investigate the impact of ignoring undercount on each of the tests. The findings are validated using real life data.

References

Anderson, C., Bratcher, T. and Kutran, K. (1994). Bayesian-estimation of population-density and visibility. *Texas Journal of Science* **46**, 1–12.

Bratcher, T. L. and Stamey, J. D. (2002). Estimation of Poisson rates with misclassified counts. *Biometrical Journal* **44**, 946–956. MR1963380 https://doi.org/10.1002/bimj.200290006

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. CRC Press. MR2682774 https://doi.org/10.1201/9781420066586

Chen, Z., Yi, G. Y. and Wu, C. (2011). Marginal methods for correlated binary data with misclassified responses. *Biometrika* **98**, 647–662. MR2836412 https://doi.org/10.1093/biomet/asr035

Claeskens, G., Nguti, R. and Janssen, P. (2008). One-sided tests in shared frailty models. *Test* **17**, 69–82. MR2393352 https://doi.org/10.1007/s11749-006-0023-9

de Oliveira Lopes, G., Loschi, H. R. and Assunção, R. M. (2017). A random-censoring Poisson model for under-reported data. *Statistics in Medicine* **36**, 4873–4892. MR3734480 https://doi.org/10.1002/sim.7456

Dvorzak, M. and Wagner, H. (2016). Sparse Bayesian modelling of underreported count data. *Statistical Modelling* **16**, 24–46. MR3457686 https://doi.org/10.1177/1471082X15588398

Eo, Y. and Morley, J. (2015). Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics* **6**, 463–497. MR3388156 https://doi.org/10.3982/QE186

Fader, P. S. and Hardie, B. G. S. (2000). A note on modelling underreported Poisson counts. *Journal of Applied Statistics* **27**, 953–964.

Fernández-Fontelo, A., Cabaña, A., Joe, H., Puig, P. and Moriña, D. (2019). Untangling serially dependent underreported count data for gender-based violence. *Statistics in Medicine* **38**, 4404–4422. MR4002439 https://doi.org/10.1002/sim.8306

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics* **9**, 861–869. MR0619289

Gravel, C. A. and Platt, R. W. (2018). Weighted estimation for confounded binary outcomes subject to misclassification. *Statistics in Medicine* **37**, 425–436. MR3748750 https://doi.org/10.1002/sim.7522

Gustafson, P. (2003). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. CRC Press. MR2005104

Hochberg, Y. (1977). On the use of double sampling schemes in analyzing categorical data with misclassification errors. *Journal of the American Statistical Association* **72** (360a), 914–921. MR0471129

Johns, G. (1994). How often were you absent? A review of the use of self-reported absence data.

Kircher, T., Nelson, J. and Burdo, H. (1985). The autopsy as a measure of accuracy of the death certificate. *The New England Journal of Medicine* **313**, 1263–1269. PMID: 4058507.

Key words and phrases. Count data, under-reporting, validation data, maximum likelihood, optimal tests.

Meyer, B. D. and Mittag, N. (2017). Misclassification in binary choice models. *Journal of Econometrics* **200**, 295–311. MR3684980 https://doi.org/10.1016/j.jeconom.2017.06.012

Poon, W. Y. and Wang, H. B. (2010). Bayesian analysis of multivariate probit models with surrogate outcome data. *Psychometrika* **75**, 498–520. MR2719940 https://doi.org/10.1007/s11336-010-9164-6

Sengupta, D., Roy, S. and Banerjee, T. (2020). Supplement to "Testing of Poisson mean with under-reported counts." https://doi.org/10.1214/20-BJPS493SUPP

Shu, D. and Yi, G. Y. (2020). Causal inference with noisy data: Bias analysis and estimation approaches to simultaneously addressing missingness and misclassification in binary outcomes. *Statistics in Medicine* **39**, 456–468. MR4052822 https://doi.org/10.1002/sim.8419

Sposto, R., Preston, D. L., Shimizu, Y. and Mabuchi, K. (1992). The effect of diagnostic misclassification on non-cancer and cancer mortality dose response in a-bomb survivors. *Biometrics* **48**, 605–617.

Stamey, J. D. and Young, D. M. (2005). Maximum likelihood estimation for a Poisson rate parameter with misclassified counts. *Australian & New Zealand Journal of Statistics* **47**, 163–172. MR2155117 https://doi.org/10.1111/j.1467-842X.2005.00381.x

Stamey, J. D., Young, D. M. and Boese, D. (2006). A Bayesian hierarchical model for Poisson rate and reporting-probability inference using double sampling. *Australian & New Zealand Journal of Statistics* **48**, 201–212. MR2253916 https://doi.org/10.1111/j.1467-842X.2006.00434.x

Stamey, J. D., Young, D. M. and Bratcher, T. L. (2004). Bayesian predictive probability functions for count data that are subject to misclassification. *Biometrical Journal* **46**, 572–578. MR2101145 https://doi.org/10.1002/bimj.200410059

Stamey, J. D., Young, D. M. and Cecchini, M. (2003). A double-sampling approach for maximum likelihood estimation for a Poisson rate parameter with visibility-biased data. *Statistica* **63**, 3–11. MR2109748

Stamey, J. D., Young, D. M. and Seaman, J. W. (2008). A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Statistics in Medicine* **27**, 2440–2452. MR2432498 https://doi.org/10.1002/sim.3134

Stoner, O., Economou, T. and da Silva, G. D. M. (2019). A hierarchical framework for correcting under-reporting in count data. *Journal of the American Statistical Association* **114**, 1481–1492. MR4047275 https://doi.org/10.1080/01621459.2019.1573732

Wu, W., Stamey, J. and Kahle, D. (2015). A Bayesian approach to account for misclassification and overdispersion in count data. *International Journal of Environmental Research and Public Health* **12**, 10648–10661.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 544–568
https://doi.org/10.1214/20-BJPS496

Nonparametric Bayesian estimation of a concave distribution function with mixed interval censored data

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

*Delft 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; ‡lixuepang3@gmail.com*

Abstract. Assume we observe a finite number of inspection times together with information on whether a specific event has occurred before each of these times. Suppose replicated measurements are available on multiple event times. The set of inspection times, including the number of inspections, may be different for each event. This is known as mixed case interval censored data. We consider Bayesian estimation of the distribution function of the event time while assuming it is concave. We provide sufficient conditions on the prior such that the resulting procedure is consistent from the Bayesian point of view. We also provide computational methods for drawing from the posterior and illustrate the performance of the Bayesian method in both a simulation study and two real datasets.

References

Calle, M. L. and Gómez, G. (2001). Nonparametric Bayesian estimation from interval-censored data using Monte Carlo methods. *Journal of Statistical Planning and Inference* **98**, 73–87. MR1860226 https://doi.org/10.1016/S0378-3758(00)00320-7

Doss, H. and Huffer, F. W. (2003). Monte Carlo methods for Bayesian analysis of survival data using mixtures of Dirichlet process priors. *Journal of Computational and Graphical Statistics* **12**, 282–307. MR1983156 https://doi.org/10.1198/1061860031680

Dümbgen, L., Freitag, S. and Jongbloed, G. (2004). Consistency of concave regression, with an application to current status data. *Mathematical Methods of Statistics* **13**, 69–81. MR2078313

Dümbgen, L., Freitag, S. and Jongbloed, G. (2006). Estimating a unimodal distribution from interval-censored data. *Journal of the American Statistical Association* **101**, 1094–1106. MR2324149 https://doi.org/10.1198/016214506000000032

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications. Vol. II*. New York: John Wiley and Sons. MR0210154

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

Finkelstein, D. M. and Wolfe, R. A. (1986). Isotonic regression for interval censored survival data using an E–M algorithm. *Communications in Statistics Theory and Methods* **15**, 2493–2505. MR0853025 https://doi.org/10.1080/03610928608829264

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge: Cambridge University Press. MR3587782 https://doi.org/10.1017/9781139029834

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics. Springer Series in Statistics*. MR1992245

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

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel: Birkhäuser. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

Hansen, M. B. and Lauritzen, S. L. (2002). Nonparametric Bayes inference for concave distribution functions. *Statistica Neerlandica* **56**, 110–127. MR1894231 https://doi.org/10.1111/1467-9574.04600

Jara, A., Hanson, T., Quintana, F., Müller, P. and Rosner, G. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software, Articles* **40**, 1–30.

Jongbloed, G., van der Meulen, F. and Pang, L. (2020). Bayesian estimation of a decreasing density. Available at arXiv:1801.02539v6.

Key words and phrases. Bayesian nonparametrics, Dirichlet process, Markov Chain Monte Carlo, posterior consistency, shape constrained inference.

Keiding, N., Begtrup, K., Scheike, T. H. and Hasibeder, G. (1996). Estimation from Current Status Data in Continuous Time. *Lifetime Data Analysis* **2**, 119–129.

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

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Haywood, CA: IMS. MR1089429

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics* **27**, 45–55. MR1774042 https://doi.org/10.1111/1467-9469.00177

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association* **71**, 897–902. MR0436445

van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* **92**, 945–959. MR1482125 https://doi.org/10.2307/2965558

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics* **28**, 779–814. MR1792787 https://doi.org/10.1214/aos/1015951998

Wu, Y. and Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics* **28**, 298–331. MR2399197 https://doi.org/10.1214/07-EJS130

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 569–589
https://doi.org/10.1214/21-BJPS498

Inference in a linear functional relationship with replications

Julio Hokama[1], Pedro Morettin[2,*], Heleno Bolfarine[2,†] and Manuel Galea[3]

[1] *Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro, Brazil, jhokama41@gmail.com*
[2] *Universidade de São Paulo, São Paulo, Brazil, [*]pamorettin@gmail.com; [†]hbolfar08@gmail.com*
[3] *Pontificia Universidad Católica de Chile, Santiago, Chile, mgalea@mat.uc.cl*

Abstract. In this paper, we consider a model for data analysis with measurement errors. The main objective of this work is to develop statistical inference tools, such as parameter estimation and hypothesis tests in a linear functional relationship with replicated observations. For this purpose, we use the maximum likelihood method in the presence of incidental parameters, and the unbiased estimating equations approach. Both approaches lead to explicit expressions for the asymptotic covariance matrices of the estimators of the model parameters. A simulation study is performed to assess the empirical behavior of estimators and of a Wald statistic. The methodology is illustrated with a real data set.

References

Arellano-Valle, R., Bolfarine, H. and Vilca-Labra, F. (1996). Ultrastructural elliptical models. *Canadian Journal of Statistics* **24**, 207–216. MR1406176 https://doi.org/10.2307/3315626

Bolfarine, H. and Galea-Rojas, M. (1995). Comments on "Functional comparative calibration using the EM-algorithm" (by D. Kimura). *Biometrics* **51**, 1579–1580.

Boos, D. D. and Stefanski, L. A. (2013). *Essential Statistical Inference, Theory and Methods*. New York: Springer. MR3024617 https://doi.org/10.1007/978-1-4614-4818-1

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods and Applications*. London: CRC Press. MR2682774 https://doi.org/10.1201/9781420066586

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. London: Chapman & Hall. MR2243417 https://doi.org/10.1201/9781420010138

Carstensen, B. (2010). *Comparing Clinical Measurement Methods: A Practical Guide*. Chichester: John Wiley.

Chan, L. K. and Mak, T. K. (1979). Maximum likelihood estimation of a linear structural relationship with replication. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **41**, 263–268. MR0547252

Cheng, C. and Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics* **48**, 511–519. MR2328620 https://doi.org/10.1198/004017006000000237

Cheng, C. and van Ness, J. (1991). On the unreplicated ultrastructural model. *Biometrika* **78**, 442–445. MR1131179 https://doi.org/10.1093/biomet/78.2.442

Cheng, C. L. and van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. London: Arnold. MR1719513

de Castro, M. and Galea, M. (2010). Robust inference in an heteroscedastic measurement error model. *Journal of the Korean Statistical Society* **39**, 439–447. MR2780215 https://doi.org/10.1016/j.jkss.2009.09.003

Demidenko, E. and Stukel, T. A. (2005). Influence analysis for linear mixed-effects models. *Statistics in Medicine* **24**, 893–909. MR2134545 https://doi.org/10.1002/sim.1974

Dorff, M. and Gurland, J. (1961). Estimation of the parameters of a linear functional relation. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **23**, 160–170. MR0124110

Everitt, B. S. (2005). *An* R *and* S-PLUS *Companion to Multivariate Analysis*. New York: Springer. MR2287653 https://doi.org/10.1007/b138954

Fuller, W. A. (1995). Estimation in the presence of measurement error. *International Statistical Review* **63**, 121–147.

Fuller, W. A. (2006). *Measurement Error Models*. New York: John Wiley. MR2301581

Key words and phrases. Linear functional relationship, replicated observations, maximum likelihood estimation, unbiased estimating equations, asymptotic normality, Wald statistic.

Galea, M., Bolfarine, H. and Giménez, P. (2006). Hypothesis testing in functional comparative calibration models. *Communications in Statistics Theory and Methods* **35**, 2021–2033. MR2338914 https://doi.org/10.1080/03610920600761980

Galea, M., de Castilho, M., Bolfarine, H. and de Castro (2003). Detection of analytical bias. *The Analyst* **128**, 1073–1081.

Galea, M. and de Castro, M. (2017). Robust inference in a linear functional model with replications using the t distribution. *Journal of Multivariate Analysis* **160**, 134–145. MR3688694 https://doi.org/10.1016/j.jmva.2017.06.008

Giménez, P. C. and Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models. *Australian Journal of Statistics* **39**, 325–344. MR1616015 https://doi.org/10.1111/j.1467-842X.1997.tb00695.x

Giménez, P. C. and Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. *Journal of Multivariate Analysis* **114**, 1–15. MR2993869 https://doi.org/10.1016/j.jmva.2012.07.002

Giménez, P. C. and Patat, M. L. (2014). Local influence for functional comparative calibration models with replicated data. *Statistical Papers* **55**, 431–454. MR3188412 https://doi.org/10.1007/s00362-012-0489-3

Gleser, L. J. (1981). Estimation in a multivariate error-in-variables regression model: Large samples results. *The Annals of Statistics* **9**, 24–44. MR0600530

Gleser, L. J. (1985). A note on G.R. Dolby's unreplicated ultrastructural error in variables model. *Biometrika* **72**, 117–124. MR0790206 https://doi.org/10.1093/biomet/72.1.117

Hokama, J. (2001). Estimation and precision in the linear regression with errors in variables model and replicate measurements. PhD Thesis, Universidade de São Paulo, Brazil. (in Portuguese).

Kimura, D. (1992). Functional comparative calibration using EM-algorithm. *Biometrics* **48**, 1263–1271.

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

Mak, T. K. (1982). Estimation in the presence of incidental parameters. *Canadian Journal of Statistics* **10**, 121–132. MR0673672 https://doi.org/10.2307/3314904

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32. MR0025113 https://doi.org/10.2307/1914288

Patefield, W. M. (1977). On the information matrix in the linear functional relationship problem. *Applied Statistics* **26**, 69–70. MR0652707 https://doi.org/10.2307/2346870

Rasekh, A. R. and Fieller, N. R. J. (2003). Influence functions in functional measurement error models with replicated data. *Statistics* **37**, 169–178. MR1979675 https://doi.org/10.1080/0233188031000112890

Ripley, B. D. and Thompson, M. (1987). Regression techniques for the detection of analytical bias. *The Analyst* **112**, 377–383.

Riquelme, M., Bolfarine, H. and Galea, M. (2015). Robust linear functional mixed models. *Journal of Multivariate Analysis* **134**, 82–98. MR3296036 https://doi.org/10.1016/j.jmva.2014.10.008

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. New York: John Wiley. MR0986070 https://doi.org/10.1002/0471725315

Solari, M. E. (1969). The maximum likelihood solution of the problem of estimating a linear functional relationship. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **31**, 372–375.

Stevens, N. T., Steiner, S. H. and MacKay, R. J. (2017). Assessing agreement between two measurement systems: An alternative to the limits of agreement approach. *Statistical Methods in Medical Research* **26**, 2487–2504. MR3738264 https://doi.org/10.1177/0962280215601133

Vilca-Labra, F., Arellano-Valle, R. and Bolfarine, H. (1998). Elliptical functional models. *Journal of Multivariate Analysis* **65**, 36–57. MR1620785 https://doi.org/10.1006/jmva.1997.1726

Wu, L. (2010). *Mixed Effects Models for Complex Data*. London: Chapman & Hall. MR2598844 https://doi.org/10.1201/9781420074086-b2

Yi, G. Y. (2017). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. New York: Springer. MR3676914 https://doi.org/10.1007/978-1-4939-6640-0

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 590–608
https://doi.org/10.1214/21-BJPS499

Behavior of the Fréchet mean and Central Limit Theorems on spheres

Do Tran

Institute for Mathematical Stochastics, University of Gïttingen, Goldschmidtstrasse 7, 37077 Göttingen, Germany, do.tranvan@uni-goettingen.de

Abstract. We compute higher derivatives of the Fréchet function on spheres with an absolutely continuous and rotationally symmetric probability distribution. Consequences include (i) a practical condition to test if the mode of the symmetric distribution is a local Fréchet mean; (ii) a central limit theorem on spheres with practical assumptions and an explicit limiting distribution; and (iii) an answer to the question of whether the smeary effect can occur on spheres with absolutely continuous and rotationally symmetric distributions: with the method presented here, it can in dimension at least 4.

References

Afsari, B. (2009). Means and averaging on Riemannian manifolds. Doctoral dissertation, University of Maryland. MR2718072

Barden, D. and Le, H. (2014). On the measure of the cut locus of a Fréchet mean. *Bulletin of the London Mathematical Society*, 323–352. MR3239607 https://doi.org/10.1112/blms/bdu025

Bhattacharya, A. and Bhattacharya, R. (2008). Statistics on Riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society* **136**, 2959–2967. MR2399064 https://doi.org/10.1090/S0002-9939-08-09445-8

Bhattacharya, R. and Lin, L. (2017). Omnibus CLT for Fréchet means and nonparametric inference on non-euclidean spaces. *Proceedings of the American Mathematical Society* **145**, 413–428. MR3565392 https://doi.org/10.1090/proc/13216

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Annals of Statistics*, 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

Carmo d., M. P. (1992). *Riemannian geometry*. Basel: Birkhäuser. MR1138207 https://doi.org/10.1007/978-1-4757-2201-7

Eltzner, B. (2019). Measure Dependent Asymptotic Rate of the Mean: Geometrical and Topological Smeariness. Preprint. Available at arXiv:1908.04233.

Eltzner, B., Galaz-Garcia, F., Huckemann, S. F. and Tuschmann, W. (2019). Stability of the Cut Locus and a Central Limit Theorem for Fréchet Means of Riemannian Manifolds. Preprint. Available at arXiv:1909.00410. MR2981621 https://doi.org/10.1007/s10463-012-0352-2

Eltzner, B. and Huckemann, S. F. (2019). A Smeary Central Limit Theorem for Manifolds with Application to High Dimensional Spheres. *The Annals of Statistics* **47**, 3360–3381. MR4025745 https://doi.org/10.1214/18-AOS1781

Eltzner, B., Hundrieser, S. and Huckemann, S. F. (2020). Finite Sample Smeariness of Fréchet Means and Application to Climate. Preprint. Available at arXiv:2005.02321.

Hillar, C. J. and Lim, L. H. (2013). Most tensor problems are NP-hard. *Journal of the ACM* **60**, 1–39. MR3144915 https://doi.org/10.1145/2512329

Hotz, T. and Huckemann, S. F. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics* **67**, 177–193. MR3297863 https://doi.org/10.1007/s10463-013-0444-7

Robeva, E. (2016). Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications* **37**, 86–102. MR3530392 https://doi.org/10.1137/140989340

Tran, D. (2020). Sampling from stratified spaces. Doctoral dissertation, Duke University. MR4119781

Key words and phrases. Spherical statistic, Central Limit Theorem, Fréchet mean, smeary.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 609–624
https://doi.org/10.1214/21-BJPS500

Graph distances of continuum long-range percolation

Ercan Sönmez

Department of Statistics, University of Klagenfurt, Universitätsstraße 65–67, 9020 Klagenfurt, Austria,
ercan.soenmez@aau.at

Abstract. We consider a version of continuum long-range percolation on finite boxes of $\mathbb{R}^d$ in which the vertex set is given by the points of a Poisson point process and each pair of two vertices at distance r is connected with probability proportional to r^{-s} for a certain constant s. We explore the graph-theoretical distance in this model. The aim of this paper is to show that this random graph model undergoes phase transitions at values $s = d$ and $s = 2d$ in analogy to classical long-range percolation on $\mathbb{Z}^d$, by using techniques which are based on an analysis of the underlying Poisson point process.

References

Aizenman, M. and Newman, C. M. (1986). Discontinuity of the percolation density in one dimensional $1/|x-y|^2$ percolation models. *Communications in Mathematical Physics* **107**, 611–647. MR0868738

Benjamini, I. and Berger, N. (2001). The diameter of long-range percolation clusters on finite cycles. *Random Structures & Algorithms* **19**, 102–111. MR1848786 https://doi.org/10.1002/rsa.1022

Benjamini, I., Kesten, H., Peres, Y. and Schramm, O. (2004). The geometry of the uniform spanning forests: Transitions in dimensions 4,8,12. *Annals of Mathematics* **160**, 465–491. MR2123930 https://doi.org/10.4007/annals.2004.160.465

Berger, N. (2004). A lower bound for the chemical distance in sparse long-range percolation models. Preprint, available at https://arxiv.org/pdf/math/0409021.pdf.

Biskup, M. and Lin, J. (2019). Sharp asymptotic for the chemical distance in long-range percolation. *Random Structures & Algorithms* **55**, 560–583. MR3997479 https://doi.org/10.1002/rsa.20849

Coppersmith, D., Gamarnik, D. and Sviridenko, M. (2002). The diameter of a long-range percolation graph. *Random Structures & Algorithms* **21**, 1–13. MR1913075 https://doi.org/10.1002/rsa.10042

Deijfen, M. and van der Hofstad, R. and Hooghiemstra, G. (2013). Scale-free percolation. *Annales de L'IHP Probabilités et Statistiques* **49**, 817–838. MR3112435 https://doi.org/10.1214/12-AIHP480

Deprez, M., Hazra, R. S. and Wüthrich, M. V. (2015). Inhomogeneous long-range percolation for real-life network modeling. *Risks* **3**, 1–23. MR3497602

Deprez, M. and Wüthrich, M. V. (2018). Scale-free percolation in continuum space. *Communications in Mathematics and Statistics* **7**, 269–308. MR3986777 https://doi.org/10.1007/s40304-018-0142-0

Kallenberg, O. (1986). *Random Measures*. Berlin: Akademie-Verlag. MR0431373

Last, G., Nestmann, F. and Schulte, M. (2019). The random connection model and functions of edge-marked Poisson processes: Second order properties and normal approximation. Preprint, available at https://arxiv.org/abs/1808.01203.

Last, G. and Penrose, M. D. (2017). *Lectures on the Poisson Process*. Cambridge: Cambridge University Press. Institute of Mathematical Statistics. MR3791470

Milgram, S. (1967). The small world problem. *Psychology Today* **1**, 60–67.

Newman, C. M. and Schulman, L. S. (1986). One dimensional $1/|j-i|^s$ percolation models: The existence of a transition for $s \leq 2$. *Communications in Mathematical Physics* **104**, 547–571. MR0841669

Penrose, M. D. (1991). On a continuum percolation model. *Advances in Applied Probability* **23**, 536–556. MR1122874 https://doi.org/10.2307/1427621

Schulman, L. S. (1983). Long-range percolation in one dimension. *Journal of Physics A, Mathematical and General* **16**, L639–L641. MR0723249

Sönmez, E. and Rousselle, A. (2021). The random walk on the random connection model. Preprint.

Watts, D. J. (1999). *Small Worlds. The Dynamics of Networks Between Order and Randomness*. *Princeton Studies in Complexity*. Princeton: Princeton University Press. MR1716136

Key words and phrases. Diameter, graph distance, random graphs, long-range percolation, random connection model, Poisson process.

Watts, D. J. (2003). *Six Degrees. The Science of a Connected Age*. New York: W. W. Norton and Co. Inc. MR2041642

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. MR1716136

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 625–629
https://doi.org/10.1214/21-BJPS501

Partitioning some multivariate distributions

Joseph B. Kadane

Department of Statistics & Data Science, Carnegie Mellon University, Baker Hall 232, Pittsburgh, Pennsylvania 15213, USA, kadane@stat.cmu.edu

Abstract. This is a study of the behavior under partition of the sample space of three multivariate distributions: multinomial, multinomial-Dirichlet, and Dirichlet. A general theorem is given, of which all three are special cases.

References

Kadane, J. B. (2020). *Principles of Uncertainty*, 2nd ed. Boca Raton: Chapman and Hall. MR2799022 https://doi.org/10.1201/b11322

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press.

Key words and phrases. Multinomial distribution, Dirichlet distribution, Dirichlet-multinomial distribution, partition.

Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 3, 630–655
https://doi.org/10.1214/21-BJPS502

Hypotheses tests on the skewness parameter in a multivariate generalized hyperbolic distribution

Manuel Galea[1]**, Filidor Vilca**[2] **and Camila Borelli Zeller**[3]

[1]*Pontificia Universidad Católica de Chile, Santiago, Chile, mgalea@mat.uc.cl*
[2]*Universidade Estadual de Campinas, São Paulo, Brazil, fily@ime.unicamp.br*
[3]*Universidade Federal de Juiz de Fora, Minas Gerais, Brazil, camila.zeller@ufjf.edu.br*

Abstract. The class of generalized hyperbolic (GH) distributions is generated by a mean-variance mixture of a multivariate Gaussian with a generalized inverse Gaussian (GIG) distribution. This rich family of GH distributions includes some well-known heavy-tailed and symmetric multivariate distributions, including the Normal Inverse Gaussian and some members of the family of scale-mixture of skew-normal distributions. The class of GH distributions has received considerable attention in finance and signal processing applications. In this paper, we propose the likelihood ratio (LR) test to test hypotheses about the skewness parameter of a GH distribution. Due to the complexity of the likelihood function, the EM algorithm is used to find the maximum likelihood estimates both in the complete model and the reduced model. For comparative purposes and due to its simplicity, we also consider the Gradient (G) test. A simulation study shows that the LR and G tests are usually able to achieve the desired significance levels and the testing power increases as the asymmetry increases. The methodology developed in the paper is applied to two real datasets.

References

Abdi, M., Madadi, M., Balakrishnan, N. and Jamalizadeh, A. (2021). Family of mean-mixtures of multivariate normal distributions: properties, inference and assessment of multivariate skewness. *Journal of Multivariate Analysis*. MR4161850 https://doi.org/10.1016/j.jmva.2020.104679

Anderson, T. W. (2003). *An Introduccion to Multivariate Statistical Analysis*. New Jersey: John Wiley. MR1990662

Arellano-Valle, R. B., Ferreira, C. S. and Genton, M. G. (2018). Scale and shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* **166**, 98–110. MR3799637 https://doi.org/10.1016/j.jmva.2018.02.007

Arslan, O. (2009). Variance-mean mixture of Kotz-type distribution. *Communications in Statistics Theory and Methods* **38**, 272–284. MR2490405 https://doi.org/10.1080/03610920802192497

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

Azzalini, A. and Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* **76**, 106–129.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics* **5**, 151–157. MR0509451

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* **24**, 1–13. MR1436619 https://doi.org/10.1111/1467-9469.00045

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

Branco, M. and Dey, D. (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

Browne, R. P. and McNicholas, P. D. (2015). A mixture of generalized hyperbolic distributions. *Canadian Journal of Statistics* **43**, 176–198. MR3353379 https://doi.org/10.1002/cjs.11246

Key words and phrases. Likelihood ratio test, gradient test, EM algorithm, generalized hyperbolic distribution, normal inverse Gaussian distribution.

Campbell, J., Lo, A. and MacKinlay, A. (1997). *Econometrics of Financial Markets*. New Jersey: Princeton University Press.

Chang, Y. P., Hung, M. C., Liu, H. and Jan, J. F. (2005). Testing symmetry of a NIG distribution. *Communications in Statistics Simulation and Computation* **34**, 851–862. MR2224264 https://doi.org/10.1080/03610910500307877

Chang, Y. P., Hung, M. C., Wang, S. F. and Yu, C. T. (2010). An EM algorithm for multivariate NIG distribution and its applications to value-at-risk. *International Journal of Information and Management Sciences* **21**, 265–283. MR2766658

Choudhary, P. K. and Nagaraja, H. N. (2017). *Measurement Agreement: Models, Methods, and Applications*. New York: John Wiley.

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* **39**, 1–38. MR0501537

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

Fuller, W. A. (2006). *Measurement Error Models*. New York: Wiley. MR2301581

Fung, T. and Seneta, E. (2010). Extending the multivariate generalised *t* and generalised VG distributions. *Journal of Multivariate Analysis* **101**, 154–164. MR2557625 https://doi.org/10.1016/j.jmva.2009.06.006

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, Forida: Chapman and Hall/CRC. MR2155324 https://doi.org/10.1201/9780203492000

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–260. MR0061330 https://doi.org/10.1093/biomet/40.3-4.237

Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*. New York: Academic Press. MR1773820

Härdle, W. K. and Simar, L. (2012). *Applied Multivariate Statistical Analysis*. New York: Springer. MR2367300

Hellmich, M. and Kassberger, S. (2011). Efficient and robust portfolio optimization in the multivariate generalized hyperbolic framework. *Quantitative Finance* **11**, 1503–1516. MR2851086 https://doi.org/10.1080/14697680903280483

Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association* **89**, 897–904. MR1294733

Hu, W. (2005). Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk. PhD Thesis, Florida State University. MR2708679

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*, 3rd ed. New Jersey: Prentice Hall. MR1168210

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Heidelberg: Springer. MR0648107

Karlis, D. and Santourian, A. (2009). Model-based clustering with non-elliptically contoured distributions. *Statistics and Computing* **19**, 73–83. MR2470559 https://doi.org/10.1007/s11222-008-9072-0

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. New York: Springer. MR2162145 https://doi.org/10.1007/1-4020-3419-9

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

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

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

Leonenko, N. N., Petherick, S. and Sikorskii, A. (2012). A normal inverse Gaussian model for a risky asset with dependence. *Statistics & Probability Letters* **82**, 109–115. MR2863031 https://doi.org/10.1016/j.spl.2011.09.007

Luciano, E. and Schoutens, W. (2006). A multivariate jump-driven financial asset model. *Quantitative Finance* **6**, 385–402. MR2261218 https://doi.org/10.1080/14697680600806275

Madan, D. B. and Seneta, E. (1990). The variance-gamma (VG) model for share market returns. *Journal of Business* **63**, 511–524.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530. MR0397994 https://doi.org/10.1093/biomet/57.3.519

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press. MR0560319

Marfè, R. (2012). A generalised variance gamma process for financial applications. *Quantitative Finance* **12**, 75–87. MR2881607 https://doi.org/10.1080/14697688.2010.505199

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. New York: John Wiley. MR2392878 https://doi.org/10.1002/9780470191613

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

O'Hagan, A., Murphy, T. B., Gormley, I. C., McNicholas, P. D. and Karlis, D. (2016). Clustering with the multivariate normal inverse Gaussian distribution. *Computational Statistics & Data Analysis* **93**, 18–30. MR3406193 https://doi.org/10.1016/j.csda.2014.09.006

Øigård, A., Hanssen, A., Hansen, R. E. and Godtliebsen, F. (2005). EM-estimation and modeling of heavy-tailed processes with the multivairate normal inverse Gaussian distribution. *Signal Processing* **85**, 1655–1673.

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

Prause, K. (1999). The Generalized hyperbolic model: estimation, financial derivatives, and risk measures. Tese de Doutorado.

Seber, G. A. (1984). *Multivariate Observations*. New York: John Wiley. MR0746474 https://doi.org/10.1002/9780470316641

Wu, L. (2010). *Mixed Effects Models for Complex Data*. London: Chapman and Hall. MR2598844 https://doi.org/10.1201/9781420074086-b2

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.

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

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

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

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

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