Contents

Special issue in Honor of Heleno Bolfarine

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

Brazilian Journal of Probability and Statistics

Volume 37 • Number 2 • June 2023

Brazilian Journal of Probability and Statistics

Volume 37 • Number 2 • June 2023

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 37, Number 2, June 2023. 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
2023, Vol. 37, No. 2, 247–248
https://doi.org/10.1214/23-BJPS372PRE

Preface to the Special Issue

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 249–262
https://doi.org/10.1214/22-BJPS543

Multiplicative errors-in-variables beta regression

Jalmar M. F. Carrasco[1,a], **Silvia L. P. Ferrari**[2,b] **and Reinaldo B. Arellano–Valle**[3,c]

[1]*Department of Statistics, Federal University of Bahia, Brazil,* [a]*carrasco.jalmar@ufba.br*
[2]*Department of Statistics, University of São Paulo, Brazil,* [b]*silviaferrari@usp.br*
[3]*Department of Statistics, Pontifical Catholic University of Chile, Chile,* [c]*reivalle@mat.uc.cl*

Abstract. This paper deals with beta regression models with a covariate that is not directly observed; instead, it is replaced by a surrogate covariate that underpredicts its actual value. We propose a multiplicative errors-in-variables model tailored for this situation and develop calibration regression and pseudo-likelihood-based inference for the unknown parameters. The impact of ignoring the measurement error and the performance of the inference methods are evaluated through simulations and a real data illustration.

References

Arellano-Valle, R. B., Bolfarine, H. and Gasco, L. (2002). Measurement error models with nonconstant covariance matrices. *Journal of Multivariate Analysis* **82**, 395–415. MR1921633 https://doi.org/10.1006/jmva.2001.2024

Bartlett, J. W. and Keogh, R. H. (2018). Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. *Statistical Methods in Medical Research* **27**, 1695–1708. MR3803260 https://doi.org/10.1177/0962280216667764

Bolfarine, H., Castro, M. and Galea, M. (2020). *Regression Models for the Comparison of Measurement Methods*. Berlin: Springer. MR4178481 https://doi.org/10.1007/978-3-030-57935-7

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

Canterle, D. R. and Bayer, F. M. (2019). Variable dispersion beta regressions with parametric link functions. *Statistical Papers* **60**, 1541–1567. MR4017022 https://doi.org/10.1007/s00362-017-0885-9

Carrasco, J. M. F., Ferrari, S. L. P. and Arellano–Valle, R. B. (2022). Supplement to "Multiplicative errors-in-variables beta regression." https://doi.org/10.1214/22-BJPS543SUPP

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

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

Cribari-Neto, F. and Zeiles, A. (2010). Beta regression in R. *Journal of Statistical Software* **34**.

da Silva, A. R. and Lima, A. O. (2017). Geographically weighted beta regression. *Spatial Statistics* **21**, 279–303. MR3692189 https://doi.org/10.1016/j.spasta.2017.07.011

Ferrari, S. L. P. (2017). Beta regression. In *Wiley StatsRef: Statistics Reference Online* (N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels, eds.).

Ferrari, S. L. P. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31**, 799–815. MR2095753 https://doi.org/10.1080/0266476042000214501

Figueroa-Zuñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. P. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis* **61**, 137–147. MR3063006 https://doi.org/10.1016/j.csda.2012.12.002

Figueroa-Zuñiga, J. I., Carrasco, J. M. F., Arellano-Valle, R. B. and Ferrari, S. L. P. (2017). A Bayesian approach to errors-in-variables beta regression. *Brazilian Journal of Probability and Statistics* **61**. In press. MR3812382 https://doi.org/10.1214/17-BJPS354

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

Guolo, A. (2011). Pseudo-likelihood inference for regression models with misclassified and mismeasured variables. *Statistica Sinica* **21**, 1639–1663. MR2895993 https://doi.org/10.5705/ss.2010.065

Key words and phrases. Beta regression, maximum pseudo-likelihood, multiplicative measurement error, product of independent beta random variables, regression calibration.

Kipnis, V., Freedman, L. S., Carroll, R. J. and Midthune, D. (2016). A bivariate measurement error model for semicontinuous and continuous variables: Application to nutritional epidemiology. *Biometrics* **72**, 106–115. MR3500579 https://doi.org/10.1111/biom.12377

Parke, W. R. (1986). Pseudo maximum likelihood estimation: The asymptotic distribution. *The Annals of Statistics* **14**, 335–357. MR0829575 https://doi.org/10.1214/aos/1176349862

Pumi, G., Prass, T. S. and Souza, R. R. (2021). Geographically weighted beta regression. *Scadinavian Journal of Statistics* **48**, 68–86. MR3692189 https://doi.org/10.1016/j.spasta.2017.07.011

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

Skrondal, A. and Kuha, J. (2012). Improved regression calibration. *Psychometrika* **77**, 649–669. MR2993527 https://doi.org/10.1007/s11336-012-9285-1

Smithson, M. and Verkuilen, J. (2006). A better lemon-squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* **11**, 54–71.

Spiegelman, D., McDermott, A. and Rosner, B. (1997). Regression calibration method for correcting measurement error bias in nutritional epidemiology. *The American Journal of Clinical Nutrition* **65**, 1179–1186.

Stranda, M., Sillauc, S., Grunwaldb, G. K. and Rabinovitchd, N. (2015). Regression calibration with instrumental variables for longitudinal models with interaction terms, and application to air pollution studies. *Environ-Metrics* **26**, 393–405. MR3395250 https://doi.org/10.1002/env.2354

Trindade, D. B., Espinheira, P. L., Vasconcellos, K. L. P., Carrasco, J. M. F. and Lima, M. C. S. (2021). Beta regression model nonlinear in the parameters with additive measurement errors in variables. *PLoS ONE* **16**, e0254103.

Wang, J. and Luo, S. (2016). Augmented beta rectangular regression models: A bayesianperspective. *Biometrical Journal* **58**, 206–221. MR3457175 https://doi.org/10.1002/bimj.201400232

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

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 263–281
https://doi.org/10.1214/22-BJPS546

Beyond the lognormal distribution with properties and applications

Emilio Gómez–Déniz[1,a], **Osvaldo Venegas**[2,b] **and Héctor W. Gómez**[3,c]

[1]*Department of Quantitative Methods in Economics and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain,* [a]*emilio.gomez-deniz@ulpgc.es*
[2]*Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco, Chile,* [b]*ovenegas@uct.cl*
[3]*Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile,* [c]*hector.gomez@uantof.cl*

Abstract. In this paper, a new family of continuous random variables with positive support is introduced. Its density function has the capacity to incorporate features of unimodality and bimodality. Special attention is paid to the lognormal distribution which is included as a particular case. Its density function is given in closed-form, allowing probabilities, moments and other related measures such as skewness and kurtosis coefficients to be computed easily. In addition, a stochastic representation of the family that enables us to generate random variates of this model is also presented. Some properties related with the right tail and actuarial aspects of the distribution are also shown. This new family of distributions is numerically illustrated with data taken from the Medical Expenditure Panel Survey (MEPS), conducted by the US Agency of Health Research and Quality and with a well-known data set which has been studied widely in the actuarial literature.

References

Albrecher, H., Beirlant, J. and Teugels, J. L. (2017). *Reinsurance: Actuarial and Statistical Aspects*. New York: Wiley. MR3791478

Boland, P. J. (2007). *Statistical and Probabilistic Methods in Actuarial Science*. London: Chapman & Hall. MR2296279

Bolfarine, H., Gómez, H. W. and Rivas, L. (2011). The log-bimodal-skew-normal model. A geochemical application. *Journal of Chemometrics* **25**, 329–332.

Brooks, C. (2009). *RATS Handbook to Accompany Introductory Econometrics for Finance*. Cambridge: Cambridge University Press. MR1923094

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York, NY, USA: Chapman and Hall/CRC. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

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

García, V., Gómez-Déniz, E. and Vázquez-Polo, F. J. (2014). On modelling insurance data by using a generalized lognormal distribution. *Journal of Quantitative Methods for Economics and Business Administration* **18**, 146–162.

García, V., Gómez-Déniz, E. and Vázquez-Polo, F. J. (2016). Marshall-Olkin family of heavy-tailed distributions which includes the lognormal one. *Communications in Statistics Theory and Methods* **45**, 146–162. MR3475009 https://doi.org/10.1080/03610926.2013.873132

Gómez-Déniz, E., Sarabia, J. M. and Calderín-Ojeda, E. (2021). Bimodal normal distribution: Extensions and applications. *Journal of Computational and Applied Mathematics* **388**, 113292. MR4185118 https://doi.org/10.1016/j.cam.2020.113292

Hogg, R. and Klugman, S. A. (1984). *Loss Distributions*. New York: Wiley. MR0747141 https://doi.org/10.1002/9780470316634

Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2008). *Loss Models. From Data to Decisions*. New Jersey: Wiley. MR2436493 https://doi.org/10.1002/9780470391341

Key words and phrases. Bimodal distribution, folded normal distribution, hyperbolic function, lognormal distribution.

Klüppelberg, C. (1988). Subexponential distributions and integrated tails. *Journal of Applied Probability* **25**, 132–141. MR0929511 https://doi.org/10.2307/3214240

Konstantinides, D. G. (2018). *Risk Theory. A Heavy Tail Approach*. Singapore: World Scientific. MR3729275

Leone, F., Nelson, L. and Nottingham, R. (1961). The folded normal distribution. *Technometrics* **3**, 543–550. MR0130737 https://doi.org/10.2307/1266560

Martínez-Flórez, G., Bolfarine, H. and Gómez, H. W. (2014). The log alpha-power asymmetric distribution with application to air pollution. *EnvironMetrics* **25**, 44–56. MR3233743 https://doi.org/10.1002/env.2256

McNeil, A. J. (1997). Estimating the tails of loss severity distribution using extreme value theory. *ASTIN Bulletin* **27**, 117–137.

Prudnikov, A. P., Brichkov, Y. A. and Marichev, O. I. (1986). *Integrals and Series. Volume 2. Special Functions*. New York: Gordon & Breach. MR0874987

Rolski, T., Schmidli, H., Schmidt, V. and Teugel, J. (1999). *Stochastic Processes for Insurance and Finance*. Chichester: Wiley. MR1680267 https://doi.org/10.1002/9780470317044

Ruskeepaa, H. (2009). *Mathematica Navigator: Mathematics, Statistics, and Graphics*. New York: Academic Press.

Venegas, O., Bolfarine, H., Gallardo, D. I., Vergara-Fernández, A. and Gómez, H. W. (2016). A note on the log-alpha-skew-normal model with geochemical applications. *Applied Mathematics & Information Sciences* **10**, 1697–1703.

Wilcox, R. (2010). *Fundamentals of Modern Statistical Methods. Substantially Improving Power and Accuracy*, 2nd ed. Berlin: Springer. MR2598252 https://doi.org/10.1007/978-1-4419-5525-8

Yang, H. (2004). Crámer–Lundberg asymptotics. In *Encyclopedia of Actuarial Science*, 1–6. New York: Wiley.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 282–312
https://doi.org/10.1214/22-BJPS551

Regression modeling of censored data based on compound scale mixtures of normal distributions

Luis Benites[1,a], **Camila B. Zeller**[2,b], **Heleno Bolfarine**[3,c] **and Víctor H. Lachos**[4,d]

[1]*Departamento de Ciencias, Pontificia Universidad Católica del Perú, Lima, Perú,* [a]*lbenitess@pucp.edu.pe*
[2]*Departamento de Estatística, Universidade Federal de Juiz de Fora, Brazil,* [b]*camila.zeller@ufjf.edu.br*
[3]*Departamento de Estatística, Universidade de São Paulo,Butanta, São Paulo, Brazil,* [c]*hbolfar@ime.usp.br*
[4]*Department of Statistics, University of Connecticut, Storrs, Connecticut, U.S.A.,* [d]*hlachos@uconn.edu*

Abstract. In the framework of censored regression models, the distribution of the error term can depart significantly from normality, for instance, due to the presence of multimodality, skewness and/or atypical observations. In this paper we propose a novel censored linear regression model where the random errors follow a finite mixture of scale mixtures of normal (SMN) distribution. The SMN is an attractive class of symmetrical heavy-tailed densities that includes the normal, Student-t, slash and the contaminated normal distribution as special cases. This approach allows us to model data with great flexibility, accommodating simultaneously multimodality, heavy tails and skewness depending on the structure of the mixture components. We develop an analytically tractable and efficient EM-type algorithm for iteratively computing the maximum likelihood estimates of the parameters, with standard errors and prediction of the censored values as a by-products. The proposed algorithm has closed-form expressions at the E-step, that rely on formulas for the mean and variance of the truncated SMN distributions. The efficacy of the method is verified through the analysis of simulated and real datasets. The methodology addressed in this paper is implemented in the R package `CensMixReg`.

References

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

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

Ateya, S. F. (2014). Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. *Statistical Papers* **55**, 311–325. MR3188406 https://doi.org/10.1007/s00362-012-0480-z

Bartolucci, F. and Scaccia, L. (2005). The use of mixtures for dealing with non-normal regression errors. *Computational Statistics & Data Analysis* **48**, 821–834. MR2133580 https://doi.org/10.1016/j.csda.2004.04.005

Basford, K., Greenway, D., McLachlan, G. and Peel, D. (1997). Standard errors of fitted component means of normal mixtures. *Computational Statistics* **12**, 1–18.

Benites, L., Lachos, V. H. and Moreno, E. (2018). CensMixReg: Censored Linear Mixture Regression Models. R package version 3.1.

Caudill, S. B. (2012). A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods & Applications* **21**, 121–137. MR2916924 https://doi.org/10.1007/s10260-011-0182-z

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics* **51**, 79–99. MR1151954 https://doi.org/10.1016/0304-4076(92)90030-U

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38. MR0501537

Key words and phrases. Censored regression model, EM-type algorithms, finite mixture models, heavy-tails distributions, limit of detection, Tobit model.

Depraetere, N. and Vandebroek, M. (2014). Order selection in finite mixtures of linear regressions literature review and a simulation study. *Statistical Papers* **55**, 871–911. MR3227556 https://doi.org/10.1007/s00362-013-0534-x

Garay, A. M., Lachos, V. H., Bolfarine, H. and Cabral, C. R. B. (2017). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers* **58**, 247–278. MR3610167 https://doi.org/10.1007/s00362-015-0696-9

Holzmann, H., Munk, A. and Gneiting, T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics* **33**, 753–763. MR2300914 https://doi.org/10.1111/j.1467-9469.2006.00505.x

Karlsson, M. and Laitila, T. (2014). Finite mixture modeling of censored regression models. *Statistical Papers* **55**, 627–642. MR3227543 https://doi.org/10.1007/s00362-013-0509-y

Lachos, V. H., Cabral, C. R. B., Prates, M. O. and Dey, D. K. (2018). Flexible regression modelling for censored data based on mixtures of student-t distributions. *Computational Statistics* **34**, 123–152. MR3920576 https://doi.org/10.1007/s00180-018-0856-1

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

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. New Jersey: John Wiley & Sons. MR2392878 https://doi.org/10.1002/9780470191613

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley. MR1789474 https://doi.org/10.1002/0471721182

McNicholas, P. and Murphy, T. (2008). Parsimonious Gaussian mixture models. *Statistics and Computing* **18**, 285–296. MR2413385 https://doi.org/10.1007/s11222-008-9056-0

Meng, X. (1994). On the global and componentswise rates of convergence of the EM algorithm. *Linear Algebra and Its Applications* **199**, 413–425. MR1274429 https://doi.org/10.1016/0024-3795(94)90363-8

Mirfarah, E., Naderi, M. and Chen, D.-G. (2021). Mixture of linear experts model for censored data: A novel approach with scale-mixture of normal distributions. *Computational Statistics & Data Analysis* **158**, 107–182. MR4210957 https://doi.org/10.1016/j.csda.2021.107182

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica: Journal of the Econometric Society*, 765–799.

Osorio, F., Paula, G. A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis* **51**, 4354–4368. MR2364450 https://doi.org/10.1016/j.csda.2006.06.004

Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association* **73**, 730–738. MR0521324

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

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.

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

Zeller, C. B., Cabral, C. R. B., Lachos, V. H. and Benites, L. (2019). Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Advances in Data Analysis and Classification* **13**, 89–116. MR3935192 https://doi.org/10.1007/s11634-018-0337-y

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 313–328
https://doi.org/10.1214/23-BJPS569

On the asymptotic distribution of sample autocovariance differences of long-memory processes

Mauricio Zevallos[a]

Department of Statistics, University of Campinas, Campinas, SP, Brazil, [a]*amadeus@unicamp.br*

Abstract. This paper presents a procedure to calculate, in terms of analytic functions, the asymptotic covariance matrix of sample autocovariance differences of stationary autoregressive fractionally integrated moving average process with Gaussian and non-Gaussian errors. Furthermore, an application of minimum distance estimation of Gaussian autoregressive fractionally integrated moving average models is presented.

References

Baillie, R. T. and Chung, H. (2001). Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis* **22**, 631–650. MR1867390 https://doi.org/10.1111/1467-9892.00245

Beran, J. (1994). *Statistics for Long-Memory Processes*. New-York: Chapman & Hall. MR1304490

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes*. Berlin-Heidelberg: Springer. MR3075595 https://doi.org/10.1007/978-3-642-35512-7

Chung, C. (2002). Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Theory* **18**, 51–78. MR1885350 https://doi.org/10.1017/S0266466602181047

Doornik, J. A. and Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* **42**, 333–348. MR2005400 https://doi.org/10.1016/S0167-9473(02)00212-8

Doukhan, P., Oppenheim, G. and Taqqu, M. S. (2003). *Theory and Applications of Long-Range Dependence*. Boston: Birkhauser. MR1956041

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Tables of Integrals, Series and Products*, 7th ed. New York: Academic Press. MR2360010

Granger, C. W. J. and Joyeux, R. (1980). An introduction to longmemory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15–29. MR0605572 https://doi.org/10.1111/j.1467-9892.1980.tb00297.x

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). *Applied Statistics* **38**, 1–21.

Horváth, L. and Kokoszka, P. (2008). Sample autocovariances of long-memory time series. *Bernoulli* **14**, 405–418. MR2544094 https://doi.org/10.3150/07-BEJ113

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–176. MR0614953 https://doi.org/10.1093/biomet/68.1.165

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariance and autocorrelation of long-memory time series. *Journal of Econometrics* **73**, 261–284. MR1410007 https://doi.org/10.1016/0304-4076(95)01740-2

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

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. New Jersey: Wiley. MR2297359 https://doi.org/10.1002/9780470131466

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Robinson, P. M. (2003). *Time Series with Long Memory*. New York: Oxford University Press. MR2077520

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* **53**, 165–188. MR1185210 https://doi.org/10.1016/0304-4076(92)90084-5

Tieslau, M. A., Schmidt, P. and Baillie, R. T. (1996). A minimum distance estimator for long-memory processes. *Journal of Econometrics* **71**, 249–264. MR1381084 https://doi.org/10.1016/0304-4076(94)01703-4

Zevallos, M. and Palma, W. (2013). Minimum distance estimation of ARFIMA processes. *Computational Statistics & Data Analysis* **58**, 242–256. MR2997939 https://doi.org/10.1016/j.csda.2012.08.005

Key words and phrases. ARFIMA, central limit theorem, minimum distance estimation.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 329–350
https://doi.org/10.1214/23-BJPS570

Likelihood-based missing data analysis in crossover trials

Savita Pareek[a], Kalyan Das[b] and Siuli Mukhopadhyay[c]

Department of Mathematics, Indian Institute of Technology Bombay, Mumbai 400 076, India,
[a]*savita@math.iitb.ac.in,* [b]*kalyan@math.iitb.ac.in,* [c]*siuli@math.iitb.ac.in*

Abstract. A multivariate mixed-effects model seems to be the most appropriate for gene expression data collected in a crossover trial. It is, however, difficult to obtain reliable results using standard statistical inference when some responses are missing. Particularly for crossover studies, missingness is a serious concern as the trial requires a small number of participants. A Monte Carlo EM (MCEM)-based technique was adopted to deal with this situation. In addition to estimation, MCEM likelihood ratio tests are developed to test fixed effects in crossover models with missing data. Intensive simulation studies were conducted prior to analyzing gene expression data.

References

Basu, S. and Santra, S. (2010). A joint model for incomplete data in crossover trials. *Journal of Statistical Planning and Inference* **140**, 2839–2845. MR2651970 https://doi.org/10.1016/j.jspi.2010.03.006

Briggs, A., Clark, T., Wolstenholme, J. and Clarke, P. (2003). Missing... presumed at random: Cost-analysis of incomplete data. *Health Economics* **12**, 377–392.

Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90**, 242–252. MR1325132

Chinchilli, V. M. and Esinhart, J. D. (1996). Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine* **15**, 1619–1634.

Chow, S. C. and Shao, J. (1997). Statistical methods for two-sequence three-period cross-over designs with incomplete data. *Statistics in Medicine* **16**, 1031–1039.

Clough, E. and Barrett, T. (2016). The gene expression omnibus database. *Methods in Molecular Biology* **1418**, 93–110.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological* **39**, 1–38. MR0501537

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society Series C Applied Statistics* **43**, 49–93. MR3405479 https://doi.org/10.1111/rssa.12132

Fort, G. and Moulines, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *The Annals of Statistics* **31**, 1220–1259. MR2001649 https://doi.org/10.1214/aos/1059655912

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409. MR1141740

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6*, 721–741.

Goetghebeur, E. and Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics* **56**, 1139–1144. MR1815593 https://doi.org/10.1111/j.0006-341X.2000.01139.x

Grender, J. M. and Johnson, W. D. (1993). Analysis of crossover designs with multivariate response. *Statistics in Medicine* **12**, 69–89.

Ho, W. K., Matthews, J. N. S., Henderson, R., Farewell, D. and Rodgers, L. R. (2012). Dropouts in the AB/BA crossover design. *Statistics in Medicine* **31**, 1675–1687. MR2947516 https://doi.org/10.1002/sim.4497

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: A review. *Test* **18**, 1–43. MR2495958 https://doi.org/10.1007/s11749-009-0138-x

Jones, B. and Kenward, M. G. (2014). *Design and Analysis of Cross-over Trials*, 3rd ed. New York: Chapman & Hall/CRC. MR1014893 https://doi.org/10.1201/b17537

Leaker, B. R., Malkov, V. A., Mogg, R., Ruddy, M. K., Nicholson, G. C., Tan, A. J., Tribouley, C. and Chen, G. (2016). The nasal mucosal late allergic reaction to grass pollen involves type 2 inflammation (IL-5 and IL-13), the inflammasome (IL-1 b), and complement. *Nature* **10**, 408–420.

Key words and phrases. Crossover trials, Monte Carlo EM algorithm, MCEM likelihood ratio tests.

Levine, R. A. and Casella, G. (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics* **10**, 422–439. MR1939033 https://doi.org/10.1198/106186001317115045

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association* **83**, 1198–1202. MR0997603

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121. MR1354029

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. New York: Wiley-Interscience. MR1925014 https://doi.org/10.1002/9781119013563

Matthews, J. N. S. and Henderson, R. (2013). Two-period, two-treatment crossover designs subject to non-ignorable missing data. *Biostatistics* **14**, 626–638.

Patel, H. I. (1985). Analysis of incomplete data in a two-period crossover design with reference to clinical trials. *Biometrika* **72**, 411–418. MR0801779 https://doi.org/10.1093/biomet/72.2.411

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation* **7**, 353–383.

Putt, M. and Chinchilli, V. M. (1999). A mixed effects model for the analysis of repeated measures cross-over studies. *Statistics in Medicine* **18**, 3037–3058.

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

Richardson, B. A. and Flack, V. F. (1996). The analysis of incomplete data in the three-period two-treatment crossover design for clinical trials. *Statistics in Medicine* **15**, 127–143.

Rosenkranz, G. K. (2015). Analysis of cross-over studies with missing data. *Statistical Methods in Medical Research* **24**, 420–433. MR3372099 https://doi.org/10.1177/0962280214521349

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods* **7**, 147–177.

Senn, S. (2002). *Cross-over Trials in Clinical Research*. New York: Wiley-Interscience.

Shao, J. (2003). *Mathematical Statistics*, 2nd ed. Berlin: Springer. MR2002723 https://doi.org/10.1007/b97553

Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* **59**, 1140–1150. MR2025699 https://doi.org/10.1111/j.0006-341X.2003.00131.x

Stubbendick, A. L. and Ibrahim, J. G. (2006). Likelihood-based inference with nonignorable missing responses and covariates in models for discrete longitudinal data. *Statistica Sinica* **16**, 1143–1167. MR2327484

Tudor, G. E., Koch, G. G. and Catellier, D. (2000). Statistical methods for crossover designs in bioenvironmental and public health studies. *Handbook of Statistics* **18**, 571–614. MR1774239 https://doi.org/10.1016/S0169-7161(00)18011-3

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

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 351–364
https://doi.org/10.1214/23-BJPS571

Maximum likelihood estimation for the reflected stochastic linear system with a large signal

Xuekang Zhang[1,a] **and Huisheng Shu**[2,b]

[1]*School of Mathematics-Physics and Finance, and Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Anhui Polytechnic University, Wuhu, 241000, China,* [a]*xkzhang@ahpu.edu.cn*

[2]*College of Science, Donghua University, Shanghai, 201620, China,* [b]*hsshu@dhu.edu.cn*

Abstract. This paper deals with maximum likelihood estimation for the drift of the reflected stochastic linear system with a large signal. The law of iterated logarithm, consistency, and the asymptotic distributions of the maximum likelihood estimators in both the stationary and the non-stationary cases are studied based on the continuous observation.

References

Ata, B., Harrison, J. and Shepp, L. (2005). Drift rate control of a Brownian processing system. *The Annals of Applied Probability* **15**, 1145–1160. MR2134100 https://doi.org/10.1214/105051604000000855

Bishwal, J. (2008). *Parameter Estimation in Stochastic Differential Equations*. Berlin Heidelberg: Springer. MR2360279 https://doi.org/10.1007/978-3-540-74448-1

Bo, L., Wang, Y. and Yang, X. (2011). Some integral functionals of reflected SDEs and their applications in finance. *Quantitative Finance* **11**, 343–348. MR2777079 https://doi.org/10.1080/14697681003785926

Bo, L., Wang, Y., Yang, X. and Zhang, G. (2011). Maximum likelihood estimation for reflected Ornstein–Uhlenbeck processes. *Journal of Statistical Planning and Inference* **141**, 588–596. MR2719520 https://doi.org/10.1016/j.jspi.2010.07.001

Borovkov, A. (1984). *Asymptotic Methods in Queueing Theory*. New York: Wiley. MR0745620

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

Han, Z., Hu, Y. and Lee, C. (2016). Optimal pricing barriers in a regulated market using reflected diffusion processes. *Quantitative Finance* **16**, 639–647. MR3473978 https://doi.org/10.1080/14697688.2015.1034163

Harrison, M. (1986). *Brownian Motion and Stochastic Flow Systems*. New York: Join Wiley Sons. MR0798279

Hu, Y., Lee, C., Lee, M. and Song, J. (2015). Parameter estimation for reflected Ornstein–Uhlenbeck processes with discrete observations. *Statistical Inference for Stochastic Processes* **18**, 279–291. MR3395608 https://doi.org/10.1007/s11203-014-9112-7

Hu, Y. and Xi, Y. (2021). Estimation of all parameters in the reflected Ornstein–Uhlenbeck process from discrete observations. *Statistics & Probability Letters* **174**, 109099. MR4241641 https://doi.org/10.1016/j.spl.2021.109099

Jiang, H. and Dong, X. (2015). Parameter estimation for the non-stationary Ornstein–Uhlenbeck process with linear drift. *Statistical Papers* **56**, 257–268. MR3300439 https://doi.org/10.1007/s00362-014-0580-z

Jiang, H. and Liu, H. (2020). Asymptotic properties for the parameter estimation in Ornstein–Uhlenbeck process with discrete observations. *Electronic Journal of Statistics* **14**, 3192–3229. MR4146360 https://doi.org/10.1214/20-EJS1738

Jiang, H. and Xie, C. (2016). Asymptotic behaviours for the trajectory fitting estimator in Ornstein–Uhlenbeck process with linear drift. *Stochastics An International Journal of Probability and Stochastic Processes* **88**, 336–352. MR3459727 https://doi.org/10.1080/17442508.2015.1066378

Jiang, H. and Yang, Q. (2022). Moderate deviations for drift parameter estimations in reflected Ornstein–Uhlenbeck process. *Journal of Theoretical Probability* **35**, 1262–1283. MR4414419 https://doi.org/10.1007/s10959-021-01096-3

Jiang, H. and Zhang, N. (2020). Cramér-type moderate deviations for statistics in the non-stationary Ornstein–Uhlenbeck process. *Stochastics An International Journal of Probability and Stochastic Processes* **92**, 478–496. 2020. MR4085761 https://doi.org/10.1080/17442508.2019.1637877

Key words and phrases. Maximum likelihood estimation, reflected stochastic linear system, large signal, law of iterated logarithm, consistency, asymptotic distributions.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

Kutoyants, Y. (1991). Minimum distance parameter estimation for diffusion type observations. *Comptes Rendus de L'Académie des Sciences Series 1, Mathematics* **312**, 637–642. MR1101048

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

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

Lee, C., Bishwal, J. and Lee, M. (2012). Sequential maximum likelihood estimation for reflected Ornstein–Uhlenbeck processes. *Journal of Statistical Planning and Inference* **142**, 1234–1242. MR2879767 https://doi.org/10.1016/j.jspi.2011.12.002

Liptser, R. and Shiryaev, A. N. (1978). *Statistics of Random Processes*. New York: Springer. MR0488267

Liptser, R. and Shiryaev, A. N. (2001) *Statistics of Random Processes: II Applications*, 2nd ed, *Applications of Mathematics*. Berlin: Springer. MR1800858

Mao, X. (2007). *Stochastic Differential Equations and Applications*, 2nd ed. UK: Horwood. MR2380366 https://doi.org/10.1533/9780857099402

Ricciardi, L. and Sacerdote, L. (1987). On the probability densities of an Ornstein–Uhlenbeck process with a reflecting boundary. *Journal of Applied Probability* **24**, 355–369. MR0889800 https://doi.org/10.1017/s0021900200031004

Shimizu, Y. (2012). Local asymptotic mixed normality for discretely observed non-recurrent Ornstein–Uhlenbeck processes. *Annals of the Institute of Statistical Mathematics* **64**, 193–211. MR2878923 https://doi.org/10.1007/s10463-010-0307-4

Ward, A. and Glynn, P. (2003). Properties of the reflected Ornstein–Uhlenbeck process. *Queueing Systems* **44**, 109–123. MR1993278 https://doi.org/10.1023/A:1024403704190

Whitt, W. (2002). *Stochastic-Process Limits*. *Springer Series in Operations Research*. New York: Springer. MR1876437

Zang, Q. and Zhang, L. (2016a). Parameter estimation for generalized diffusion processes with reflected boundary. *Science China Mathematics* **59**, 1163–1174. MR3505047 https://doi.org/10.1007/s11425-015-5112-3

Zang, Q. and Zhang, L. (2016b). A general lower bound of parameter estimation for reflected Ornstein–Uhlenbeck processes. *Journal of Applied Probability* **53**, 22–32. MR3540777 https://doi.org/10.1017/jpr.2015.5

Zang, Q. and Zhang, L. (2019). Asymptotic behaviour of the trajectory fitting estimator for reflected Ornstein–Uhlenbeck processes. *Journal of Theoretical Probability* **32**, 183–201. MR3908911 https://doi.org/10.1007/s10959-017-0796-7

Zang, Q. and Zhu, C. (2016). Asymptotic behaviour of parametric estimation for nonstationary reflected Ornstein–Uhlenbeck processes. *Journal of Mathematical Analysis and Applications* **444**, 839–851. MR3535738 https://doi.org/10.1016/j.jmaa.2016.06.067

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 365–392
https://doi.org/10.1214/23-BJPS572

Longitudinal binary response models using alternative links for medical data

Alex de la Cruz Huayanay[1,a], **Jorge L. Bazán**[2,b] **and Carlos A. Ribeiro Diniz**[3,c]

[1]*Inter-Institutional Graduation Program in Statistics, USP/UFSCar, São Carlos, Brazil,* [a]*aldehu@usp.br*
[2]*Department of Applied Mathematics and Statistics, University of São Paulo, São Carlos, Brazi,* [b]*jlbazan@icmc.usp.br*
[3]*Department of Statistics, Federal University of São Carlos, São Carlos, Brazil,* [c]*dcad@ufscar.br*

Abstract. Motivated for a medical data about schizophrenia symptoms where an imbalanced binary response is observed, we introduce a broad class of link functions, called power and reverse power, as an alternative to analyse longitudinal binary data, particularly when it is imbalanced as is common in medical data. Bayesian estimation using an MCMC procedure through the No-U-Turn Sampler algorithm is proposed. Posterior predictive checks, Bayesian randomized quantile residuals, and a Bayesian influence measures are considered for model diagnostics. Different models are compared using selection model criteria. A simulation study is developed to analyse the prior sensitivity of the variance of the random effect and to assess the performance of the proposed model in the presence of imbalanced data. Finally, an application of the methodology studied in a set of medical data on the presence of schizophrenia symptom "thought disorder" is considered. In this data set, the presence of symptoms is much less than the absence, thus we show, in practice, the usefulness of using alternative link functions in imbalanced data.

References

Abanto-Valle, C. A., Dey, D. K. and Jiang, X. (2015). Binary state space mixed models with flexible link functions: A case study on deep brain stimulation on attention reaction time. *Statistics and Its Interface* **8**, 187–194. MR3322165 https://doi.org/10.4310/SII.2015.v8.n2.a6

Albert, J. and Chib, S. (2018). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–769. MR1380812 https://doi.org/10.1093/biomet/82.4.747

Ando, T. (2011). Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences* **31**, 13–38. MR2976700 https://doi.org/10.1080/01966324.2011.10737798

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

Bazán, J. L., Romeo, J. S. and Rodrigues, J. (2014). Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics* **28**, 467–482. MR3263060 https://doi.org/10.1214/13-BJPS218

Bazán, J. L., Torres-Avilés, F., Suzuki, A. and Louzada, F. (2017). Power and reversal power links for binary regressions: An application for motor insurance policyholders. *Applied Stochastic Models in Business and Industry* **33**, 22–34. MR3615258 https://doi.org/10.1002/asmb.2215

Bermúdez, L., Pérez, J. M., Ayuso, M., Gómez, E. and Vázquez, F. J. (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics* **42**, 779–786.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25. MR1394064 https://doi.org/10.2307/2291379

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

Cho, H., Ibrahim, J. G., Sinha, D. and Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics* **65**, 116–124. MR2665851 https://doi.org/10.1111/j.1541-0420.2008.01037.x

Key words and phrases. Asymmetric link, Bayesian diagnostic, binary response, imbalanced data, mixed-effects model, longitudinal data, health data.

Coelho, F. R., Russo, C. M. and Bazán, J. L. (2022). On outliers detection and prior distribution sensitivity in standard skew-probit regression models. *Brazilian Journal of Probability and Statistics* **36**, 441–462. MR4489175 https://doi.org/10.1214/22-bjps534

Collet, D. (2003). *Modelling Binary Data*, 2nd ed. Boca Raton, USA: Chapman & Hall/CRC. MR1999899

Costa, P. C., Bochkina, N. and Diniz, A. R. (2018). Bayesian truncated beta nonlinear mixed-effects models. *Journal of Applied Statistics* **45**, 320–346. MR3764730 https://doi.org/10.1080/02664763.2016.1276891

De la Cruz Huayanay, A., Bazán, J. L., Cancho, V. G. and Dey, D. K. (2019). Performance of asymmetric links and correction methods for imbalanced data in binary regression. *Journal of Statistical Computation and Simulation* **89**, 1694–1714. MR3934752 https://doi.org/10.1080/00949655.2019.1593984

Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press. MR2049007

Diggle, P., Liang, K. Y. and Zeger, S. L. (1994). *Longitudinal Data Analysis*. Oxford University Press.

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

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2012). *Applied Longitudinal Analysis, Vol. 998*. John Wiley & Sons. MR2830137

Fong, Y., Rue, H. and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics* **11**, 397–412.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160. MR0529531

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Technical report no. 462, Stanford Univ. CA Dept. of Statistics. MR1380275

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* **1**, 515–534. MR2221284 https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC. MR3235677

Gelman, A., Goegebeur, Y., Tuerlinckx, F. and Van Mechelen, I. (2000). Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society Series C Applied Statistics* **49**, 247–268. MR1821324 https://doi.org/10.1111/1467-9876.00190

Genest, C., Nikoloulopoulos, A. K., Rivest, L. P. and Fortin, M. (2013). Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Brazilian Journal of Probability and Statistics* **27**, 265–284. MR3064724 https://doi.org/10.1214/11-BJPS165

Gibbons, R. D. and Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics* **53**, 1527–1537.

Gómez-Déniz, E., Calderín-Ojeda, E. and Gómez, H. W. (2022). Asymmetric versus symmetric binary regresion: A new proposal with applications. *Symmetry* **14**, 733.

Hamming, R. W. (1970). On the distribution of numbers. *The Bell System Technical Journal* **49**, 1609–1625. MR0267809 https://doi.org/10.1002/j.1538-7305.1970.tb04281.x

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623. MR3214779

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*. John Wiley & Sons. MR1326603

Karimi, B. and Lavielle, M. (2019). Efficient Metropolis–Hastings sampling for nonlinear mixed effects models. In *BAYSM 2018*. Cham: Springer. MR4043149 https://doi.org/10.1007/978-3-030-30611-3_9

Komori, O., Eguchi, S., Ikeda, S., Okamura, H., Ichinokawa, M. and Nakayama, S. (2016). An asymmetric logistic regression model for ecological data. *Methods in Ecology and Evolution* **7**, 249–260.

Lemonte, A. J. and Bazán, J. L. (2018). New links for binary regression: An application to coca cultivation in Peru. *Test* **27**, 597–617. MR3858449 https://doi.org/10.1007/s11749-017-0563-1

Lemonte, A. J. and Moreno-Arenas, G. (2020). Improved estimation for a new class of parametric link functions in binary regression. *Sankhya Series B* **82**, 84–110. MR4093931 https://doi.org/10.1007/s13571-018-0179-9

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

Masuda, M. M. and Stone, R. P. (2015). Bayesian logistic mixed-effects modelling of transect data: Relating red tree coral presence to habitat characteristics. *ICES Journal of Marine Science* **72**, 2674–2683.

Nishio, M. and Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn sampler for estimating genetic parameters and breeding values. *Genetics Selection Evolution* **51**, 1–12.

Ordoñez, J. A., Prates, M. O., Bazán, J. L. and Lachos, V. H. (2023). Penalized complexity priors for the skewness parameter of power links. *Canadian Journal of Statistics*. https://doi.org/10.1002/cjs.11769

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G. M., Mallick, B. K. and Ibrahim, J. G. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *Annals of Applied Statistics* **5**, 449–467. MR2810405 https://doi.org/10.1214/10-AOAS390

Qiu, Z., Li, H., Su, H., Ou, G. and Wang, T. (2013). Logistic regression bias correction for large scale data with rare events. In *International Conference on Advanced Data Mining and Applications*, 133–144. Berlin, Heidelberg: Springer.

Scurrah, K. J., Palmer, L. J. and Burton, P. R. (2000). Variance components analysis for pedigree-based censored survival data using generalized linear mixed models (glmms) and Gibbs sampling in bugs. *Genetic Epidemiology* **19**, 127–148.

Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* **20**, 435–452.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **64**, 583–639. MR1979380 https://doi.org/10.1111/1467-9868.00353

Stan Development Team (2017). Stan modeling language users guide and reference manual. Version 2.16.0. https://mc-stan.org.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40**, 961–971.

Thara, R., Henrietta, M., Joseph, A., Rajkumar, S. and Eaton, W. W. (1994). Ten-year course of schizophrenia—the madras longitudinal study. *Acta Psychiatrica Scandinavica* **90**, 329–336.

Thomas, R., Have, T., Kunselman, A. R., Pulkstenis, E. P. and Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367–383.

Van Niekerk, J. and Rue, H. (2021). Skewed probit regression—identifiability, contraction and reformulation. *REVSTAT Statistical Journal* **19**, 1–22. MR4241745

Van Rossum, G. and Drake, F. L. (2009). Python 3 reference manual. Scotts Valley, CA: CreateSpace. Version 3.7.13. https://docs.python.org/.

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

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

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594. MR2756194

Wolfinger, R. and O'connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 393–411
https://doi.org/10.1214/23-BJPS573

Influence diagnostics for the power-normal Tobit model

Guillermo Martínez-Flórez[1,a], **Mario Pacheco**[2,b] **and Artur J. Lemonte**[3,c]

[1]*Departamento de Matemáticas y Estadística, Universidad de Córdoba, Montería, Colombia,* [a]*gmartinezflorez11@gmail.com*
[2]*Facultad de Estadística, Universidad Santo Tomás, Bogotá, Colombia,* [b]*mariopachecolopez@gmail.com*
[3]*Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal/RN, Brazil,* [c]*arturlemonte@gmail.com*

Abstract. Diagnostic analysis tools are studied for the censored power-normal Tobit regression model. We follow the Cook's (*J. R. Stat. Soc., Ser. B, Stat. Methodol.* **48** (1986) 133–169) approach, and several perturbation schemes are considered to detect influential observations. In particular, closed-form expressions of the normal curvatures for studying local influence are obtained under some perturbation schemes. The approach pursued also considers separate analysis for regression and scale-asymmetry parameters. Further, we define residuals to identify departures from the model assumptions, as well as to assess the overall goodness-of-fit of the censored power-normal Tobit regression model. The diagnostic measures developed are applied in a real data set for illustrative purposes.

References

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica* **41**, 997–1016. MR0440773 https://doi.org/10.2307/1914031

Atkinson, A. C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika* **68**, 13–20. MR0614937 https://doi.org/10.1093/biomet/68.1.13

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Cook, R. D. (1977). Detection of influential observations in linear regression. *Tecnometric* **19**, 15–18. MR0436478 https://doi.org/10.2307/1268249

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **48**, 133–169. MR0867994

Davison, A. C. and Gigli, A. (1989). Desviance residuals and normal scores plots. *Biometrika* **76**, 211–221.

Durrans, S. R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research* **28**, 1649–1655.

Galea, M., Paula, G. A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71–79.

Galea, M., Paula, G. A. and Uribe-Opazo, M. (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers* **44**, 23–45. MR1963364 https://doi.org/10.1007/s00362-002-0132-9

Galea, M., Riquelme, M. and Paula, G. A. (2000). Diagnostics methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics* **14**, 167–184. MR1860055

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

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**, 570–582.

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

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* **55**, 765–799.

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

Key words and phrases. Influence diagnostic, deviance residual, local influence, power-normal model.

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica* **46**, 1211–1215. MR0508692 https://doi.org/10.2307/1911445

Ortega, E. M. (2001). Influence analysis in generalized log-gamma regression models. Doctor Thesis, Department of Statistics, Universitu of São Paulo, Brazil. (in portuguese). MR1963013 https://doi.org/10.1016/S0167-9473(02)00104-4

Pewsey, A., Gómez, H. W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions. *Test* **21**, 775–789. MR2992092 https://doi.org/10.1007/s11749-011-0280-0

Poon, W. Y. and Poon, Y. S. (1999). Conformal normal curvature and assessement of local influence. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **61**, 51–61. MR1664096 https://doi.org/10.1111/1467-9868.00162

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

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–160. MR1049416 https://doi.org/10.1093/biomet/77.1.147

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

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer. MR1880596 https://doi.org/10.1007/978-1-4419-0300-6

Wei, B. C., Hu, Y. Q. and Fung, W. K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25–37. MR1614235 https://doi.org/10.1111/1467-9469.00086

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 412–430
https://doi.org/10.1214/23-BJPS574

High-dimensional regime for Wishart matrices based on the increments of the solution to the stochastic heat equation

Julie Gamain[1,a], **David A. C. Mollinedo**[2,c] **and Ciprian A. Tudor**[1,3,b]

[1]*CNRS, Université de Lille, Laboratoire Paul Painlevé UMR 8524, F-59655 Villeneuve d'Ascq, France,*
[a]*julie.gamain@univ-lille.fr,* [b]*ciprian.tudor@univ-lille.fr*
[2]*Universidade Tecnológica Federal do Paraná, Brazil*
[3]*Simion Stoilow Institute of Mathematics of the Romanian Academy, Bucharest, Romania,* [c]*davida@utfpr.edu.br*

Abstract. We consider a $n \times d$ random matrix $\mathcal{X}_{n,d}$ whose entries are the spatial increments of the solution to the stochastic heat equation with space-time white noise. We analyze the limit behavior of the associated Wishart matrix, by showing that it converges almost surely to a diagonal matrix (with equal diagonal terms) and the renormalized Wishart matrix satisfies a central limit theorem. Our techniques are based on the analysis on Wiener chaos, Malliavin calculus and Stein's method.

References

Araya, H. and Tudor, C. A. (2021). Asymptotic expansion for the quadratic variations of the solution to the heat equation with additive white noise. *Stochastics and Dynamics* **21**, 2150010. MR4192911 https://doi.org/10.1142/S0219493721500106

Bishop, A. N., Del Moral, P. and Niclas, A. (2018). An introduction to Wishart matrix moments. *MAL* **11**, 97–218.

Bourguin, S. and Dang, T. (2022). High-dimensional regimes for sequences of non-stationary Gaussian correlated Wishart matrices. *Random Matrices: Theory and Applications* **1**, 43.

Bourguin, S., Diez, C.-P. and Tudor, C. A. (2021). Limiting behavior of large correlated Wishart matrices with chaotic entries. *Bernoulli* **27**, 1077–1102. MR4255227 https://doi.org/10.3150/20-bej1266

Bubeck, S., Ding, J., Eldan, R. and Rácz, M. Z. (2016). Testing high-dimensional geometry in random graphs. *Random Structures & Algorithms* **49**, 503–532. MR3545825 https://doi.org/10.1002/rsa.20633

Bubeck, S. and Ganguly, S. (2018). Entropic CLT and phase transition in high dimensional Wishart matrices. *International Mathematics Research Notices* **2018**, 588–606. MR3801440 https://doi.org/10.1093/imrn/rnw243

Chow, P.-L. (2007). *Stochastic Partial Differential Equations*. *Chapman and Hall. CRC Applied Mathematics and Nonlinear Science Series*. Boca Raton, FL: Chapman and Hall. CRC. MR2295103

Dalang, R. (1999). Extending the martingale measure stochastic integral with applications to spatially homogeneous SPDE's. *Electronic Journal of Probability* **4**, 1–29 Erratum in Electr. J. Probab. 6 (2001), 5 pp. MR1825714

Dalang, R., Khoshnevisan, D., Mueller, C., Nualart, D. and Xiao, Y. (2006). *A Minicourse on Stochastic Partial Differential Equations*. *Lecture Notes in Mathematics, 1962*. Berlin: Springer. MR1500166

Diez, C.-P. and Tudor, C. A. (2021b). Limit behavior for Wishart matrices with Skorohod integrals. *ALEA—Latin American Journal of Probability and Mathematical Statistics* **18**, 1625–1641. MR4291035 https://doi.org/10.30757/alea.v18-59

Diez, D.-P. and Tudor, C. A. (2021). Non-central limit theorem for large Wishart matrices with Hermite entries. *Journal of Stochastic Analysis* **2**, 2.

Fang, X. and Koike, Y. (2022). New error bounds in multivariate normal approximations with exchangeable pairs with applications to Wishart matrices and fourth moment theorems. *The Annals of Applied Probability* **32**, 602–631.

Gamain, J. and Tudor, C. A. (2021). Random matrices and the stochastic wave equation. Preprint.

Jiang, T. and Li, D. (2015). Approximation of rectangular beta-Laguerre ensembles and large deviations. *Journal of Theoretical Probability* **28**, 804–847. MR3413957 https://doi.org/10.1007/s10959-013-0519-7

Johnstone, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I*, 307–333. Zürich: Eur. Math. Soc. MR2334195 https://doi.org/10.4171/022-1/13

Key words and phrases. Wishart matrix, stochastic heat equation, Wiener chaos, multiple stochastic integrals, Malliavin calculus, high-dimensional regime.

Krylov, N. (2006). On the foundation of the Lp-theory of stochastic partial differential equations. In *Stochastic Partial Differential Equations and Applications-VII*. *Lect. Notes Pure Appl. Math.* **245**, 179–191. Boca Raton: Chapman and Hall/CRC. MR2227229 https://doi.org/10.1201/9781420028720.ch16

Mikulincer, D. (2022). A CLT in Stein's distance for generalized Wishart matrices and higher order tensors. *International Mathematics Research Notices* **2022**, 7839–7872. MR4418721 https://doi.org/10.1093/imrn/rnaa336

Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus from Stein's Method to Universality*. Cambridge: Cambridge University Press. MR2962301 https://doi.org/10.1017/CBO9781139084659

Nourdin, I. and Pu, F. (2022). Gaussian fluctuation for Gaussian Wishart matrices of overall correlation. *Statistics & Probability Letters* **181**, 109269. MR4329126 https://doi.org/10.1016/j.spl.2021.109269

Nourdin, I. and Zheng, G. (2022). Asymptotic behavior of large Gaussian correlated Wishart matrices. *Journal of Theoretical Probability* **35**, 2239–2268. MR4509068 https://doi.org/10.1007/s10959-021-01133-1

Nualart, D. (2006). *Malliavin Calculus and Related Topics*, 2nd ed. Berlin: Springer. MR2200233

Rácz, M. Z. and Richey, J. (2019). A smooth transition from Wishart to GOE. *Journal of Theoretical Probability* **32**, 898–906. MR3959632 https://doi.org/10.1007/s10959-018-0808-2

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2514435

Tudor, C. A. (2013). *Analysis of Variations for Self-Similar Processes. A Stochastic Calculus Approach*. *Probability and Its Applications (New York)*. Cham: Springer. MR3112799 https://doi.org/10.1007/978-3-319-00936-0

Tudor, C. A. (2022). *Stochastic Partial Differential Equations with Additive Gaussian Noise. Analysis and Inference*. Hackensack, NJ: World Scientific.

Walsh, J. B. (1986). An introduction to stochastic partial differential equations. In *Lecture Notes in Mathematics, Vol. 1180*, 265–439. Berlin, Heidelberg, New York: Springer. MR0876085 https://doi.org/10.1007/BFb0074920

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20A**, 32–52.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 431–454
https://doi.org/10.1214/23-BJPS575

Dependent percolation on $\mathbb{Z}^2$

Bernardo N. B. de Lima[1,a], Vladas Sidoravicius[2] and Maria Eulália Vares[3,b]

[1]*Departamento de Matemática, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, CEP 30123-970, Belo Horizonte, MG, Brazil,* [a]*bnblima@mat.ufmg.br*
[2]*Shanghai New York University, Pudong New District, Shanghai, China*
[3]*Instituto de Matemática, Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos 149, CEP 21941-909, Rio de Janeiro, RJ, Brazil,* [b]*eulalia@im.ufrj.br*

Abstract. We consider a dependent percolation model on the square lattice $\mathbb{Z}^2$. The range of dependence is infinite in vertical and horizontal directions. In this context, we prove the existence of a phase transition. The proof exploits a multi-scale renormalization argument that is defined once the environment configuration is suitably good and, which, together with the main estimate for the induction step, comes from Kesten, Sidoravicius and Vares (*Electronic Journal of Probability* **27** (2022) 1–49). This paper is inspired by de Lima (Ph.D.Thesis, *Informes de Matemática. IMPA*, Série C-26/2004) where the simpler case of a deterministic environment was considered. It has various applications, including an alternative proof for the phase transition on the two dimensional random stretched lattice proved by Hoffman (*Comm. Math. Phys.* **254** (2005) 1–22).

References

Aizenman, M., Chayes, J. T., Chayes, L. and Newman, C. M. (1988). Discontinuity of the magnetization in one-dimendional $\frac{1}{|x-y|^2}$ Ising and Potts models. *Journal of Statistical Physics* **50**, 1–40. MR0939480 https://doi.org/10.1007/BF01022985

Basu, R. and Sly, A. (2014). Lipschitz embeddings of random sequences. *Probability Theory and Related Fields* **159**, 721–775. MR3230007 https://doi.org/10.1007/s00440-013-0519-7

Benjamini, I. and Kesten, H. (1995). Percolation of arbitrary words in $\{0, 1\}^{\mathbb{N}}$. *Annals of Probability* **23**, 1024–1060. MR1349161

Bramson, M., Durrett, R. T. and Schonmann, R. H. (1991). The contact process in a random environment. *Annals of Probability* **19**, 960–983. MR1112403

Chayes, L. and Schonmann, R. H. (2000). Mixed percolation as a bridge between site and bond percolation. *The Annals of Applied Probability* **10**, 1182–1196. MR1810870 https://doi.org/10.1214/aoap/1019487612

de Lima, B. N. B. (2003). Percolação de Bernoulli dependente em $\mathbb{Z}^2$. Ph.D. Thesis, Informes de Matemática. IMPA. Série C-26/2004.

Duminil-Copin, H., Hilário, M. R., Kozma, G. and Sidoravicius, V. (2018). Brochette percolation. *Israel Journal of Mathematics* **225**, 479–501. MR3805656 https://doi.org/10.1007/s11856-018-1678-0

Georgii, H. O., Häggström, O. and Maes, C. (2001). The random geometry of the equilibrium phases. In *Phase Transition and Critical Phenomena, Vol. 18* (C. Domb and J. Lebowitz, eds.) 1–142. London: Academic Press. MR2014387 https://doi.org/10.1016/S1062-7901(01)80008-2

Grimmett, G. R. (1999). *Percolation*, 2nd ed. Berlin: Springer. MR1707339 https://doi.org/10.1007/978-3-662-03981-6

Hoffman, C. (2005). Phase transition in dependent percolation. *Communications in Mathmatical Physics* **254**, 1–22. MR2116736 https://doi.org/10.1007/s00220-004-1240-2

Jahnel, B., Jhawar, S. K. and Vu, A. D. (2023). Continuum percolation in a nonstabilizing environment. Available at arXiv:2205.15366v2.

Jonasson, J., Mossel, E. and Peres, Y. (2000). Percolation in dependent random environment. *Random Structures & Algorithms* **16**, 333–343. MR1761579 https://doi.org/10.1002/1098-2418(200007)16:4<333::AID-RSA3>3.3.CO;2-3

Key words and phrases. Dependent percolation, multiscale renormalization, random environment.

Kesten, H., de Lima, B. N. B., Sidoravicius, V. and Vares, M. E. (2014). On the compatibility of binary sequences. *Communications on Pure and Applied Mathematics* **67**, 871–905. MR3193962 https://doi.org/10.1002/cpa.21486

Kesten, H., Sidoravicius, V. and Vares, M. E. (2022). Oriented percolation in a random environment. *Electronic Journal of Probability.* **27** 49 pp. MR4442896 https://doi.org/10.1214/22-ejp791

Kesten, H., Sidoravicius, V. and Zhang, Y. (1998). Almost all words are seen in critical site percolation on the triangular lattice. *Electronic Journal of Probability* **3**, 75 pp. MR1637089 https://doi.org/10.1214/EJP.v3-32

Kesten, H., Sidoravicius, V. and Zhang, Y. (2001). Percolation of arbitrary words on the closed packed graph of $\mathbb{Z}^2$. *Eletronic Journal of Probability* **6**, 27 pp. MR1825711 https://doi.org/10.1214/EJP.v6-77

Klein, A. (1994). Multiscale analysis in disorder systems: Percolation and contact process in random environments. In *Probability and Phase Transitions*. *Nato ASI Series*. Dordecht: Kluwer Academic Publishers. MR1283179

Liggett, T. (1992). The survival of the one dimensional contact process in random environments. *Annals of Probability* **20**, 696–723. MR1159569

McCoy, B. M. (1969). Theory of a two-dimensional Ising model with random impurities. III. Boundary efects. *Physical Review* **188**, 1014–1031. MR0286433

McCoy, B. M. (1970). Theory of a two-dimensional Ising model with random impurities. IV. Generalizations. *Physical Review B* **2**, 2795–2803. MR0334792

McCoy, B. M. and Wu, T. T. (1968). Theory of a two-dimensional Ising model with random impurities. I. Thermodynamics. *Physical Review* **176**, 631–643. MR0286431

McCoy, B. M. and Wu, T. T. (1969). Theory of a two-dimensional Ising model with random impurities. II. Spin correlation functions. *Physical Review* **188**, 982–1013. MR0286432

Nolin, P., Tassion, V. and Teixeira, A. (2022). No exceptional words for Bernoulli percolation. *Journal of the European Mathematical Society* To appear.

Winkler, P. (2000). Dependent percolation and colliding random walks. *Random Structures & Algorithms* **16**, 58–84. MR1728353 https://doi.org/10.1002/(SICI)1098-2418(200001)16:1<58::AID-RSA5>3.0.CO;2-E

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 2, 455–462
https://doi.org/10.1214/23-BJPS576

On the two-point function of the one-dimensional KPZ equation

Sergio I. López[1,a] **and Leandro P. R. Pimentel**[2,b]

[1]*Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico,* [a]*silo@ciencias.unam.mx*

[2]*Institute of Mathematics, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil,* [b]*leandro@im.ufrj.br*

Abstract. In this short communication, we show that basic tools from Malliavin calculus can be applied to derive the two-point function of the slope of the one-dimensional KPZ equation, starting from a two-sided Brownian motion with an arbitrary diffusion parameter, in terms of the polymer end-point annealed distribution associated to the stochastic heat equation. We also prove that this distribution is given in terms of the derivative of the variance of the solution of the KPZ equation.

References

Alberts, T., Khanin, K. and Quastel, J. (2014). The continuum directed random polymer. *Journal of Statistical Physics* **154**, 305–326. MR3162542 https://doi.org/10.1007/s10955-013-0872-z

Amir, G., Corwin, I. and Quastel, J. (2011). Probability distribution of the free energy of the continuum directed random polymer in 1 + 1 dimensions. *Communications on Pure and Applied Mathematics* **64**, 466–537. MR2796514 https://doi.org/10.1002/cpa.20347

Balázs, M., Quastel, J. and Seppäläinen, T. (2011). Fluctuation exponent of the KPZ/stochastic Burgers equation. *Journal of the American Mathematical Society* **24**, 683–708. MR2784327 https://doi.org/10.1090/S0894-0347-2011-00692-9

Bertini, L. and Giacomin, G. (1997). Stochastic Burgers and KPZ equations from particle systems. *Communications in Mathematical Physics* **183**, 571–607. MR1462228 https://doi.org/10.1007/s002200050044

Borodin, A., Corwin, I., Ferrari, P. L. and Vëto, B. (2015). Height fluctuations for the stationary KPZ equation. *Mathematical Physics, Analysis and Geometry* **18**, 20. MR3366125 https://doi.org/10.1007/s11040-015-9189-2

Corwin, I. (2012). The Kardar–Parisi–Zhang equation and universality class. *Random Matrices: Theory and Applications* **1**, 1130001. MR2930377 https://doi.org/10.1142/S2010326311300014

Das, S. and Zhu, W. (2022). Localization of the continuum directed random polymer. Available at https://arxiv.org/abs/2203.03607. MR4603188 https://doi.org/10.1007/s00023-023-01288-z

Dunlap, A., Gu, Y. and Komorowski, T. (2021). Fluctuations of the KPZ equation on a large torus. *Comm. Pure and Appl. Math.* To appear, available at https://arxiv.org/abs/2111.03650.

Forster, D., Nelson, D. R. and Stephen, M. J. (1977). Large-distance and long-time properties of a randomly stirred fluid. *Physical Review A* **16**, 732–749. MR0459274 https://doi.org/10.1103/PhysRevA.16.732

Gu, Y. and Komorowski, T. (2023). Another look at Balázs–Quastel–Seppäläinen theorem. *Transactions of the American Mathematical Society* **376**, 2947–2962. MR4557886 https://doi.org/10.1090/tran/8847

Hairer, M. (2013). Solving the KPZ equation. *Annals of Mathematics* **178**, 559–664. MR3071506 https://doi.org/10.4007/annals.2013.178.2.4

Kardar, M., Parisi, G. and Zhang, Y. Z. (1986). Dynamic scaling of growing interfaces. *Physical Review Letters* **56**, 889–892.

Le Doussal, P. (2017). Maximum of an Airy process plus Brownian motion and memory in Kardar–Parisi–Zhang growth. *Physical Review E* **96**, 060101.

Maes, C. and Thiery, T. (2017). Midpoint distribution of directed polymers in the stationary regime: Exact result through linear response. *Journal of Statistical Physics* **168**, 937–963. MR3682959 https://doi.org/10.1007/s10955-017-1839-2

Matetski, K., Quastel, J. and Remenik, D. (2021). The KPZ fixed point. *Acta Mathematica* **227**, 115–203. MR4346267 https://doi.org/10.4310/acta.2021.v227.n1.a3

Nualart, D. (2019). Malliavin calculus and normal approximations. *Ensaios Matemáticos* **34**, 1–74. MR4573937

Key words and phrases. KPZ equation, two-point function, Malliavin calculus.

Nualart, D. and Nualart, E. (2018). *Introduction to Malliavin Calculus. IMS Textbooks*. Cambridge: Cambridge University Press. MR3838464 https://doi.org/10.1017/9781139856485

O'Connell, N. and Yor, M. (2001). Brownian analogues of Burke's theorem. *Stochastic Processes and Their Applications* **96**, 285–304. MR1865759 https://doi.org/10.1016/S0304-4149(01)00119-3

Pimentel, L. P. R. (2020). Integration by parts and the KPZ two-point function. *Annals of Probability* **50**, 1755–1780. MR4474501 https://doi.org/10.1214/22-aop1564

Quastel, J. and Sarkar, S. (2023). Convergence of exclusion processes and the KPZ equation to the KPZ fixed point. *Journal of the American Mathematical Society* **36**, 251–289. MR4495842 https://doi.org/10.1090/jams/999

Seppäläinen, T. and Valkó, B. (2010). Bounds for scaling exponents for a $1 + 1$ dimensional directed polymer in a Brownian environment. *ALEA* **7**, 451–476. MR2741194

Virag, B. (2020). The heat and the landscape I. Available at https://arxiv.org/pdf/2008.07241.

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: https://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 https://www.e-publications.org/ims/submission/.

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

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

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

Individual and General Subscriptions:
https://imstat.org/journals-and-publications/subscriptions-and-orders/

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

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