Brazilian Journal of Probability and Statistics

Volume 31 • Number 3 • August 2017

www.imstat.org/bjps

Contents

BJPS is an IMS supported journal

ISSN 0103-0752

Brazilian Statistical Association

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 413
DOI: 10.1214/17-BJP313ED

Note from the Editor

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 414
DOI: 10.1214/17-BJPS313PRE

Preface

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 415–475
DOI: 10.1214/16-BJPS320

Probabilistic models for the (sub)tree(s) of life

Amaury Lambert

UPMC Univ Paris 06

Abstract. The goal of these lectures is to review some mathematical aspects of random tree models used in evolutionary biology to model species trees.

We start with stochastic models of tree shapes (finite trees without edge lengths), culminating in the β-family of Aldous' branching models.

We next introduce real trees (trees as metric spaces) and show how to study them through their contour, provided they are properly measured and ordered.

We then focus on the reduced tree, or coalescent tree, which is the tree spanned by species alive at the same fixed time. We show how reduced trees, like any compact ultrametric space, can be represented in a simple way via the so-called comb metric. Beautiful examples of random combs include the Kingman coalescent and coalescent point processes.

We end up displaying some recent biological applications of coalescent point processes to the inference of species diversification, to conservation biology and to epidemiology.

References

Aldous, D. (1991). The continuum random tree. I. *The Annals of Probability* **19**, 1–28. MR1085326
Aldous, D. (1993). The continuum random tree. III. *The Annals of Probability* **21**, 248–289. MR1207226
Aldous, D. (1996). Probability distributions on cladograms. In *Random Discrete Structures* (A. Friedman, W. Miller, D. Aldous and R. Pemantle, eds.) **76** 1–18. New York: Springer. MR1395604
Aldous, D. and Popovic, L. (2005). A critical branching process model for biodiversity. *Advances in Applied Probability* **37**, 1094–1115. MR2193998
Aldous, D. J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science* **16**, 23–34. MR1838600
Barthélémy, J.-P. and Guénoche, A. (1991). *Trees and Proximity Representations*. New York: Wiley. MR1138723
Bertoin, J. (1996). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge Univ. Press. MR1406564
Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. *Cambridge Studies in Advanced Mathematics* **102**. Cambridge: Cambridge Univ. Press. MR2253162
Blum, M. G. and François, O. (2006). Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology* **55**, 685–691.
Brown, J. K. M. (1994). Probabilities of evolutionary trees. *Systematic Biology* **43**, 78–91.
Burago, D., Burago, Y. and Ivanov, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Providence, RI: American Mathematical Society. MR1835418

Key words and phrases. Random tree, tree shape, real tree, reduced tree, branching process, coalescent, comb, phylogenetics, population dynamics, population genetics.

Champagnat, N. and Lambert, A. (2012). Splitting trees with neutral Poissonian mutations I: Small families. In *Stochastic Processes and Their Applications* **122**, 1003–1033. MR2891445

Champagnat, N. and Lambert, A. (2013). Splitting trees with neutral Poissonian mutations II: Largest and oldest families. *Stochastic Processes and their Applications* **123**, 1368–1414. MR3016227

Delaporte, C., Achaz, G. and Lambert, A. (2016). Mutational pattern of a sample from a critical branching population. *Journal of Mathematical Biology Journal of Mathematical Biology* **73**, 627–664. MR3535416

Dress, A., Moulton, V. and Terhalle, W. (1996). T-theory: An overview. *European Journal of Combinatorics* **17**, 161–175. MR1379369

Duquesne, T. (2006). The coding of compact real trees by real valued functions. Preprint. Available at arXiv:math/0604106.

Duquesne, T. and Le Gall, J.-F. (2002). Random trees, Lévy processes and spatial branching processes. *Asterisque—Société Mathématique de France* **281**. Paris: Société Mathématique de France.

Etienne, R. S., Morlon, H. and Lambert, A. (2014). Estimating the duration of speciation from phylogenies. *Evolution* **68**, 2430–2440.

Etienne, R. S. and Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. *Systematic Biology* **61**, 204–213.

Evans, S. N. (2008). *Probability and Real Trees. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Lecture Notes in Mathematics* **1920**. Berlin: Springer. MR2351587

Evans, S. N., Pitman, J. and Winter, A. (2005). Rayleigh processes, real trees, and root growth with re-grafting. *Probability Theory and Related Fields* **134**, 81–126. MR2221786

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112. Erratum *Theoretical Population Biology* **3** 240, 376. MR0325177

Geiger, J. (1996). Size-biased and conditioned random splitting trees. In *Stochastic Processes and Their Applications* **65**, 187–207. MR1425355

Geiger, J. and Kersting, G. (1997). Depth-first search of random trees, and Poisson point processes. In *Classical and Modern Branching Processes Minneapolis, MN, 1994*. *IMA Vol. Math. Appl.* **84**, 111–126. New York: Springer. MR1601713

Haas, B. (2016). Scaling limits of Markov-branching trees and applications. Preprint. Available at arXiv:1605.07873.

Haas, B., Miermont, G., Pitman, J. and Winkel, M. (2008). Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability* **36**, 1790–1837. MR2440924

Hagen, O., Hartmann, K., Steel, M. and Stadler, T. (2015). Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic Biology* **64**, 432–440.

Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability* **3**, 44–77. MR0282451

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444–448.

Kingman, J. (1982). The coalescent. *Stochastic Processes and Their Applications* **13**, 235–248. MR0671034

Knuth, D. E. (1997). *The Art of Computer Programming*. Reading, MA: Addison-Wesley. MR0378456

Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Berlin: Springer. MR2250061

Lambert, A. (2008). Population dynamics and random genealogies. *Stochastic Models* **24**, 45–163. MR2466449

Lambert, A. (2009). The allelic partition for coalescent point processes. *Markov Processes and Related Fields* **15**, 359–386. MR2554367

Lambert, A. (2010). The contour of splitting trees is a Lévy process. *The Annals of Probability* **38**, 348–395. MR2599603

Lambert, A. (2011). Species abundance distributions in neutral models with immigration or mutation and general lifetimes. *Journal of Mathematical Biology* **63**, 57–72. MR2806489

Lambert, A., Alexander, H. K. and Stadler, T. (2014a). Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *Journal of Theoretical Biology* **352**, 60–70. MR3198053

Lambert, A., Morlon, H. and Etienne, R. S. (2014b). The reconstructed tree in the lineage-based model of protracted speciation. *Journal of Mathematical Biology* **70**, 367–397. MR3294977

Lambert, A. and Popovic, L. (2013). The coalescent point process of branching trees. *Annals of Applied Probability* **23**, 99–144. MR3059232

Lambert, A., Simatos, F. and Zwart, B. (2013). Scaling limits via excursion theory: Interplay between Crump–Mode–Jagers branching processes and processor-sharing queues. *The Annals of Applied Probability* **23**, 2357–2381. MR3127938

Lambert, A. and Stadler, T. (2013). Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology* **90**, 113–128.

Lambert, A. and Steel, M. (2013). Predicting the loss of phylogenetic diversity under non-stationary diversification models. *Journal of Theoretical Biology* **337**, 111–124. MR3115320

Lambert, A. and Trapman, P. (2013). Splitting trees stopped when the first clock rings and Vervaat's transformation. *Journal of Applied Probability* **50**, 208–227. MR3076782

Lambert, A. and Uribe Bravo, G. (2016a). The comb representation of compact ultrametric spaces. Preprint. Available at arXiv:1602.08246.

Lambert, A. and Uribe Bravo, G. (2016b). Totally ordered, measured trees and splitting trees with infinite variation. Preprint. Available at arXiv:1607.02114.

Le Gall, J.-F. (1993). The uniform random tree in a Brownian excursion. *Probability Theory and Related Fields* **96**, 369–383. MR1231930

Le Gall, J.-F. (2005). Random trees and applications. *Probability Surveys* **2**, 245–311. MR2203728

Le Gall, J.-F. and Miermont, G. (2012). Scaling limits of random trees and planar maps. In *Probability and Statistical Physics in Two and More Dimensions* (D. Ellwood, ed.). *Clay Math. Proc.* **15**, 155–211. Providence, RI: American Mathematical Society. MR3025391

Manceau, M., Lambert, A. and Morlon, H. (2015). Phylogenies support out-of-equilibrium models of biodiversity. *Ecology Letters* **18**, 347–356.

Mooers, A., Gascuel, O., Stadler, T., Li, H. and Steel, M. (2012). Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Systematic Biology* **61**, 195–203.

Murtagh, F. (1984). Counting dendrograms: A survey. *Discrete Applied Mathematics* **7**, 191–199. MR0727923

Nee, S. (2006). Birth-death models in macroevolution. *Annual Review of Ecology, Evolution and Systematics* **37**, 1–17.

Nee, S., May, R. and Harvey, P. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **344**, 305–311.

Nee, S. and May, R. M. (1997). Extinction and the loss of evolutionary history. *Science* **278**, 692–694.

Paulin, F. (1989). The Gromov topology on R-trees. *Topology and its Applications* **32**, 197–221. MR1007101

Pitman, J. (2006). *Combinatorial Stochastic Processes*. In *Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics* **1875**. Berlin: Springer. MR2245368

Popovic, L. (2004). Asymptotic genealogy of a critical branching process. *Annals of Applied Probability* **14**, 2120–2148. MR2100386

Richard, M. (2014). Splitting trees with neutral mutations at birth. In *Stochastic Processes and Their Applications* **124**, 3206–3230. MR3231617

Semple, C. and Steel, M. A. (2003). *Phylogenetics*. *Oxford Lecture Series in Mathematics and its Applications* **24**. Oxford: Oxford Univ. Press. MR2060009

Slowinski, J. B. (1990). Probabilities of n-trees under two models: A demonstration that asymmetrical interior nodes are not improbable. *Systematic Biology* **39**, 89–94.

Stadler, T. (2010). Sampling-through-time in birth–death trees. *Journal of Theoretical Biology* **267**, 396–404. MR2974417

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* **108**, 6187–6192.

Stanley, R. P. (1999). *Enumerative Combinatorics*. *Cambridge Studies in Advanced Mathematics* **62**. Cambridge: Cambridge Univ. Press. MR1676282

Trapman, P. and Bootsma, M. C. J. (2009). A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection. *Mathematical Biosciences* **219**, 15–22. MR2518217

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 476–501
DOI: 10.1214/16-BJPS342

Finite-size corrections to the speed of a branching-selection process

Francis Comets and Aser Cortines

Université Paris Diderot—Paris 7

Abstract. We consider a particle system studied by E. Brunet and B. Derrida (*Phys. Rev. E* **70** (2004) 016106), which evolves according to a branching mechanism with selection of the fittest keeping the population size fixed and equal to N. The particles remain grouped and move like a travelling front driven by a random noise with a deterministic speed. Because of its mean-field structure, the model can be further analysed as $N \to \infty$. We focus on the case where the noise lies in the max-domain of attraction of the Weibull extreme value distribution and show that under mild conditions the correction to the speed has universal features depending on the tail probabilities.

References

Aïdékon, E. (2013). Convergence in law of the minimum of a branching random walk. *Ann. Probab.* **41**, 1362–1426. MR3098680

Athreya, K. B. and Ney, P. E. (2004). *Branching Process*. Mineola: Dover Publications. MR2047480

Bérard, J. and Gouéré, J.-B. (2010). Brunet–Derrida behavior of branching-selection particle systems on the line. *Comm. Math. Phys.* **298**, 323–342. MR2669438

Berestycki, J., Berestycki, N. and Schweinsberg, J. (2013). The genealogy of branching Brownian motion with absorption. *Ann. Probab.* **41**, 527–618. MR3077519

Biggins, J. D. (1977). Chernoff's theorem in the branching random walk. *J. Appl. Probab.* **14**, 630–636. MR0464415

Bramson, M. (1983). Convergence of solutions of the Kolmogorov equation to travelling waves. *Mem. Amer. Math. Soc.* **44**, iv+190. MR0705746

Brunet, E. and Derrida, B. (1997). Shift in the velocity of a front due to a cut-off. *Phys. Rev. E* **56**, 2597–2604. MR1473413

Brunet, E. and Derrida, B. (1999). Microscopic models of traveling wave equations. *Comput. Phys. Commun.* **121–122**, 376–381.

Brunet, E. and Derrida, B. (2004). Exactly soluble noisy traveling-wave equation appearing in the problem of directed polymers in a random medium. *Phys. Rev. E* **70**, 016106. MR2125704

Comets, F., Quastel, J. and Ramírez, A. F. (2013). Last passage percolation and traveling fronts. *J. Stat. Phys.* **152**, 419–451. MR3082639

Cook, J. and Derrida, B. (1990). Directed polymers in a random medium: 1/d expansion and the n-tree approximation. *J. Phys. A* **23**, 1523–1554. MR1048783

Cortines, A. (2014). Front velocity and directed polymers in random medium. *Stochastic Process. Appl.* **124**, 3698–3723. MR3249352

Cortines, A. (2016). The genealogy of a solvable population model under selection with dynamics related to directed polymers. *Bernoulli* **22**, 2209–2236. MR3498028

Key words and phrases. Front propagation, branching random walk, selection, extreme value theory, first-passage percolation, finite-size corrections, propagation speed, mean-field.

Couronné, O. and Gerin, L. (2014). A branching-selection process related to censored Galton–Watson process. *Ann. Inst. Henri Poincaré Probab. Stat.* **50**, 84–94. MR3161523

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*. New York: Springer. MR0950166

Durrett, R. and Remenik, D. (2011). Brunet–Derrida particle systems, free boundary problems and Wiener–Hopf equations. *Ann. Probab.* **39**, 2043–2078. MR2932664

Gantert, N., Hu, Y. and Shi, Z. (2011). Asymptotics for the survival probability in a killed branching random walk. *Ann. Inst. Henri Poincaré Probab. Stat.* **47**, 111–129. MR2779399

Maillard, P. (2016). Speed and fluctuations of N-particle branching Brownian motion with spatial selection. *Probab. Theory Related Fields* **166**, 1061–1173. MR3568046

Mallein, B. N-Branching random walk with α-stable spine. *Theory Probab. Appl.* To appear. Preprint. Available at arXiv:1503.03762.

Mueller, C., Mytnik, L. and Quastel, J. (2011). Effect of noise on front propagation in reaction-diffusion equations of KPP type. *Invent. Math.* **184**, 405–453. MR2793860

Panja, D. (2004). Effects of fluctuations on propagating fronts. *Phys. Rep.* **393**, 87–174.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer. MR0900810

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 502–515
DOI: 10.1214/16-BJPS321

Statistical inference for the parameter of Lindley distribution based on fuzzy data

Abbas Pak

Shahrekord University

Abstract. In many practical situations, we face data which are not only random but vague as well. To deal with these two types of uncertainties, it is necessary to incorporate fuzzy concept into statistical technique. In this paper, we investigate the maximum likelihood estimation and Bayesian estimation for Lindley distribution when the available observations are reported in the form of fuzzy data. We employ the EM algorithm to determine the maximum likelihood estimate (MLE) of the parameter and construct approximate confidence interval by using the asymptotic normality of the MLE. In the Bayesian setting, we use an approximation based on the Laplace approximation as well as a Markov Chain Monte Carlo technique to compute the Bayes estimate of the parameter. In addition, the highest posterior density credible interval of the unknown parameter is obtained. Extensive simulations are performed to compare the performances of the different proposed methods.

References

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

Gil, M. A., López-Diaz, M. and Ralescu, D. A. (2006). Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems* **157**, 2546–2557. MR2328381

Ghitany, M. E., Al-Mutairi, D. K. and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation* **79**, 279–287. MR2477530

Gupta, P. K. and Singh, B. (2012). Parameter estimation of Lindley distribution with hybrid censored data. *International Journal of System Assurance Engineering and Management*. doi:10.1007/s13198-012-0120-y.

Gupta, P. K., Singh, B. Krishna, H., and Kumar, K. (2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. *Mathematics and Computers in Simulation* **82**, 281–294. MR2846043

Lindley, D. V. (1958). Fudicial distributions and Bayesian theorem. *Journal of the Royal Statistical Society B* **20**, 102–107. MR0095550

Mazucheli, J. and Achcarb, J. A. (2011). The Lindley distribution applied to competing risks lifetime data. *Computer Methods and Programs in Biomedicine* **104**, 188–192.

Pak, A., Parham, G. H. and Saraj, M. (2013). Inferences on the competing risk reliability problem for exponential distribution based on fuzzy data. *IEEE Transactions on Reliability* **63**, 2–13.

Rubinstein, R. Y. and Kroese, D. P. (2006). *Simulation and the Monte Carlo Method*, 2nd ed. New Jersey: Wiley. MR2365210

Key words and phrases. Fuzzy data analysis, maximum likelihood principle, Bayesian estimation, asymptotic confidence interval.

Shafiq, M. and Viertl, R. (2014). Maximum likelihood estimation for Weibull distribution in case of censored fuzzy life time data, 1–17. Available at http://www.statistik.tuwien.ac.at/forschung/SM/SM-2014-2complete.pdf.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86. MR0830567

Viertl, R. (2006). Univariate statistical analysis with fuzzy data. *Computational Statistics & Data Analysis* **55**, 133–147. MR2297592

Wu, H. C. (2004). Fuzzy Bayesian estimation on lifetime data. *Computational Statistics* **19**, 613–633. MR2108493

Zadeh, L. A. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications* **10**, 421–427. MR0230569

Zarei, R., Amini, M., Taheri, S. M. and Rezaei, A. H. (2012). Bayesian estimation based on vague lifetime data. *Soft Computing* **16**, 165–174.

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 516–541
DOI: 10.1214/16-BJPS322

A new lifetime model with variable shapes for the hazard rate

Ahmed Z. Afify[a], Gauss M. Cordeiro[b], Nadeem Shafique Butt[c], Edwin M. M. Ortega[d] and Adriano K. Suzuki[d]

[a]*Benha University*
[b]*Universidade Federal de Pernambuco*
[c]*King Abdulaziz University*
[d]*Universidade de São Paulo*

Abstract. We define and study a new generalization of the complementary Weibull geometric distribution introduced by Tojeiro et al. (*J. Stat. Comput. Simul.* **84** (2014) 1345–1362). The new lifetime model is referred to as the Kumaraswamy complementary Weibull geometric distribution and includes twenty three special models. Its hazard rate function can be constant, increasing, decreasing, bathtub and unimodal shaped. Some of its mathematical properties, including explicit expressions for the ordinary and incomplete moments, generating and quantile functions, Rényi entropy, mean residual life and mean inactivity time are derived. The method of maximum likelihood is used for estimating the model parameters. We provide some simulation results to assess the performance of the proposed model. Two applications to real data sets show the flexibility of the new model compared with some nested and non-nested models.

References

Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters* **39**, 35–42. MR1649319

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

Afify, A. Z., Nofal, Z. M. and Ebraheim, A. N. (2015). Exponentiated transmuted generalized Rayleigh distribution: A new four parameter Rayleigh distribution. *Pakistan Journal of Statistics and Operation Research* **11**, 115–134. MR3638751

Afify, A. Z., Cordeiro, G. M., Yousof, H. M., Alzaatreh, A. and Nofal, Z. M. (2016). The Kumaraswamy transmuted-*G* family of distributions: Properties and applications. *Journal of Data Science* **14**, 245–270.

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

Chen, G. and Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology* **27**, 154–161.

Cordeiro, G. M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation* **81**, 883–898. MR2806932

Key words and phrases. Censored data, complementary Weibull geometric, generating function, maximum likelihood, order statistic.

Cordeiro, G. M., Ortega, E. M. M. and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of The Franklin Institute* **347**, 1399–1429. MR2720934
Cordeiro, G. M., Nadarajah, S. and Ortega, E. M. M. (2013). General results for the beta Weibull distribution. *Journal of Statistical Computation and Simulation* **83**, 1082–1114. MR3169222
Cordeiro, G. M., Hashimoto, E. M. and Ortega, E. M. M. (2014). The McDonald Weibull model. *Statistics: A Journal of Theoretical and Applied Statistics* **48**, 256–278. MR3175769
Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics—Theory and Methods* **31**, 497–512. MR1902307
Gomes, A. E., da-Silva, C. Q., Cordeiro, G. M. and Ortega, E. M. M. (2014). A new lifetime model: The Kumaraswamy generalized Rayleigh distribution. *Journal of Statistical Computation and Simulation* **84**, 290–309. MR3169327
Guess, F. and Proschan, F. (1988). Mean residual life, theory and applications. *Handbook of Statistics* **7**, 215–224.
Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions: Theory and methods. *Australian and New Zealand Journal of Statistics* **41**, 173–188. MR1705342
Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal* **43**, 117–130. MR1847768
Khan, M. N. (2015). The modified beta Weibull distribution. *Hacettepe Journal of Mathematics and Statistics* **44**, 1553–1568.
Kundu, D. and Raqab, M. Z. (2005). Generalized Rayleigh distribution: Different methods of estimations. *Computational Statistics and Data Analysis* **49**, 187–200. MR2129172
Kundu, D. and Raqab, M. Z. (2009). Estimation of $R = P(Y < X)$ for three parameter Weibull distribution. *Statistics and Probability Letters* **79**, 1839–1846. MR2560482
Lai, C. D. and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. *Springer Science and Business Media*. Berlin: Springer. MR2223811
Lee, C., Famoye, F. and Olumolade, O. (2007). Beta-Weibull distribution: Some properties and applications to censored data. *Journal of Modern Applied Statistical Methods* **6**, 173–186.
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
Louzada, F., Marchi, V. and Carpenter, J. (2013). The complementary exponentiated exponential geometric lifetime distribution. *Journal of Probability and Statistics* **2013**, Article ID 502159. MR3037955
Mudholkar, G. S., Srivastava, D. K. and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association* **91**, 1575–1583. MR1439097
Nadarajah, S., Cordeiro, G. M. and Ortega, E. M. M. (2013). The exponentiated Weibull distribution: A survey. *Statistical Papers* **54**, 839–877. MR3072904
Nassar, M. M. and Eissa, F. H. (2003). On the exponentiated Weibull distribution. *Communications in Statistics—Theory and Methods* **32**, 1317–1336. MR1985853
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.
Perdoná, G. S. C. (2006). Modelos de Riscos Aplicados a Análise de Sobrevivência. Doctoral thesis, Institute of Computer Science and Mathematics, University of São Paulo, Brasil (in Portuguese).
Rayleigh, J. W. S. (1880). On the resultant of a large number of vibration of the same pitch and arbitrary phase. *Philosophical Magazine, 5th Series* **10**, 73–78.
Rinne, H. (2009). *The Weibull Distribution: A Handbook*. London: CRC Press. MR2477856
Silva, A. N. F. (2004). Estudo evolutivo das crianças expostas ao HIV e notificadas pelo núcleo de vigilância epidemiológica do HCFMRP-USP. M.Sc. thesis, University of São. Paulo, Brasil (in Portuguese).

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

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, Transactions, ASME* **18**, 293–297.

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 542–560
DOI: 10.1214/16-BJPS323

Bias correction in power series generalized nonlinear models

Priscila G. Silva, Audrey H. M. A. Cysneiros and Gauss M. Cordeiro

Universidade Federal de Pernambuco

Abstract. Power series generalized nonlinear models [*Comput. Statist. Data Anal.* **53** (2009) 1155–1166] can be used when the Poisson assumption of equidispersion is not valid. In these models, we consider a more general family of discrete distributions for the response variable and a nonlinear structure for the regression parameters, although the dispersion parameter and other shape parameters are assumed known. We derive a general matrix formula for the second-order bias of the maximum likelihood estimate of the regression parameter vector in these models. We use the results by [*J. Roy. Statist. Soc. B* **30** (1968) 248–275] and bootstrap technique [*Ann. Statist.* **7** (1979) 1–26] to obtain the bias-corrected maximum likelihood estimate. Simulation studies are performed using different estimates. We also present an empirical application.

References

Botter, D. A. and Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Journal of Statistical Computation and Simulation* **62**, 91–104. MR1682556

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, 434 p. New York: Cambridge University Press. MR1648274

Consul, P. C. (1990). New class of location-parameter discrete probability distributions and their characterizations. *Communications in Statistics, Theory and Methods* **19**, 4653–4666. MR1114865

Cook, D. R., Tsai, C. L. and Wei, B. C. (1986). Bias in nonlinear regression. *Biometrika* **73**, 615–623. MR0897853

Cordeiro, G. M., Andrade, M. G. and De Castro, M. (2009). Power series generalized nonlinear models. *Computational Statistics and Data Analysis* **53**, 1155–1166. MR2657079

Cordeiro, G. M., Ferrari, S. L. P., Uribe-Opazo, M. A. and Vasconcellos, K. L. P. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters* **46**, 317–328. MR1743990

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

Cordeiro, G. M. and Paula, G. A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* **76**, 93–100. MR0991426

Cordeiro, G. M. and Vasconcellos, K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statistics and Probability Letters* **35**, 155–164. MR1483269

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

Key words and phrases. Bias correction, discrete distribution, maximum likelihood, scoring method, nonlinear model.

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

Doornik, J. A. (2009). *An Object-Oriented Matrix Programming Language Ox 6*. London: Timberlake Consultants.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1–26. MR0515681

Gupta, R. C. (1974). Modified power series distribution and some of its applications. *Sankhyâ B* **36**, 288–298. MR0391334

Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237

Ospina, R., Cribari–Neto, F. and Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics and Data Analysis* **51**, 960–981. MR2297500

Vasconcellos, K. L. P. and Silva, S. G. (2005). Corrected estimates for student t regression models with unknown degrees of freedom. *Journal of Statistical Computation and Simulation* **75**, 409–423. MR2154788

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 561–568
DOI: 10.1214/16-BJPS324

A note on curvature influence diagnostics in elliptical regression models

Mauricio Zevallos and Luiz Koodi Hotta

University of Campinas

Abstract. In this paper, we derive analytical expressions for the curvature influence statistic proposed by Cook [*J. Roy. Statist. Soc. Ser. B* **48** (1986) 133–169] in elliptical regression models under a data perturbation scheme. A relationship between the curvature statistics and the residuals is established and the effects of the shape parameter are assessed. The results reveal the role of the shape parameter in applying the curvature influence diagnostics technique.

References

Atkinson, A. C. (1985). *Plots, Transformations and Regression*. Oxford: Clarendon.

Carroll, R. J. and Ruppert, D. (1985). Transformations in regression: A robust analysis. *Technometrics* **27**, 1–12. MR0772893

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

Davison, A. C. and Tsai, C. L. (1992). Regression model diagnostics. *International Statistical Review* **60**, 337–353.

Fang, K. T. and Anderson, T. W. (1990). *Statistical Inferences in Elliptical Contoured and Related Distributions*. New York: Allerton Press. MR1066887

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

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

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers* **41**, 211–224. MR1769062

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* **75**, 828–838. MR0600964

Schwarzmann, B. (1991). A connection between local-influence analysis and residual diagnostics. *Technometrics* **33**, 103–104.

Zhu, H., Ibrahim, J., Lee, S. and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics* **35**, 2565–2588. MR2382658

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association* **71**, 400–405. MR0405699

Key words and phrases. Data perturbation, local influence, influential observations.

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 569–617
DOI: 10.1214/16-BJPS325

A brief tutorial on transformation based Markov Chain Monte Carlo and optimal scaling of the additive transformation

Kushal Kr. Dey[a] **and Sourabh Bhattacharya**[b]

[a]*University of Chicago*
[b]*Indian Statistical Institute*

Abstract. We consider the recently introduced Transformation-based Markov Chain Monte Carlo (TMCMC) (*Stat. Methodol.* **16** (2014) 100–116), a methodology that is designed to update all the parameters simultaneously using some simple deterministic transformation of a one-dimensional random variable drawn from some arbitrary distribution on a relevant support. The additive transformation based TMCMC is similar in spirit to random walk Metropolis, except the fact that unlike the latter, additive TMCMC uses a single draw from a one-dimensional proposal distribution to update the high-dimensional parameter. In this paper, we first provide a brief tutorial on TMCMC, exploring its connections and contrasts with various available MCMC methods.

Then we study the diffusion limits of additive TMCMC under various set-ups ranging from the product structure of the target density to the case where the target is absolutely continuous with respect to a Gaussian measure; we also consider the additive TMCMC within Gibbs approach for all the above set-ups. These investigations lead to appropriate scaling of the one-dimensional proposal density. We also show that the optimal acceptance rate of additive TMCMC is 0.439 under all the aforementioned set-ups, in contrast with the well-established 0.234 acceptance rate associated with optimal random walk Metropolis algorithms under the same set-ups. We also elucidate the ramifications of our results and clear advantages of additive TMCMC over random walk Metropolis with ample simulation studies and Bayesian analysis of a real, spatial dataset with which 160 unknowns are associated.

References

Bédard, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *The Annals of Applied Probability* **17**, 1222–1244. MR2344305

Bédard, M. (2008a). Efficient sampling using Metropolis algorithms: Applications of optimal scaling results. *Journal of Computational and Graphical Statistics* **17**, 312–332. MR2439962

Bédard, M. (2008b). Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Processes and their Applications* **118**, 2198–2222. MR2474348

Bédard, M., Douc, R. and Moulines, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Processes and their Applications* **122**, 758–786. MR2891436

Key words and phrases. Additive transformation, diffusion limit, high dimension, optimal scaling, random walk, transformation-based Markov Chain Monte Carlo.

Bédard, M. and Rosenthal, J. S. (2008). Optimal scaling of Metropolis algorithms: Heading toward general target distributions. *Canadian Journal of Statistics* **36**, 483–503. MR2532248

Bélisle, C. J. P., Romeijn, H. E. and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research* **18**, 255–266. MR1250117

Berbee, H. C. P., Boender, C. G. E., Rinnooy Kan, A. H. G., Scheffer, C. L., Smith, R. L. and Telgen, J. (1987). Hit-and-run algorithms for the identification of nonredundant linear inequalities. *Mathematical Programming* **37**, 184–207. MR0883020

Beskos, A., Roberts, G. O. and Stuart, A. M. (2009). Optimal scalings for local Metropolis–Hastings chains on non-product targets in high dimensions. *The Annals of Applied Probability* **19**, 863–898. MR2537193

Beskos, A. and Stuart, A. M. (2009). MCMC methods for sampling function space. In *ICIAM07: 6th International Congress on Industrial and Applied Mathematics* (R. Jeltsch and G. Wanner, eds.) 337–364. Zürich: European Mathematical Society. MR2588600

Christensen, O. F. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics* **15**, 1–17. MR2269360

Das, M. and Bhattacharya, S. (2016). Transdimensional transformation based Markov chain Monte Carlo. Preprint. Available at https://arxiv.org/abs/1403.5207.

Dey, K. K. and Bhattacharya, S. (2016). Adaptive transformation based Markov chain Monte Carlo. Manuscript under preparation.

Dey, K. K. and Bhattacharya, S. (2016a). On geometric ergodicity of additive and multiplicative transformation based Markov chain Monte Carlo in high dimensions. *Brazilian Journal of Probability and Statistics*. To appear. Available at https://arxiv.org/abs/1312.0915.

Dey, K. K. and Bhattacharya, S. (2016b). Supplement to “A brief tutorial on transformation based Markov Chain Monte Carlo and optimal scaling of the additive transformation.” DOI:10.1214/16-BJPS325SUPP.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* **47**, 299–350. MR1626544

Dutta, S. (2012). Multiplicative random walk Metropolis–Hastings on the real line. *Sankhya B* **74**, 315–342. MR3046902

Dutta, S. and Bhattacharya, S. (2014). Markov chain Monte Carlo based on deterministic transformations. *Statistical Methodology* **16**, 100–116. Also available at arXiv:1306.6684. Supplement available at arXiv:1106.5850. MR3110892

Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) 3–48. New York: Chapman & Hall/CRC. MR2858443

Gilks, W. R., Roberts, G. O. and George, E. I. (1994). Adaptive direction sampling. *The Statstician* **43**, 179–189.

Johnson, L. T. and Geyer, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *The Annals of Statistics* **40**, 3050–3076. MR3097969

Jourdain, B., Lelièvre, T. and Miasojedow, B. (2013). Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit. Preprint. Available at arXiv:1210.7639v2. MR3349007

Koralov, L. B. and Sinai, Y. G. (2007). *Theory of Probability and Random Processes*. New York: Springer. MR2343262

Kou, S. C., Xie, X. S. and Liu, J. S. (2005). Bayesian analysis of single-molecule experimental data. *Applied Statistics* **54**, 469–506. MR2137252

Liang, F., Liu, C. and Caroll, R. (2010). *Advanced Markov chain Monte Carlo methods: Learning from past samples*. New York: Wiley. MR2828488

Liu, J. S., Liang, F. and Wong, W. H. (2000). The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association* **95**, 121–134. MR1803145

Liu, J. S. and Sabatti, S. (2000). Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika* **87**, 353–369. MR1782484

Liu, J. S. and Yu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**, 1264–1274. MR1731488

Martino, L. and Read, J. (2013). On the flexibility of the design of multiple try Metropolis schemes. *Computational Statistics* **28**, 2797–2823. MR3141364

Mattingly, J. C., Pillai, N. S. and Stuart, A. M. (2011). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *The Annals of Applied Probability* **22**, 881–930. MR2977981

Neal, P. and Roberts, G. O. (2006). Optimal scaling for partially updating MCMC. *Algorithms. The Annals of Applied Probability* **16**, 475–515. MR2244423

Prato, G. D. and Zabczyk, J. (1992). *Stochastic Equations in Infinite Dimensions*. *Encylopedia of Mathematics and Its Applications* **44**. Cambridge: Cambridge University Press. MR1207136

Roberts, G., Gelman, A. and Gilks, W. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120. MR1428751

Roberts, G. O. and Gilks, W. R. (1994). Convergence of adaptive direction sampling. *Journal of Multivariate Analysis* **49**, 287–298. MR1276441

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statistical Science* **16**, 351–367. MR1888450

Roberts, G. O. and Rosenthal, R. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367. MR2749836

Romeijn, H. E. and Smith, R. L. (1994). Simulated annealing for constrained global optimization. *Journal of Global Optimization* **5**, 101–126. MR1291094

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theory of Probability and its Applications* **1**, 261–290. MR0084897

Smirnov, N. V. (1948). Tables for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* **19**, 279–281. MR0025109

Smith, R. L. (1996). The hit-and-run sampler: A globally reaching Markov sampler for generating arbitrary multivariate distributions. In *Proceedings of the 1996 Winter Simulation Conference* (J. M. Charnes, D. J. Morrice, D. T. Brunner and J. J. Swain, eds.), 260–264.

Storvik, G. (2011). On the flexibility of Metropolis–Hastings acceptance probabilities in auxiliary variable proposal generation. *Scandinavian Journal of Statistics* **38**, 342–358. MR2829604

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 618–639
DOI: 10.1214/16-BJPS326

Bayesian analysis of flexible measurement error models

Luz Marina Rondon[a,b] **and Heleno Bolfarine**[a]

[a]*Universidade de São Paulo*
[b]*Universidad Nacional de Colombia*

Abstract. This paper proposes the Bayesian inference for flexible measurement error models, in which their systematic components include explanatory variable vectors with and without measurement errors, as well as nonlinear effects that are approximated by using B-splines. The model investigated is the structural version, as the error-prone variables follow scale mixtures of normal distributions such as Student-t, slash, contaminated normal, Laplace and symmetric hyperbolic distributions. To draw samples of the posterior distribution of the model parameters, an MCMC algorithm is proposed. The performance of this algorithm is assessed through simulations. In addition, the function `fmem()` of the R package **BayesGESM** is presented, which provides an easy way to apply the methodology presented in this paper. The proposed methodology is applied to a real data set, which shows that ignoring measurement errors (i.e., analyze the data by using the traditional methodology) can lead to wrong conclusions.

References

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

Arellano-Valle, R. B., Bolfarine, H. and Labra, V. (1996). Ultrastructural elliptical models. *Canadian Journal of Statistics* **24**, 207–216. MR1406176

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society. Series A, Mathematical, Physical and Engineering Sciences* **353**, 401–419.

Belsley, D. A., Kuh, E. and Welsch, R. E. (2005). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley. MR0576408

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.

Cao, C. Z., Lin, J. G. and Zhu, X. X. (2012). On estimation of a heteroscedastic measurement error model under heavy-tailed distributions. *Computational Statistics and Data Analysis* **56**, 438–448. MR2842351

Carroll, R. J., Roeder, K. and Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics* **55**, 44–54.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. Boca Raton: Chapman and Hall. MR2243417

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

Key words and phrases. Bayesian analysis, measurement error models, semi-parametric models, MCMC algorithm, B-splines, scale mixtures of normal distributions.

De Boor, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences.* New York: Springer. MR0507062

de Castro, M., Bolfarine, H. and Galea, M. (2013). Bayesian inference in measurement error models for replicated data. *Environmetrics* **24**, 22–30. MR3042271

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121. MR1435485

Fuller, M. A. (1987). *Measurement Error Models*. New York: Wiley. MR0898653

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*, 2nd ed. Boca Raton, FL: Chapman and Hall. MR2260716

Gelfand, A., Dey, D. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics* **4**, 147–167. MR1380275

Harrison, D. Jr. and Rubinfeld, D. L (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management* **5**, 81–102.

He, X., Fung, W. K. and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association* **100**, 1176–1184. MR2236433

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

Kelly, B. C. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal* **665**, 1489–1506.

Kulathinal, S. B., Kuulasmaa, K. and Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement errors vary betwwen the observations. *Statistics in Medicine* **21**, 1089–1101.

Lemonte, A. J. and Patriota, A. G. (2011). Multivariate elliptical models with general parameterization. *Statistical Methodology* **8**, 389–400. MR2800359

Li, L., Palta, M. and Shao, J. (2004). A measurement error model with a Poisson distributed surrogate. *Statistics in Medicine* **23**, 2527–2536.

Maronna, R. A., Martin, D. R. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. New York: Wiley. MR2238141

Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae* **92**, 97–111. MR2265333

Patriota, A. G., Bolfarine, H. and de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology* **6**, 408–423. MR2751083

Peng, F. and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics* **23**, 199–213.

Rondon, L. M. and Bolfarine, H. (2014). BayesGESM: Bayesian Analysis of Generalized Elliptical Semiparametric Models. R package version 1.1. http://CRAN.R-project.org/package=BayesGESM.

Rogers, W. H. and Tukey, J. W. (1972). Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica* **26**, 211–226. MR0383605

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Var der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Methodological)* **64**, 583–640. MR1979380

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550. MR0898357

Weiss, R. and Cook, R. (1992). A grafical case statistic for assessing posterior influence. *Biometrika* **79**, 51–55.

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 640–652
DOI: 10.1214/16-BJPS327

Boosting, downsizing and optimality of test functions of Markov chains

Thomas R. Boucher

Department of Mathematics, Texas A&M University-Commerce

Abstract. Test functions play an important role in Markov chain theory. Stability of a Markov chain can be demonstrated by constructing a test function of the chain that satisfies a stochastic drift criterion. The test function defines a class of functions of the process for which limit laws hold, yields bounds on the convergence of the Markov chain transition probabilities to the stationary distribution, and provides information concerning the mixing properties of the chain. Under certain conditions, these results can be improved by using a new test function derived from a known test function of a Markov chain.

References

Borovkov, A. A. and Hordijk, A. (2004). Characterization and sufficient conditions for normed ergodicity of Markov chains. *Adv. in Appl. Probab.* **36**, 227–242. MR2035781

Boucher, T. R. and Cline, D. B. H. (2007). Stability of cyclic threshold autoregressive time series models. *Statist. Sinica* **17**, 43–62. MR2352503

Chan, K. S., Petruccelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22**, 267–279. MR0789351

Cline, D. B. H. and Pu, H. (2001). Stability of nonlinear time series: What does noise have to do with it? In *Selected Proceedings of the Symposium on Inference for Stochastic Processes, Vol. 37*, 151–170. MR2002508

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101–121. MR1389882

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. London: Springer. MR1287609

Petruccelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *J. Appl. Probab.* **21**, 270–286. MR0741130

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1**, 20–71. MR2095565

Rosenthal, J. S. (2003). Asymptotic variance and convergence rates of nearly-periodic MCMC algorithms. *J. Amer. Statist. Assoc.* **98**, 169–177. MR1965683

Key words and phrases. Markov chain, convergence, ergodicity, mixing, test function.

Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 653–665
DOI: 10.1214/16-BJPS328

Second-order autoregressive Hidden Markov Model

Daiane Aparecida Zuanetti and Luis Aparecido Milan

Departamento de Estatística, UFSCar—Brazil

Abstract. We propose an extension of Hidden Markov Model (HMM) to support second-order Markov dependence in the observable random process. We propose a Bayesian method to estimate the parameters of the model and the non-observable sequence of states. We compare and select the best model, including the dependence order and number of states, using model selection criteria like Bayes factor and deviance information criterion (DIC). We apply the procedure to several simulated datasets and verify the good performance of the estimation procedure. Tests with a real dataset show an improved fitting when compared with usual first order HMMs demonstrating the usefulness of the proposed model.

References

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics* **37**, 1554–1563. MR0202264

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* **41**, 164–171. MR0287613

Biblio, M., Monfort, A. and Robert, C. P. (1999). Bayesian estimation of switching ARMA models. *Journal of Econometrics* **93**, 229–255. MR1721099

Boys, R. and Henderson, D. (2002). On determining the order of Markov dependence of an observed process governed by a hidden Markov model. *Scientifc Programming* **10**, 241–251.

Boys, R. and Henderson, D. (2004). A Bayesian approach to DNA sequence segmentation. *Biometrics* **60**, 573–588. MR2089432

Boys, R., Henderson, D. and Wilkinson, D. (2000). Detecting homogeneous segments in DNA sequences by using hidden Markov models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **49**, 269–285. MR1765825

Braun, J. V., Braun, R. K. and Muller, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika* **87**, 301–314. MR1782480

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75**, 79–97. MR1414504

Churchill, G. (1989). Stochastic models for heterogeneous DNA sequences. *Bulletin of Mathematical Biology* **51**, 79–94. MR0978904

Churchill, G. (1992). Hidden Markov chains and the analysis of genome structure. *Computers and Chemistry* **16**, 107–115.

da-Silva, C. Q. (2003). Hidden Markov models applied to a subsequence of the *Xylella fastidiosa* genome. *Genetics and Molecular Biology* **26**, 529–535.

Key words and phrases. Hidden Markov model, second-order dependence, Markov chain Monte Carlo (MCMC), gene modeling, bacteriophage *lambda* genome.

Djurić, P. M., Kotecha, J. H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F. and Miguez, J. (2003). Particle filtering. *Signal Processing Magazine, IEEE* **20**, 19–38.

Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Media: Springer.

Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering* **12**, 656–704.

du Preez, J. A. (1998). Efficient higher-order hidden Markov modeling. Ph.D. thesis, University of Stellenbosch. Available: www.ussigbase.org/downloads/jadp_phd.pdf.

Gassiat, E. and Kéribin, C. (2000). The likelihood ratio test for the number of components in a mixture with Markov regime. *ESAIM. Probabilités Et Statistique* **4**, 25–52. MR1780964

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **17**, 457–472.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology* **313**, 903–919.

Hadar, U. and Messer, H. (2009). High-order hidden Markov models—estimation and implementation. In *Statistical Signal Processing*. *Cardiff, UK*. doi:10.1109/SSP.2009.5278591.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrika* **57**, 357–384. MR0996941

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions. Lecture Notes in Economic and Mathematical Systems 454*. New York: Springer. MR1473720

Leea, S. Y., Leea, J. Y., Jungb, K. S. and Ryu, K. H. (2009). A 9-state hidden Markov model using protein secondary structure information for protein fold recognition. *Computers in Biology and Medicine* **39**, 527–534.

Martino, L., Read, J., Elvira, V. and Louzada, F. (2015). Cooperative parallel particle filters for online model selection and applications to urban mobility. Available at viXra:1512.0420.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society. Series C* **36**, 318–324.

Muri, F. (1998). Modelling bacterial genomes using hidden Markov models. In *COMPSTAT'98 Proceedings in Computational Statistics* (R. W. Payne and P. J. Green, eds.) 89–100. Heidelberg: Physica.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257–285.

Ristic, B., Arulampalam, S. and Gordon, N. J. (2004). Beyond the Kalman filter: Particle filters for tracking applications. Artech house.

Robert, C. P., Celeux, G. and Diebolt, J. (1993). Bayesian estimation of hidden Markov chains: A stochastic implementation. *Statist. Prob. Letters* **16**, 77–83. MR1208503

Robert, C. P. and Titterington, D. M. (1998). Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. *Statistics and Computing* **8**, 145–158.

Rydén, T., Teräsvirta, T. and Asbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* **13**, 217–244.

Schimert, J. (1992). A high order hidden Markov model. Ph.D. thesis, University of Washington. MR2688816

Seifert, M. (2010). Extensions of Hidden Markov Models for the analysis of DNA microarray data. Ph.D. thesis, University of Halle-Wittenberg. Available at http://nbn-resolving.de/urn:nbn:de:gbv:3:4-4110.

Seifert, M., Abou-El-Ardat, K., Friedrich, B., Klink, B. and Deutsch, A. (2014). Autoregressive higher-order hidden Markov models: Exploiting local chromosomal dependencies in the analysis of tumor expression profiles. *PLoS ONE* **9**, e100295.

Seifert, M., Gohr, A., Strickert, M. and Grosse, I. (2012). Parsimonious higher-order hidden Markov models for improved array-CGH analysis with applications to Arabidopsis thaliana. *PLoS Computational Biology* **8**, e1002286.

Skalka, A., Burgi, E. and Hershey, A. D. (1968). Segmental distribution of nucleotides in the DNA of Bacteriophage lambda. *Journal of Molecular Biology* **34**, 1–16.

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960.

Spiegelhalter, D., et al. (2002). Bayesian measures of model complexity and fit. *Royal Statistical Society* **64**, 583–639. MR1979380