Director, Hannah M.; **Gattiker, James**; **Lawrence, Earl**; **van der Wiel, Scott**

Efficient sampling on the simplex with a self-adjusting logit transform proposal. (English)

Zbl 07192130
J. Stat. Comput. Simulation 87, No. 18, 3521-3536 (2017).

Summary: A vector of k positive coordinates lies in the k-dimensional simplex when the sum of all coordinates in the vector is constrained to equal 1. Sampling distributions efficiently on the simplex can be difficult because of this constraint. This paper introduces a transformed logit-scale proposal for Markov Chain Monte Carlo that naturally adjusts step size based on the position in the simplex. This enables efficient sampling on the simplex even when the simplex is high dimensional and/or includes coordinates of differing orders of magnitude. Implementation of this method is shown with the SALTSampler R package and comparisons are made to other simpler sampling schemes to illustrate the improvement in performance this method provides. A simulation of a typical calibration problem also demonstrates the utility of this method.

MSC:

62 Statistics

Keywords:

simplex; Metropolis-Hastings algorithm; transformations; compositional data

Software:

BayesDA; compositions; hitandrun; R; SALTSampler; Stan

Full Text: DOI

References:

[1] Aitchison J.The statistical analysis of compositional data. J R Stat Soc Ser B Methodol. 1982;44(2):139-177. [Google Scholar] · Zbl 0491.62017

[2] Aitchison J. The statistical analysis of compositional data. London: Chapman and Hall; 1986. [Crossref], [Google Scholar] · Zbl 0688.62004

[3] Pawlowsky-Glahn V, Egozcue JJ.Geometric approach to statistical analysis on the simplex. Stoch Environ Res Risk Assess. 2001;15(5):384-398. doi: 10.1007/s004770100077[Crossref], [Web of Science ®], [Google Scholar] · Zbl 0987.62001

[4] Pawlowsky-Glahn V, Buccianti A. Compositional data analysis: theory and applications. Chichester: John Wiley \& Sons; 2011. [Crossref], [Google Scholar] · Zbl 1103.62111

[5] Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Modeling and analysis of compositional data. Chichester: John Wiley \& Sons; 2015. [Google Scholar]

[6] Van den Boogaart KG, Tolosana-Delgado R. Analyzing compositional data with R. Berlin: Springer; 2013. [Crossref], [Google Scholar] · Zbl 1276.62011

[7] van den Boogaart KG, Tolosana-Delgado R.Compositions: a unified R package to analyze compositional data. Comput Geosci. 2008;34(4):320-338. doi: 10.1016/j.cageo.2006.11.017[Crossref], [Web of Science ®], [Google Scholar]

[8] Billheimer D, Guttorp P, Fagan WF.Statistical interpretation of species composition. J Amer Statist Assoc. 2001;96(456):1205-1214. doi: 10.1198/016214501753381850[Taylor \& Francis Online], [Web of Science ®], [Google Scholar] · Zbl 1073.62573

[9] Thomas CW, Aitchison J.Compositional data analysis of geological variability and process: a case study. Math Geol. 2005;37(7):753-772. doi: 10.1007/s11004-005-7378-4[Crossref], [Google Scholar] · Zbl 1151.86307

[10] Fry JM, Fry TRL, McLaren KR.Compositional data analysis and zeros in micro data. Appl Econ. 2000;32(8):953-959. doi: 10.1080/000368400322002[Taylor \& Francis Online], [Web of Science ®], [Google Scholar]

[11] Aitchison J, Greenacre M.Biplots of compositional data. J R Stat Soc Ser C. 2002;51(4):375-392. doi: 10.1111/1467-9876.00275[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1111.62300

[12] Filzmoser P, Hron K, Reimann C.Principal component analysis for compositional data with outliers. Environmetrics. 2009;20(6):621-632. doi: 10.1002/env.966[Crossref], [Web of Science ®], [Google Scholar]

[13] Geyer C. Introduction to Markov chain Monte Carlo. In: Brooks S, Gelman A, Jones G, Meng X-L, editors. Handbook of

Markov chain Monte Carlo. Boca Raton (FL): Chapman \& Hall/ CRC Press; 2011. p. 3-48. [Google Scholar]

[14] Stan Development Team. Stan modeling language users guide and reference manual, Version 2.5.0, 2014. Available from: http://mc-stan.org/users/documentation/. [Google Scholar]

[15] Betancourt M. Cruising the simplex: Hamiltonian Monte Carlo and the Dirichlet distribution. In: 31st AIP conference proceedings 31st, AIP. Vol. 1443; 2012. p. 157-164. [Google Scholar]

[16] van Valkenhoef G, Tervonen T. hitandrun: 'Hit and Run' and 'Shake and Bake' for sampling uniformly from convex shapes, 2015. R package version 0.5-1. Available from: https://cran.r-project.org/web/packages/hitandrun/index.html. [Google Scholar]

[17] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. [Google Scholar]

[18] Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. Vol. 2. Boca Raton (FL): CRC/Chapman Hall; 2014. [Google Scholar] · Zbl 1279.62004

[19] Kass RE, Carlin BP, Gelman A, et al. Markov chain Monte Carlo in practice: a roundtable discussion. Am Stat. 1998;52(2):93-100. [Taylor \& Francis Online], [Web of Science ®], [Google Scholar]

[20] Bayarri MJ, Berger JO, Paulo R, et al. A framework for validation of computer models. Technometrics. 2012;49(2):138-154. doi: 10.1198/004017007000000092[Taylor \& Francis Online], [Web of Science ®], [Google Scholar]

[21] Kennedy MC, O'Hagan A.Bayesian calibration of computer models. J R Stat Soc Ser B Statist Methodol. 2001;63(3):425-464. doi: 10.1111/1467-9868.00294[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1007.62021

[22] Jasra A, Stephens DA, Holmes CC.Population-based reversible jump Markov chain Monte Carlo. Biometrika. 2007;94(4):787-807. doi: 10.1093/biomet/asm069[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1158.62019

[23] Liang F, Wong WH.Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. J Amer Statist Assoc. 2001;96(454):653-666. doi: 10.1198/016214501753168325[Taylor \& Francis Online], [Web of Science ®], [Google Scholar] · Zbl 1017.62022

[24] Martino L, Elvira V, Luengo D, et al. Orthogonal parallel MCMC methods for sampling and optimization. Digit Signal Process. 2016;58:64-84. doi: 10.1016/j.dsp.2016.07.013[Crossref], [Web of Science ®], [Google Scholar]

[25] Jasra A, Stephens DA, Holmes CC.On population-based simulation for static inference. Stat Comput. 2007;17(3):263-279. doi: 10.1007/s11222-007-9028-9[Crossref], [Web of Science ®], [Google Scholar]

[26] Chopin N.A sequential particle filter method for static models. Biometrika. 2002;89(3):539-552. doi: 10.1093/biomet/89.3.539[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1036.62062