

**Mossel, Elchanan; Vigoda, Eric**

**Limitations of Markov chain Monte Carlo algorithms for Bayesian inference of phylogeny.**

(English) [Zbl 1121.60078](#)

*Ann. Appl. Probab.* 16, No. 4, 2215-2234 (2006).

Summary: Markov chain Monte Carlo algorithms play a key role in the Bayesian approach to phylogenetic inference. We present the first theoretical work analyzing the rate of convergence of several Markov chains widely used in phylogenetic inference. We analyze simple, realistic examples where these Markov chains fail to converge quickly. In particular, the data studied are generated from a pair of trees, under a standard evolutionary model. We prove that many of the popular Markov chains take exponentially long to reach their stationary distribution. Our construction is pertinent since it is well known that phylogenetic trees for genes may differ within a single organism. Our results shed a cautionary light on phylogenetic analysis using Bayesian inference and highlight future directions for potential theoretical work.

**MSC:**

[60J10](#) Markov chains (discrete-time Markov processes on discrete state spaces)

Cited in **6** Documents

[92D15](#) Problems related to evolution

**Keywords:**

Markov chain Monte Carlo; phylogeny; tree space

**Software:**

[MrBayes](#)

**Full Text:** [DOI](#)

**References:**

- [1] Bhatnagar, N. and Randall, D. (2004). Torpid mixing of simulated tempering on the Potts model. In Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms ( SODA ) 478–487. · [Zbl 1318.82023](#)
- [2] Cavender, J. A. (1978). Taxonomy with confidence. *Math. Biosci.* 40 271–280. · [Zbl 0391.92015](#) · [doi:10.1016/0025-5564\(78\)90089-5](#)
- [3] Chor, B., Hendy, M. D., Holland, B. R. and Penny, D. (2000). Multiple maxima of likelihood in phylogenetic trees: An analytic approach. *Mol. Biol. Evol.* 17 1529–1541.
- [4] Durbin, R., Eddy, S., Krogn, A. and Mitchison, G. (1998). *Biological Sequence Analysis : Probabilistic Models of Proteins and Nucleic Acids* . Cambridge Univ. Press. · [Zbl 0929.92010](#) · [doi:10.1017/CBO9780511790492](#)
- [5] Dyer, M., Frieze, A. and Jerrum, M. (2002). On counting independent sets in sparse graphs. *SIAM J. Comput.* 31 1527–1541. · [Zbl 1041.68045](#) · [doi:10.1137/S0097539701383844](#)
- [6] Diaconis, P. and Holmes, S. P. (2002). Random walks on trees and matchings. *Electron. J. Probab.* 7 . · [Zbl 1007.60071](#)
- [7] Develin, M. and Sturmfels, B. (2004). Tropical convexity. *Doc. Math.* 9 1–27. · [Zbl 1054.52004](#)
- [8] Farris, J. S. (1973). A probability model for inferring evolutionary trees. *Syst. Zool.* 22 250–256.
- [9] Felsenstein, J. (2004). *Inferring Phylogenies* . Sinauer Associates, Inc., Sunderland, MA.
- [10] Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. *Computing Science and Statistics : Proc. 23rd Symp. on the Interface* 156–163. Interface Foundation, Fairfax Station, VA.
- [11] Graur, D. and Li, W.-H. (1999). *Fundamentals of Molecular Evolution* , 2nd ed. Sinauer Associates, Inc., Sunderland, MA.
- [12] Huelsenbeck, J. P., Larget, B., Miller, R. E. and Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst Biol.* 51 673–688.
- [13] Huelsenbeck, J. P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17 754–755.
- [14] Huelsenbeck, J. P., Ronquist, F., Nielsen, R. and Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294 2310–2314.
- [15] Janson, S., Łuczak, T. and Ruciński, A. (2000). *Random Graphs* . Wiley, New York.
- [16] Larget, B. and Simon, D. L. (1999). Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16 750–759.
- [17] Li, S., Pearl, D. K. and Doss, H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* 95 493–508.

- [18] Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford Univ. Press.
- [19] Neyman, J. (1971). Molecular studies of evolution: A source of novel statistical problems. In *Statistical Decision Theory and Related Topics* (S. S Gupta and J. Yackel, eds.) 1–27. Academic Press, New York. · [Zbl 0231.62010](#)
- [20] Mossel, E. and Vigoda, E. (2005). Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* 309 2207–2209.
- [21] Rannala, B. and Yang, Z. (1996). Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J. Mol. Evol.* 43 304–311.
- [22] Simon, D. L. and Larget, B. (2000). Bayesian analysis in molecular biology and evolution (BAMBE). Version 2.03 beta, Dept. Mathematics and Computer Science, Duquesne Univ., Pittsburgh, PA.
- [23] Speyer, D. and Sturmfels, B. (2004). The tropical Grassmannian. *Adv. Geom.* 4 389–411. · [Zbl 1065.14071](#) · [doi:10.1515/adv.2004.023](#)
- [24] Yang, Z. (2000). Complexity of the simplest phylogenetic estimation problem. *Proc. R. Soc. Lond. B Biol. Sci.* 267 109–116.
- [25] Yang, Z. and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14 717–724.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.