Benchmarking MCMC Samplers on Challenging Synthetic Posteriors

Fabian Kohrs
University of Lübeck

Abstract

Bayesian inference provides a statistical method for building models under uncertainty. In practice, applications often rely on Markov Chain Monte Carlo (MCMC) methods whose performance depends strongly on the geometry of the posterior distribution. Attributes such as multimodality, correlation, dimensionality, and tail weight can substantially affect sampler efficiency and accuracy, yet a comprehensive understanding of their effects across different samplers remains limited. Existing benchmarking frameworks, such as PosteriorDB [2] and MCBench [1], have advanced reproducibility and comparability but either lack systematic control of posterior attributes or cover only a limited set of target distributions.

This thesis introduces MC-FiT (Monte Carlo Fixed Target), a synthetic benchmarking framework that enables direct control of posterior geometry and provides independent and identically distributed reference samples for evaluation. Using MC-FiT, a series of experiments are conducted to investigate how individual attributes and selected combinations influence MCMC sampler performance. Performance is assessed in terms of distributional accuracy, convergence diagnostics, and computational efficiency.

The results provide empirical insights into how different samplers respond to challenging posterior structures, highlight attribute thresholds where performance degrades sharply, and suggest practical guidelines for aligning sampler choice with anticipated posterior characteristics. Beyond these findings, MC-FiT establishes a reusable benchmark framework that supports systematic, reproducible evaluation of MCMC methods.

References

[1] Z. Ding, C. Grunwald, K. Ickstadt, K. Kröninger, and S. L. Cagnina, *Mcbench: A benchmark suite for monte carlo sampling algorithms*, 2025. arXiv: `2501.03138 [stat.CO]`. [Online]. Available: `https://arxiv.org/abs/2501.03138`.

[2] M. Magnusson, J. Torgander, P.-C. Bürkner, L. Zhang, B. Carpenter, and A. Vehtari, *Posteriordb: Testing, benchmarking and developing bayesian inference algorithms*, 2024. arXiv: 2407.04967 [stat.CO]. [Online]. Available: https://arxiv.org/abs/2407.04967.