Scalable Automation of Monte Carlo methods

Luke Bornn[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 6

[1] Harvard University, Cambridge, MA 02138
bornn@stat.harvard.edu

Abstract

While statisticians are well-accustomed to performing exploratory analysis in the modeling stage of an analysis, the notion of conducting preliminary general-purpose exploratory analysis in the Monte Carlo stage (or more generally, the model-fitting stage) of an analysis is an area which we feel deserves much further attention. Towards this aim, this paper proposes a general-purpose algorithm for automatic density exploration. The proposed exploration algorithm combines and expands upon components from various adaptive Markov chain Monte Carlo methods, with the Wang-Landau algorithm at its heart. Additionally, the algorithm is run on interacting parallel chains – a feature which both decreases computational cost as well as stabilizes the algorithm, improving its ability to explore the density. Performance of this new parallel adaptive Wang-Landau (PAWL) algorithm is studied in several applications. Through a Bayesian variable selection example, the authors demonstrate the convergence gains obtained with interacting chains. The ability of the algorithm's adaptive proposal to induce mode-jumping is illustrated through a Bayesian mixture modeling application. Lastly, through a 2D Ising model, the authors demonstrate the ability of the algorithm to overcome the high correlations encountered in spatial models. Supplemental materials are available online.

Keywords: Markov chain Monte Carlo, Wang-Landau, Parallelization, Adaptive Monte Carlo

1 Introduction

The Wang-Landau algorithm generates a time-inhomogeneous Markov chain that admits a distribution $\tilde{\pi}_t$ as the invariant distribution at iteration t. The biased distribution $\tilde{\pi}_t$ targeted by the algorithm at iteration t is based on the

target distribution π, and modified such that a) the generated chain visits all the sets $(\mathcal{X}_i)_{i=1}^d$ equally, that is the proportion of visits in each set is converging to d^{-1} when t goes to infinity; and b) the restriction of the modified distribution $\tilde{\pi}_t$ to each set $\mathcal{X}_i$ coincides with the restriction of the target distribution π to this set, up to a multiplicative constant. The modification (a) is crucial, as inducing uniform exploration of the sets is the biasing mechanism which improves exploration. Ideally the biased distribution $\tilde{\pi}$ would not depend on t, and would be available analytically as:

$$\tilde{\pi}(x) = \pi(x) \times \frac{1}{d} \sum_{i=1}^{d} \frac{\mathcal{I}_{\mathcal{X}_i}(x)}{\psi(i)} \tag{1}$$

where $\psi(i) = \int_{\mathcal{X}_i} \pi(x)\mathrm{d}x$ and $\mathcal{I}_{\mathcal{X}_i}(x)$ is equal to 1 if $x \in \mathcal{X}_i$ and 0 otherwise. Checking that using $\tilde{\pi}$ as the invariant distribution of a MCMC algorithm would validate points a) and b) is straightforward. Figure 1 illustrates a univariate target distribution π and its corresponding biased distribution $\tilde{\pi}$ under two different partitions of the state space. In this paper, we explore the Wang-Landau algorithm as an automatic density exploration method, parallelizing it and developing an adaptive binning strategy and proposal mechanism to minimize user input.

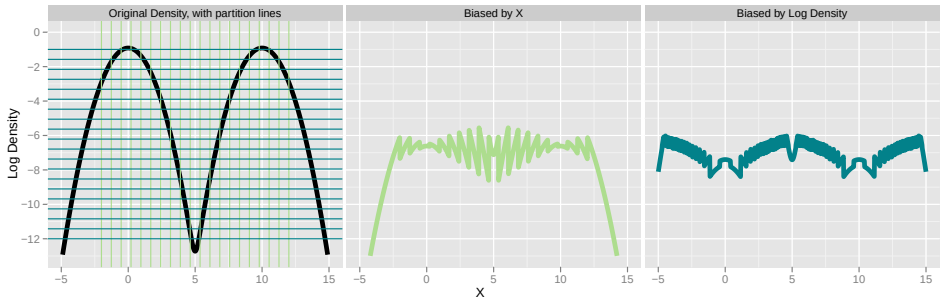

Figure 1: Probability density functions for a univariate distribution π and its biased version $\tilde{\pi}$ when partitioning the state space along the x-axis ($\xi(x) = x$, middle) and the log density ($\xi(x) = -\log \pi(x)$, right). The left-most plot also shows the partitioning of the state space with $\xi(x)$; in both cases $d = 20$. The biasing is done such that the integral $\int_{\mathcal{X}_i} \tilde{\pi}(x)\mathrm{d}x$ is the same for all $\mathcal{X}_i$ (areas under the curve for each set) and such that π and $\tilde{\pi}$ coincide on each set $\mathcal{X}_i$, up to a multiplicative constant.