A review of nested sampling

Abstract
Consensus Monte Carlo [1] was introduced as a general-purpose mechanism for distributed Bayesian computation for continuous data sets. In a nutshell, independent shards of data are first assigned to a set of workers, which sample from the corresponding subposterior via MCMC. Then, the obtained samples are communicated to a central server and aggregated to obtain an approximation of the resulting posterior. The aggregation scheme assumes (based on the Berstein-von Mises theorem) that the resulting posterior is approximately multivariate Gaussian.

1. Introduction

Multi-threaded algorithms often require specialized programming skills and the development of computer programs that are challenging even for expert software engineers. On the other hand, multi-machine code can be easily scaled and executed without any algorithmic modification of the program. In this case, however, the cost of between-machine communication and program reinialization due to catasthropic failures in hardware becomes a bottleneck. To address these issues, embarrassingly parallel algorithms enable distributed Bayesian inference with minimal communication.

2. Method

Firstly, we partition the posterior distribution as

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \prod_{1 \leq s \leq S} p(\boldsymbol{y}_s \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})^{\frac{1}{S}}. \tag{1}$$

Then, each worker $s \in [[1, S]]$ samples G parameters $\theta_{s1}, ..., \theta_{sG}$ from the corresponding subposterior distribution, $p(\boldsymbol{y}_s \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})^{\frac{1}{S}}$. Finally, these samples are aggregated according to the rule for $g \in [[1, G]]$

$$\boldsymbol{\theta}_g = \left(\sum_s W_s \right)^{-1} \sum W_s \theta_{sg}, \tag{2}$$

in which $W_s = \Sigma_s^{-1}$ are weights fixed as the inverse of the covariance matrix of each worker. For Gaussian models, this approach yields asymptotically exact samples. For non-Gaussian models, the algorithm is inherently biased. For hierarchical models, the within-group samples should not be partitioned to avoid issues emanating from non-independent samples.

3. Empirical analysis

In spite of the posited Gaussian nature of the posterior distribution, which is roughly assured by Berstein-von Mises theorem under sufficiently regular conditions, the resulting aggregated distribution is shown to accurately approximate even non-Gaussian targets. The absence of theoretical assurances, however, is quite troublesome for the wide acceptance of such a method.

Bibliography

[1] S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch, “Bayes and big data: the consensus Monte Carlo algorithm,” Int. J. Manag. Sci. Eng. Manag., vol. 11, no. 2, pp. 78–88, Apr. 2016.