

Sequential Monte Carlo to Estimate the Partition Function of Rectangular Graphical Models

Christian A. Naesseth¹, Fredrik Lindsten², Thomas Schön³

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

¹ Linköping University, Department of Electrical Engineering, Linköping, Sweden
`christian.a.naesseth@liu.se`

² University of Cambridge, Department of Engineering, Cambridge, United Kingdom
`fredrik.lindsten@eng.cam.ac.uk`

³ Uppsala University, Department of Information Technology, Uppsala, Sweden
`thomas.schon@it.uu.se`

Abstract

We propose a new sequential-Monte-Carlo-based sampler to estimate the partition function of certain graphical models. The sampler is profiled against a state-of-the-art method on a problem from information theory, calculating the capacity of the finite-size 2-D $(1, \infty)$ run-length limited constrained channel.

Keywords: discrete graphical models; normalization constant; 2-D channel capacity

1 Introduction

Estimating the partition function (normalization constant) of probabilistic graphical models (PGM) is ubiquitous in applications of Bayesian statistics. To give a few examples, it relates to the Bayes factor used in model comparison, evaluation methods of models on previously unseen data and derived quantities such as the capacity of communication channels.

Here we propose to use a fully adapted sequential Monte Carlo (SMC) algorithm to estimate the partition function of rectangular graphical models with discrete variables. We show how to make this algorithm computationally efficient by combining the SMC sampler with ideas from Forward Filtering/Backward Sampling (FF/BS) proposed in [1, 2]. Our example comes from information theory, estimating the noiseless capacity of the 2-D $(1, \infty)$ run-length limited constrained channel. We profile our algorithm against a state-of-the-art Monte Carlo estimation method proposed in [3].

2 Method

We consider rectangular PGM with pair-wise interaction between discrete variables. That means that the joint probability mass function (PMF) of the set of random variables, $X := \{x_{1,1}, \dots, x_{1,J}, x_{2,J}, \dots, x_{M,J}\}$, can be represented as a product of factors

over the pairs of variables in the graph:

$$p(X) = \frac{1}{Z} \prod_{(\ell j, mn) \in \mathcal{E}} \psi(x_{\ell, j}, x_{m, n}). \quad (1)$$

Here, Z —the partition function—is given by

$$Z = \sum_X \prod_{(\ell j, mn) \in \mathcal{E}} \psi(x_{\ell, j}, x_{m, n}), \quad (2)$$

and $\psi(x_{\ell, j}, x_{m, n})$ denotes the so-called potential function encoding the pair-wise interaction between $x_{\ell, j}$ and $x_{m, n}$.

It has been shown that the constrained channel can be modeled as a square lattice PGM [3], which is a special case of (1) with $J = M$. The channel consists of random variables $x_{\ell, j} \in \{0, 1\}$ with a constraint that no two horizontally or vertically adjacent bits may both equal 1, i.e. $\psi(x_{\ell, j}, x_{m, n}) = 0$ if $x_{\ell, j} = x_{m, n} = 1$ else $\psi(x_{\ell, j}, x_{m, n}) = 1$. The finite-size capacity C_M is can then be described as a mapping of the partition function,

$$C_M = \frac{1}{M^2} \log_2 Z. \quad (3)$$

However, computing Z is intractable for these types of models; whereby we propose to estimate it using a fully adapted SMC [5]. That the sampler is *fully adapted* means that the proposal distributions for the resampling and propagation steps are optimally chosen with respect to minimizing the variance of the N importance weights. We define \mathbf{x}_k to be the M -dimensional variable corresponding to all original variables in column k ,

$$\mathbf{x}_k = \{x_{1, k}, \dots, x_{M, k}\}, \quad k = 1, \dots, M. \quad (4)$$

This results in an undirected chain giving us a natural sequence of target distributions

$$\gamma_1(\mathbf{x}_1) = \phi(\mathbf{x}_1), \quad (5a)$$

$$\gamma_k(\mathbf{x}_{1:k}) = \gamma_{k-1}(\mathbf{x}_{1:k-1}) \phi(\mathbf{x}_k) \psi(\mathbf{x}_k, \mathbf{x}_{k-1}), \quad (5b)$$

where $\phi(\cdot), \psi(\cdot)$ are the in-column and between-column interactions respectively. We illustrate this in Figure 1 for $M = 6$. Using a standard implementation of the SMC sampler is in this case computationally prohibitive, the complexity grows as $\mathcal{O}(NM2^M)$. The key enabler for our approach is observing that conditionally on previous iterations our proposal distribution, $\phi(\mathbf{x}_k) \psi(\mathbf{x}_k, \mathbf{x}_{k-1})$, is described by a chain. Sampling and calculating the weights can be performed efficiently using FF/BS, resulting in a complexity of $\mathcal{O}(NM^2)$.

For a full derivation of the algorithm and discussions see the companion paper [4].

3 Experiments

We compare the proposed algorithm to the state-of-the-art Monte Carlo approximation algorithm proposed in [3], in the sequel referred to as tree sampler. We display, for

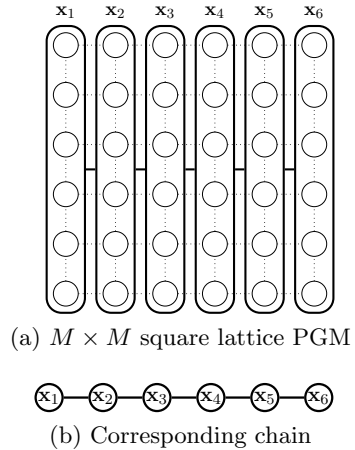


Figure 1: Square lattice graphical model converted to an undirected chain.

$M = 60$, the estimated mean square error (MSE) of the capacity in Figure 2a based on 10 independent runs of the algorithms. The proposed SMC sampler performs very well and on average has more than an order-of-magnitude smaller error than the tree sampler.

To see how our proposed method scales with size M we fix the number of particles $N = 1000$ and estimate the variance of $\log_2(\hat{Z})$ based on 20 independent runs for $M \in \{2, 4, \dots, 60\}$. The results seem to indicate that the variance scales approximately as M^3 , see Figure 2b.

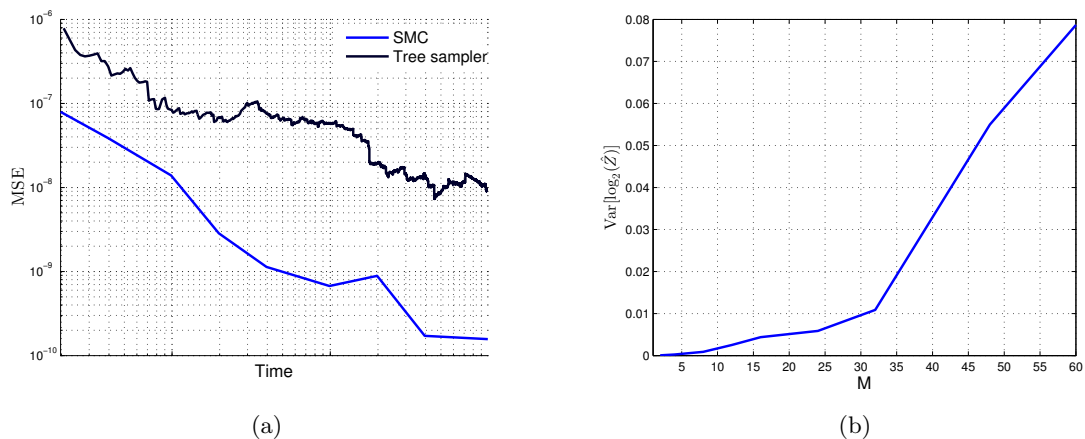


Figure 2: *a)* C_{60} estimated MSE based on 10 independent runs of the SMC and tree samplers. Plotted versus wall-clock time in loglog-scale. *b)* Estimated variance of the SMC estimator $\log_2(\hat{Z})$ plotted versus square lattice size M .

References

- [1] Chris K Carter and Robert Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.
- [2] Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.
- [3] M. Molkaiaie and H.-A. Loeliger. Monte Carlo algorithms for the partition function and information rates of two-dimensional channels. *IEEE Transactions on Information Theory*, 59(1):495–503, 2013.
- [4] C. A. Naesseth, F. Lindsten, and T. B. Schön. Capacity estimation of two-dimensional channels using Sequential Monte Carlo. *ArXiv e-prints 1405.0102*, 2014.
- [5] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.