Proceedings of the 1996 Conference on Information Sciences and Systems, Princeton, NJ, vol. II, pp. 1212-1215, 1996

EFFICIENT MONTE CARLO ESTIMATION OF PARTITION FUNCTION RATIOS OF MARKOV RANDOM FIELDS

Gerasimos Potamianos

Center for Language and Speech Processing
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218, U.S.A.
e-mail: makis@cspjhu.ece.jhu.edu

Abstract

The problem of estimating the partition function ratio of two Markov random fields via Monte Carlo simulations is studied. Such an estimation is central to maximum likelihood based Markov random field statistical inference. Two partition function ratio estimation algorithms are considered. Both involve choosing an appropriate probability mass function. The optimal choice as well as suboptimal choices for this function are presented. The latter result in efficient algorithms that achieve significant improvements over methods reported elsewhere, as it is demonstrated in supporting simulations.

1. Introduction

Markov random fields (*MRFs*) constitute a popular class of statistical models for images [1], [2]. However, the lack of a closed form solution of the *partition function* (*PF*) and hence of the *likelihood function* [3] of a general MRF [4] imposes restrictions to *maximum likelihood* statistical inference of *fully* or *partially observed* (hidden) data, modeled via MRFs [3], [5]. To avoid this limitation, a number of *stochastic approximation* (i.e., *Monte Carlo*) techniques for PF estimation have been proposed [6]-[10]. A unified presentation, analysis, and comparative study of them can be found in [10].

As a result of the study in [10], the *Jerrum-Sinclair* (*J-S*) [6], [10] algorithm was shown to be the most preferable PF estimation method, resulting in *asymptotically unbiased* and *consistent* PF estimators. Such estimators can sufficiently well approximate the true PF in *polynomial time* with respect to the number of MRF sites, *Turing reducible* to the problem of sampling from the MRF probability mass function [10]. The J-S algorithm estimates the PF of the *original* MRF as a product of the a-priori known PF of a *reference* MRF and Monte Carlo estimated PF ratios of *intermediate* MRFs that "span" the distance between the probability mass functions of the original and the reference MRFs. Three are the important elements of the J-S PF estimation algorithm: (a) The choice of the reference MRF; (b) the choice of the intermediate MRFs; and (c) the Monte Carlo algorithm associated with PF ratio estimation. The current paper addresses issue (c), based on related work in [11] and [12]. Issues (a) and (b) have been successfully addressed in [10].

The paper is structured as follows: Section 2 establishes necessary background and notation, Section 3 gives a unified presentation of two methods for PF ratio estimation, first suggested in [11] and [12]. An appropriate probability mass function is introduced, and its optimal choice is established. Section 4 presents sub-optimal choices of this function that result in practical PF ratio estimation algorithms, the efficiency of which is demonstrated in simulation experiments, reported in Section 5. Conclusions are drawn in Section 6. To simplify notation, we exclusively consider MRFs defined over two-dimensional rectangular grids (images).

2. Background and Notation

Consider a collection of $M \times N$ sites $\Lambda = \{(i,j) : 1 \le i \le M,\ 1 \le j \le N\}$. A discrete-valued random variable H_{ij} is assigned at each site $(i,j) \in \Lambda$, taking values h_{ij} from a finite state-space E, with $|E| = R : 2 \le R < +\infty$, where $|\bullet|$ denotes cardinality of set $\bullet$. The resulting random field $\mathbf{H} = \{H_{ij} : (i,j) \in \Lambda\}$ can take any of the R^{MN} possible states $\mathbf{h} = \{h_{ij} : (i,j) \in \Lambda\} \in E^{MN}$ with probability mass function $Pr[\mathbf{H} = \mathbf{h}]$. Let us denote the *neighborhood* of a site $(i,j) \in \Lambda$, by $\mathbb{N}_{ij} \subseteq \Lambda$. Let $\mathbf{h}_{ij} = \{h_{kl} : (k,l) \in \mathbb{N}_{ij},\ k \le i \text{ and } l \le j\}$, for every $(i,j) \in \Lambda$, and $J = |\mathbf{h}_{ij}|$, independent of $(i,j) \in \Lambda$, assuming a *homogeneous* neighborhood [1], [2]. We restrict $\mathbf{H}$ to be a MRF whose probability mass function is given by the *Gibbs distribution* [1], [9]

$$\pi_\sigma(\mathbf{h}) = Pr[\mathbf{H} = \mathbf{h}] = \frac{1}{Z(\sigma)} A_\sigma(\mathbf{h}), \tag{1a}$$

for all states $\mathbf{h} \in E^{MN}$, where

$$A_\sigma(\mathbf{h}) = \exp\left\{-\frac{1}{T} U_\sigma(\mathbf{h})\right\} = \prod_{i=1}^{M} \prod_{j=1}^{N} \sigma(h_{ij}, \mathbf{h}_{ij}), \tag{1b}$$

and

$$Z(\sigma) = \sum_{\mathbf{h}} A_\sigma(\mathbf{h}). \tag{1c}$$

In (1), $Z(\sigma)$ is the *partition function* (*PF*), T is a positive parameter known as the *temperature*, $U_\sigma(\bullet)$ is the *energy function*, and $\sigma = \{\sigma(x_o, x_1, \ldots, x_J) : x_o, x_1, \ldots, x_J \in E\}$ is the *local transfer function* (*LTF*), which depends on T and is positive and finite for every $0 < T \le +\infty$ [9]. For simplicity only we assume that the LTF is homogeneous [10].

As an example of (1), consider the *Ising model* [4]. For such a MRF, $\mathbf{h}_{ij} = \{h_{i-1,j}, h_{i,j-1}\}$, $J = 2$, $E = \{-1, +1\}$, and

$$\sigma(x_o, x_1, x_2) = \exp\left\{-\frac{1}{T}(a_v x_o x_1 + a_h x_o x_2)\right\}, \tag{2}$$

for all $x_o, x_1, x_2 \in E$, where a_v and a_h are two parameters.

A useful special MRF is the *Markov mesh* [9], characterized by a probability mass function $P_\tau(\mathbf{h})$, $\mathbf{h} \in E^{MN}$, of the form

$$P_\tau(\mathbf{h}) = \prod_{i=1}^{M} \prod_{j=1}^{N} \tau(h_{ij}, \mathbf{h}_{ij}) > 0, \tag{3a}$$

where the LTF $\tau = \{\tau(x_o, x_1, \ldots, x_J) : x_o, x_1, \ldots, x_J \in E\}$ satisfies

$$\sum_{x \in E} \tau(x, x_1, \ldots, x_J) = 1, \quad \text{for every } (x_1, \ldots, x_J) \in E^J, \tag{3b}$$

is positive and finite. In this case, the PF equals [9]

$$Z(\tau) = \sum_{\mathbf{h}} \prod_{i=1}^{M} \prod_{j=1}^{N} \tau(h_{ij}, \mathbf{h}_{ij}) = 1. \tag{4}$$

In general, we are interested in estimating the PF of a general MRF (1), given its LTF σ. This can be achieved by means of the *Jerrum-Sinclair* (*J-S*) algorithm, as follows [6], [10]: First, define a reference MRF with LTF given by

$$\tau(x_o, x_1, \ldots, x_J) = \frac{\sigma(x_o, x_1, \ldots, x_J)}{\sum_x \sigma(x, x_1, \ldots, x_J)}, \tag{5}$$

for every $x_o, x_1, \dots, x_J \in E$, that clearly satisfies (3b). Then, define a MRF parametric class with LTFs given by

$$\sigma_\beta(x_o,\dots,x_J) = \left[\sigma(x_o,\dots,x_J)\right]^{\beta}\left[\tau(x_o,\dots,x_J)\right]^{1-\beta}, \qquad (6)$$

for all $x_o, x_1, \dots, x_J \in E$ and $\beta \in [0,1]$. Choose any integer $\eta \geq 1$, and denote $\sigma_{(i)}(x_o,x_1,\dots,x_J) = \sigma_{i/\eta}(x_o,x_1,\dots,x_J)$ for all x_o, $x_1,\dots,x_J \in E$ and $i = 0, 1, \dots, \eta$. Then (see also (4)),

$$Z(\sigma) = \prod_{i=0}^{\eta-1} \frac{Z(\sigma_{(i+1)})}{Z(\sigma_{(i)})}. \qquad (7)$$

Clearly, in the J-S algorithm, the problem of PF estimation reduces to that of PF ratio estimation; i.e., *given, two MRFs with LTFs* σ_o *and* σ_1 *estimate their PF ratio* $Z(\sigma_1)/Z(\sigma_o)$ *by means of an efficient Monte Carlo algorithm.* The study of such algorithms is the subject of this paper.

Traditionally, PF ratio $Z(\sigma_1)/Z(\sigma_o)$ is rewritten as [6], [10]

$$\frac{Z(\sigma_1)}{Z(\sigma_o)} = \sum_{\mathbf{h}} \left[\frac{A_{\sigma_1}(\mathbf{h})}{A_{\sigma_o}(\mathbf{h})}\right]\pi_{\sigma_o}(\mathbf{h}) = \mathbf{E}_{\pi_{\sigma_o}}\left[\frac{A_{\sigma_1}(\mathbf{H})}{A_{\sigma_o}(\mathbf{H})}\right]. \qquad (8a)$$

Therefore, an asymptotically unbiased and consistent Monte Carlo estimator of $Z(\sigma_1)/Z(\sigma_o)$ will be given by [6], [10], [13]

$$\hat{\Delta Z}_o(K) = \frac{1}{K}\sum_{k=1}^{K}\frac{A_{\sigma_1}(\mathbf{H}_k)}{A_{\sigma_o}(\mathbf{H}_k)}, \qquad (8b)$$

where $\{\mathbf{H}_k,\ k=1,2,\dots,K\}$ is a collection of MRFs obtained by means of a *Markov Chain Monte Carlo sampling* scheme (*MCMCS*) such as the *Gibbs sampler* [2] and satisfies $\lim_{k\to+\infty} Pr[\mathbf{H}_k = \mathbf{h}] = \pi_{\sigma_o}(\mathbf{h})$, for every $\mathbf{h} \in E^{MN}$. The variance of estimator (8b) can be rather large, in the case that the Gibbs distributions π_{σ_o} and π_{σ_1} significantly "differ"[1] [10]. This can seriously affect the accuracy of the J-S PF estimator (7). In this work, we study two alternative ways of estimating ratio $Z(\sigma_1)/Z(\sigma_o)$ that were originally proposed in [11] and [12]. These methods incorporate information from both π_{σ_o} and π_{σ_1} in the choice of the probability mass function, which Monte Carlo samples are drawn from, in an effort to achieve *importance sampling* [13].

3. Two Partition Function Ratio Estimators

Let $\omega : E^{MN} \to \mathbf{R}^+$ be a function such that $0 < \omega(\mathbf{h}) < +\infty$, for all $\mathbf{h} \in E^{MN}$. With an appropriate renormalization ω becomes a probability mass function on E^{MN}, therefore we assume that $\sum_{\mathbf{h}} \omega(\mathbf{h}) = 1$. Using (1) we obtain [11],

$$\frac{Z(\sigma_1)}{Z(\sigma_o)} = \frac{\sum_{\mathbf{h}}\left[\frac{A_{\sigma_1}(\mathbf{h})}{\omega(\mathbf{h})}\right]\pi_{\sigma_o}(\mathbf{h})}{\sum_{\mathbf{h}}\left[\frac{A_{\sigma_o}(\mathbf{h})}{\omega(\mathbf{h})}\right]\pi_{\sigma_1}(\mathbf{h})} = \frac{\mathbf{E}_{\pi_{\sigma_o}}\left[\frac{A_{\sigma_1}(\mathbf{H})}{\omega(\mathbf{H})}\right]}{\mathbf{E}_{\pi_{\sigma_1}}\left[\frac{A_{\sigma_o}(\mathbf{H})}{\omega(\mathbf{H})}\right]}. \qquad (9a)$$

Therefore, an asymptotically unbiased and consistent Monte Carlo estimator of $Z(\sigma_1)/Z(\sigma_o)$ will be [11], [13]

$$\hat{\Delta Z}_1(K) = \frac{\frac{1}{K}\sum_{k=1}^{K} A_{\sigma_1}(\mathbf{H}_k^{(o)}) \,/\, \omega(\mathbf{H}_k^{(o)})}{\frac{1}{K}\sum_{k=1}^{K} A_{\sigma_o}(\mathbf{H}_k^{(1)}) \,/\, \omega(\mathbf{H}_k^{(1)})}, \qquad (9b)$$

where we assumed that we use an equal number of samples in the Monte Carlo estimators of both the numerator and denominator of (9a), whereas $\{\mathbf{H}_k^{(i)},\ k=1,2,\dots,K\}$, $i = 0, 1$, are two collections of MRFs obtained by MCMCS that satisfy $\lim_{k\to+\infty} Pr[\mathbf{H}_k^{(i)} = \mathbf{h}] = \pi_{\sigma_i}(\mathbf{h})$, for every $\mathbf{h} \in E^{MN}$.

Alternatively, we can write [12]

$$\frac{Z(\sigma_1)}{Z(\sigma_o)} = \frac{\sum_{\mathbf{h}}\left[\frac{A_{\sigma_1}(\mathbf{h})}{\omega(\mathbf{h})}\right]\omega(\mathbf{h})}{\sum_{\mathbf{h}}\left[\frac{A_{\sigma_o}(\mathbf{h})}{\omega(\mathbf{h})}\right]\omega(\mathbf{h})} = \frac{\mathbf{E}_{\omega}\left[\frac{A_{\sigma_1}(\mathbf{H})}{\omega(\mathbf{H})}\right]}{\mathbf{E}_{\omega}\left[\frac{A_{\sigma_o}(\mathbf{H})}{\omega(\mathbf{H})}\right]}. \qquad (10a)$$

Therefore, an asymptotically unbiased and consistent Monte Carlo estimator of $Z(\sigma_1)/Z(\sigma_o)$ will be [12], [13]

$$\hat{\Delta Z}_2(K) = \frac{\frac{1}{K}\sum_{k=1}^{K} A_{\sigma_1}(\mathbf{H}_k^{(o)}) \,/\, \omega(\mathbf{H}_k^{(o)})}{\frac{1}{K}\sum_{k=1}^{K} A_{\sigma_o}(\mathbf{H}_k^{(1)}) \,/\, \omega(\mathbf{H}_k^{(1)})}, \qquad (10b)$$

where $\{\mathbf{H}_k^{(i)},\ k=1,2,\dots,K\}$, $i = 0, 1$, are two mutually *independent* collections of random fields drawn from $\omega(\mathbf{h})$, $\mathbf{h} \in E^{MN}$, and we assumed that an equal number of Monte Carlo samples are used in estimations in (10b).

Notice that both (9) and (10) reduce to (8), if we set $\omega(\mathbf{h}) = \pi_{\sigma_o}(\mathbf{h})$, for all $\mathbf{h} \in E^{MN}$. We would like now to determine the optimal choice for the probability mass function $\omega(\mathbf{h})$, $\mathbf{h} \in E^{MN}$, in (9) and (10). Such a choice should minimize the mean square error (*MSE*) of estimators (9b) and (10b), respectively. To simplify our derivations, we choose instead, to minimize their asymptotic mean square errors, $MSE_1(\omega)$ and $MSE_2(\omega)$, defined by

$$MSE_i(\omega) = \lim_{K\to+\infty} K\,\mathbf{E}\left[\ln \hat{\Delta Z}_i(K) - \ln\frac{Z(\sigma_1)}{Z(\sigma_o)}\right], \qquad (11)$$

for $i = 1, 2$ (see (9), (10), [5], [10], [11], and [14]). Furthermore, we assume that in (9b) and (10b), *exact* and independent samples, drawn from probability mass functions ω, π_{σ_o}, and π_{σ_1}, are used.[2] Let us denote the *divergence* of probability mass function $P_2(\mathbf{h})$ from $P_1(\mathbf{h})$ defined on E^{MN} by [14]

$$d(P_1, P_2) = \sum_{\mathbf{h}} P_1(\mathbf{h}) \ln\frac{P_1(\mathbf{h})}{P_2(\mathbf{h})}. \qquad (12)$$

In addition, let us define probability mass functions

$$\bar{\omega}(\mathbf{h}) = \frac{\pi_{\sigma_o}(\mathbf{h}) + \pi_{\sigma_1}(\mathbf{h})}{2} = \frac{A_{\sigma_1}(\mathbf{h}) + \frac{Z(\sigma_1)}{Z(\sigma_o)}A_{\sigma_o}(\mathbf{h})}{2\,Z(\sigma_1)} \qquad (13a)$$

and

$$\bar{\bar{\omega}}(\mathbf{h}) = \sqrt{\pi_{\sigma_o}^2(\mathbf{h}) + \pi_{\sigma_1}^2(\mathbf{h})} \Big/ \sum_{\mathbf{h}'} \sqrt{\pi_{\sigma_o}^2(\mathbf{h}') + \pi_{\sigma_1}^2(\mathbf{h}')}, \qquad (13b)$$

for all $\mathbf{h} \in E^{MN}$. We have the following theorem:

Theorem 1: *Probability mass function* $\bar{\omega}$, $\mathbf{h} \in E^{MN}$ *is the optimal choice to be used in (9), i.e.,*

$$\bar{\omega} = \arg\min_{\omega} MSE_1(\omega), \qquad (14a)$$

whereas, it satisfies the "divergence optimality criterion"

$$\bar{\omega} = \arg\min_{\omega}\{d(\pi_{\sigma_o}, \omega) + d(\pi_{\sigma_1}, \omega)\}. \qquad (14b)$$

In addition, probability mass function $\bar{\bar{\omega}}(\mathbf{h})$, $\mathbf{h} \in E^{MN}$, *is the optimal choice to be used in (10), i.e.,*

$$\bar{\bar{\omega}} = \arg\min_{\omega} MSE_2(\omega). \qquad (15)$$

[1] In terms of divergence (see (12) and [10]).

[2] A more general treatment is currently being studied, along the lines of the work in [6] and [10].

Furthermore,

$$MSE_2(\bar{\bar{\omega}}) \le MSE_2(\bar{\omega}) \le MSE_1(\bar{\omega})\,. \tag{16}$$

Proof: Briefly, (14a) follows from the minimization of

$$MSE_1(\omega) \;=\; \frac{\sum\limits_{\mathbf{h}} \dfrac{\pi_{\sigma_1}(\mathbf{h})\,\pi_{\sigma_o}(\mathbf{h})}{\omega(\mathbf{h})}\,\dfrac{\pi_{\sigma_1}(\mathbf{h})+\pi_{\sigma_o}(\mathbf{h})}{\omega(\mathbf{h})}}{\Big[\sum\limits_{\mathbf{h}} \dfrac{\pi_{\sigma_1}(\mathbf{h})\,\pi_{\sigma_o}(\mathbf{h})}{\omega(\mathbf{h})}\Big]^2} \;-\; 2\,, \tag{17a}$$

with respect to ω over the E^{MN}-dimensional simplex, whereas (14b) and (15) follow from the minimization of the second and third terms in

$$MSE_2(\omega)+2 = \Big[\sum_{\mathbf{h}} \frac{\pi^2_{\sigma_o}(\mathbf{h})+\pi^2_{\sigma_1}(\mathbf{h})}{\omega(\mathbf{h})}\Big] \ge d(\pi_{\sigma_o},\omega) + d(\pi_{\sigma_1},\omega)\,, \tag{17b}$$

respectively, with respect to ω over the same simplex (see (9)-(13), and the *central limit theorem* and *delta method* in [5]). In addition, it can be shown that both $\bar{\omega}$ and $\bar{\bar{\omega}}$ constitute global minima in (14) and (15). Finally, (16) follows from (15) and the fact that (see (13a) and (17))

$$\begin{aligned} MSE_2(\bar{\omega}) \;&=\; 2\sum_{\mathbf{h}} \frac{\pi^2_{\sigma_o}(\mathbf{h})+\pi^2_{\sigma_1}(\mathbf{h})}{\pi_{\sigma_o}(\mathbf{h})+\pi_{\sigma_1}(\mathbf{h})} \;-\; 2 \\ &\le\; \Big[\sum_{\mathbf{h}} \frac{\pi_{\sigma_1}(\mathbf{h})\,\pi_{\sigma_o}(\mathbf{h})}{\pi_{\sigma_o}(\mathbf{h})+\pi_{\sigma_1}(\mathbf{h})}\Big]^{-1} - \;2 \;=\; MSE_1(\bar{\omega})\,. \quad \square \end{aligned}$$

Due to (16), we consider $\bar{\omega}$ to be "optimal" for both estimators (9b) and (10b). Theorem 1 shows that, theoretically, (10b) is a better estimator than (9b). However, neither (13a) nor (13b) can be computed in general (though, a MCMCS algorithm can be easily devised to draw samples that are asymptotically distributed according to $\bar{\omega}$). In order for (9b) and (10b) to be of any practical use a good choice of ω that lies "close" to the "optimal" $\bar{\omega}$ has to be determined. This is discussed in the following section.

4. Practical Algorithms

As we mentioned above, probability mass function $\bar{\omega}(\mathbf{h})$, $\mathbf{h} \in E^{MN}$, is not computable in general, due to the presense of PF ratio $Z(\sigma_1)/Z(\sigma_o)$ in (13a). However, we can replace this PF ratio by its Monte Carlo estimate $\Delta\hat{Z}_o(K)$, and thus obtain an approximation of $\bar{\omega}$ as

$$\hat{\omega}_1(\mathbf{h}) \;=\; \frac{A_{\sigma_1}(\mathbf{h}) + \Delta\hat{Z}_o(K)A_{\sigma_o}(\mathbf{h})}{Z(\sigma_1) + \Delta\hat{Z}_o(K)\,Z(\sigma_o)}\,, \tag{18}$$

for all $\mathbf{h} \in E^{MN}$. Clearly, $\hat{\omega}_1$ can be used instead of $\bar{\omega}$ in (9), since the uncomputable denominator of (18) cancels out in (9). However, use of (18) is not possible in (10), since (10b) would require drawing samples from $\hat{\omega}_1$. Such a task requires the exact value of $Z(\sigma_1)/Z(\sigma_o)$.

Equations (8b), (9b), and (18) give rise to the following practical algorithm for PF ratio estimation:

Practical PF Ratio Estimation Algorithm 1.

- *Run a MCMCS, generating the MRF collection* $\{\mathbf{H}_k^{(o)}, k = 1,2,\ldots,K\}$, *that satisfies* $\lim\limits_{k\to+\infty} Pr[\mathbf{H}_k^{(o)} = \mathbf{h}] = \pi_{\sigma_o}(\mathbf{h})$, *for every* $\mathbf{h} \in E^{MN}$.
- *Store ratios* $B(\mathbf{H}_k^{(o)}) = A_{\sigma_1}(\mathbf{H}_k^{(o)})/A_{\sigma_o}(\mathbf{H}_k^{(o)})$, $k = 1,2,\ldots,K$.
- *Compute (8b), as* $\Delta\hat{Z}_o(K) \;=\; 1/K\sum_{k=1}^{K} B(\mathbf{H}_k^{(o)})$.
- *Run a MCMCS, generating the MRF collection* $\{\mathbf{H}_k^{(1)}, k = 1,2,\ldots,K\}$, *that satisfies* $\lim\limits_{k\to+\infty} Pr[\mathbf{H}_k^{(1)} = \mathbf{h}] = \pi_{\sigma_1}(\mathbf{h})$, *for every* $\mathbf{h} \in E^{MN}$.
- *Estimate the desired PF ratio as*

$$\Delta\hat{Z}'_1(K) = \frac{\dfrac{1}{K}\sum\limits_{k=1}^{K} \dfrac{B(\mathbf{H}_k^{(o)})}{B(\mathbf{H}_k^{(o)}) + \Delta\hat{Z}_o(K)}}{\dfrac{1}{K}\sum\limits_{k=1}^{K} \big[A_{\sigma_1}(\mathbf{H}_k^{(1)})/A_{\sigma_o}(\mathbf{H}_k^{(1)}) + \Delta\hat{Z}_o(K)\big]^{-1}}\,. \tag{19}$$

As we demonstrate in Section 5, Algorithm 1 results in more accurate PF ratio estimates than (8b), due to superior importance sampling, though at a significantly increased computational and storage requirement cost. Notice, that Algorithm 1 can be easily incorporated within the J-S PF estimation algorithm framework (5)-(7).

We would like now to propose a practical choice of ω to be used in (10). Due to Theorem 1, it is natural to seek an ω that lies "close" to the "optimal" $\bar{\omega}$, given by (13a), and, in addition, enables estimating (10). That would require ω to be restricted within a class C that consists of ω's such that $\omega(\mathbf{h})$ be computable up to a multiplicative constant, for all $\mathbf{h} \in E^{MN}$, and sampling from ω is possible. It is then natural to seek

$$\hat{\omega}_2 \;=\; \arg\min_{\omega\in C}\,\{\, d(\bar{\omega},\omega)\,\}\,, \tag{20}$$

as a "reasonable" choice of an ω to be used in (10).

We first restrict $\omega(\mathbf{h}) = P_\tau(\mathbf{h})$, i.e., we parametrize $\omega(\mathbf{h})$ in terms of the LTF τ that satisfies (3b). We then have the following theorem:

Theorem 2: *The optimal* $\omega(\mathbf{h})$ *within the class of Markov meshes is given by (3), with LTF given by* $\tau(x_o,x_1,\ldots,x_J) =$

$$\frac{\mathbf{E}_{\pi_{\sigma_o}}[\nu_{\mathbf{H}}(x_o,x_1,\ldots,x_J)] \;+\; \mathbf{E}_{\pi_{\sigma_1}}[\nu_{\mathbf{H}}(x_o,x_1,\ldots,x_J)]}{\sum\limits_{x}\mathbf{E}_{\pi_{\sigma_o}}[\nu_{\mathbf{H}}(x,x_1,\ldots,x_J)] \;+\; \sum\limits_{x}\mathbf{E}_{\pi_{\sigma_1}}[\nu_{\mathbf{H}}(x,x_1,\ldots,x_J)]}\,, \tag{21}$$

for all $x_o,x_1,\ldots,x_J \in E$, *where* $\nu_{\mathbf{H}}(x_o,x_1,\ldots,x_J)$ *denotes the number of occurrences* $(h_{ij},\mathbf{h}_{ij}) = (x_o,x_1,\ldots,x_J)$ *that appear in MRF* $\mathbf{H}$.

Proof: See (3), (13a), (20) and the proof of Theorem 3.1, in [10]. $\square$

In (21), $\mathbf{E}_{\pi_{\sigma_i}}[\nu_{\mathbf{H}}(x_o,x_1,\ldots,x_J)]$, $x_o,x_1,\ldots,x_J \in E$, $i = 0,1$, can be estimated by means of a suitable MCMCS scheme. Notice, however, that Markov mesh based PF estimation was shown in [10] to often be unreliable, since Markov meshes often provide poor approximations to random fields (here to $\bar{\omega}$). Therefore, we do not further pursue the use of (21) for PF ratio estimation.

Instead, we choose to restrict (see also (6))

$$\omega(\mathbf{h}) \;=\; \pi_{\sigma_\beta}(\mathbf{h}) \;=\; \frac{1}{Z(\sigma_\beta)}\,\big[A_{\sigma_1}(\mathbf{h})\big]^{\beta}\,\big[A_{\sigma_o}(\mathbf{h})\big]^{1-\beta}\,, \tag{22a}$$

for all $\mathbf{h} \in E^{MN}$, where

$$Z(\sigma_\beta) \;=\; \sum_{\mathbf{h}} \big[A_{\sigma_1}(\mathbf{h})\big]^{\beta}\,\big[A_{\sigma_o}(\mathbf{h})\big]^{1-\beta}\,, \tag{22b}$$

for $\beta \in [0,1]$. We now have the following theorem:

Theorem 3: *The optimal* $\hat{\omega}_2(\mathbf{h})$ *that satisfies (20) and (22), is given by (22) with* $\beta = \hat{\beta}$, *where* $\hat{\beta}$ *is the unique solution of*

$$\mathbf{E}_{\bar{\omega}}\Big[\log\frac{A_{\sigma_1}(\mathbf{H})}{A_{\sigma_o}(\mathbf{H})}\Big] \;=\; \mathbf{E}_{\pi_{\sigma_\beta}}\Big[\log\frac{A_{\sigma_1}(\mathbf{H})}{A_{\sigma_o}(\mathbf{H})}\Big]\,, \tag{23}$$

for $\beta \in [0,1]$.

Proof: See (13a), (20), and (22). In addition, notice that the

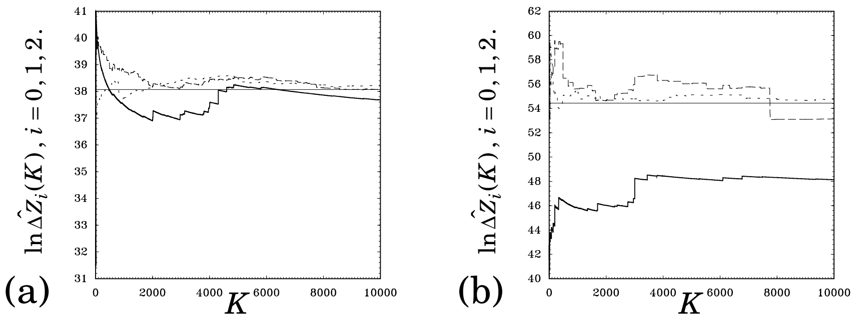

Figure: Convergence of estimators (8b) (solid thick line), (19) (dotted line), and (24) (dashed line), to the exact result (solid thin line), as a function of Monte Carlo iterations, for the estimation of the partition function ratio of two Ising models (2) with $a_v = a_h = 1.0$ considered at a 16×16 point lattice at temperatures: (a) $T = 2.0$ and $T = 2.5$, (b) $T = 2.0$ and $T = 3.0$ (vertical axis is in logarithmic scale).

right hand side of (23) is an increasing function of β. □

Clearly, (23) can be solved using Monte Carlo estimates of the left hand side and, for various values of $\beta \in [0, 1]$, of the right hand side of (23). However, such an approach would defeat the goal of computational efficiency in PF ratio estimation. We, instead, choose the ad-hoc value $\beta = 0.5$, which gives satisfactory results in our simulations. In addition, for the shake of computational efficiency, we use the same Monte Carlo samples in both numerator and denominator in (10b). The resulting PF ratio estimation algorithm is now straightforward:

Practical PF Ratio Estimation Algorithm 2.

- *Run a MCMCS, generating the MRF collection* $\{\mathbf{H}_k, k = 1, 2, \ldots, K\}$, *that satisfies* $\lim_{k \to +\infty} Pr[\mathbf{H}_k = \mathbf{h}] = \pi_{\sigma_{0.5}}(\mathbf{h})$, *for every* $\mathbf{h} \in E^{MN}$, *where* $\pi_{\sigma_{0.5}}(\mathbf{h})$ *is the MRF with LTF*
$$\sigma_{0.5}(x_o, \ldots, x_J) = \sqrt{\sigma_o(x_o, \ldots, x_J)\, \sigma_1(x_o, \ldots, x_J)},$$
for all $x_o, x_1, \ldots, x_J \in E$.
- *Estimate the desired PF ratio as*
$$\Delta \hat{Z}_2'(K) = \frac{\frac{1}{K} \sum_{k=1}^{K} \left[A_{\sigma_1}(\mathbf{H}_k) / A_{\sigma_o}(\mathbf{H}_k) \right]^{0.5}}{\frac{1}{K} \sum_{k=1}^{K} \left[A_{\sigma_o}(\mathbf{H}_k) / A_{\sigma_1}(\mathbf{H}_k) \right]^{0.5}}. \qquad (24)$$

As we demonstrate in Section 5, Algorithm 2 results in significantly more accurate PF ratio estimates than (8), with practically no additional computations. Furthermore, it gives more accurate estimation results and is computationally less intensive than Algorithm 1. Finally, it can be easily incorporated within the J-S PF estimation algorithm (5)-(7).

5. Simulation Results

We have compared the accuracy of estimators (8b), (19), and (24) in a variety of PF ratio estimation tasks. As a result of our simulation experiments, we have concluded that:

1. Both (19) and (24) provide significantly more accurate PF ratio estimates than (8), especially when probability mass functions π_{σ_o} and π_{σ_1} are "far" apart.
2. Furthermore, (24) provides more accurate PF ratio estimates than (19) with significantly less computations.
3. The use of the same Monte Carlo samples in both the numerator and denominator of (24) does not cause a significant deterioration in the PF ratio estimation accuracy, as opposed to the use of two independent Monte Carlo sample sequences (as in (10b)).
4. The accuracy of the J-S PF estimation algorithm dramatically improves by using (24), instead of (8b), when estimating the PF ratios in (7).

As an example of the relative performance of (8a), (19), and (24), we consider the PF ratio estimation problem of two Ising models (2) that are "close" to each other (Figure (a)) and of two Ising models that are "far" apart (Figure (b)). The figure clearly supports Conclusions 1 and 2.

6. Conclusions and Future Work

This paper provides a unified presentation and novel results about the statistical efficiency and optimality of two partition function ratio estimators reported in the literature. Practical algorithms for the sub-optimal implementation of the methods are described. Both algorithms result in significantly more accurate PF ratio estimation than a traditional PF ratio estimation approach, with Algorithm 2 being the most preferable. Algorithm 2 leads to a more efficient PF estimation via the J-S algorithm, therefore it is appropriate for use in Monte Carlo maximum likelihood statistical inference of MRFs [8]. Future work will provide theoretical comparisons between the MSE's of (8b), (19), and (24), attack the issue of a better choice for β in (22), and take into consideration the effect of sample correlations in (24).

References

[1] R.C. Dubes and A.K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, pp. 131-164, 1989.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 6, pp. 721-741, 1984.

[3] W. Qian and D.M. Titterington, "Estimation of parameters in hidden Markov models," *Phil. Trans. R. Soc. London, Ser. A*, vol. 337, pp. 407-428, 1991.

[4] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*. London, England: Academic Press, 1982.

[5] P.J. Bickel and K.A. Doksum, *Mathematical Statistics. Basic Ideas and Selected Topics*. Oakland, CA: Holden-Day, 1977.

[6] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithms for the Ising model," *SIAM J. Comp.*, vol. 22, pp. 1087-1116, 1993.

[7] Y. Ogata, "A Monte Carlo method for an objective Bayesian procedure," *Ann. Inst. Stat. Math.*, vol. 42, pp. 403-433, 1990.

[8] C.J. Geyer and E.A. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data," (with Discussion), *J. R. Stat. Soc., Ser. B*, vol. 54, pp. 657-699, 1992.

[9] G. Potamianos and J. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simulations," *IEEE Trans. Information Th.*, vol. 39, pp. 1322-1332, 1993.

[10] G. Potamianos, *Stochastic Simulation Algorithms for Partition Function Estimation of Markov Random Field Images*. Ph.D. Thesis (J. Goutsias, Adv.), Dept. of ECE, Johns Hopkins Univ., Baltimore, MD, 1994.

[11] C.H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *J. Comp. Physics*, vol. 22, pp. 245-268, 1976.

[12] G.M. Torrie and J.P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comp. Physics*, vol. 23, pp. 187-199, 1977.

[13] M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods. Volume I: Basics*. New York, NY: John Wiley and Sons, 1986.

[14] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*. New York City, NY: Springer-Verlag, 1985.