AN ANALYSIS OF MONTE CARLO METHODS FOR LIKELIHOOD ESTIMATION OF GIBBSIAN IMAGES

Gerasimos Potamianos and *John Goutsias*

Department of Electrical and Computer Engineering
Image Analysis and Communications Laboratory
The Johns Hopkins University
Baltimore, MD 21218, USA

ABSTRACT

We present a unified analysis of two Monte Carlo algorithms for estimating the likelihood function of Gibbs random field images. We show that such an estimation reduces to estimating the partition functions of suitably chosen Gibbs random fields. The first algorithm requires drawing samples from a mutually compatible Gibbs random field, and provides unbiased and consistent estimators of these partition functions. The second algorithm uses samples which are approximately drawn from the Gibbs distribution, and results in asymptotically unbiased and consistent estimators of the partition functions. We introduce a measure of the computational complexity of these algorithms, which enables us to compare them. We conclude that the first algorithm is superior, especially for models with strong interactions.

1. INTRODUCTION

Gibbs random fields (GRF's) constitute a popular class of parametric random field models in image processing. However, given a *full*, or *partial*, observation of a GRF, it is impossible to calculate its *likelihood function* exactly, except in very simple cases [1]. This poses serious limitations to parameter estimation and hypothesis testing tasks. Alternative approaches, that do not require explicit likelihood calculations, do exist. However, they are statistically inefficient, cannot be generalized in the presence of noise [2], or are inappropriate for hypothesis testing [3]. Therefore, very often, one has to resort to Monte Carlo estimation of the likelihood.

Many such algorithms exist, scattered in the literature [4-7]. In this paper, we analyze two of these algorithms. In Section 2, we introduce all necessary notation, and we show that the problem of likelihood estimation is reduced to that of partition function calculation. In Section 3, we review the first class of partition function Monte Carlo estimation algorithms, proposed by us in [4] and [5], and we provide a measure of their computational complexity. In Section 4, we study the properties of the estimation scheme suggested in [6], which requires using samples drawn from the Gibbs distribution. We discuss an approach for obtaining and using such samples, and we show that the resulting estimator is inappropriate for partition function calculations. Finally, in Section 5 we draw our conclusions.

2. THE GRF LIKELIHOOD FUNCTION

Consider $M \times N$ sites which form a rectangular grid $\Lambda = \{(i,j): 1 \le i \le M, 1 \le j \le N\}$. 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, which contains $R \ge 2$ distinct values. The resulting random field $\mathbf{H} = \{H_{ij} : 1 \le i \le M, 1 \le j \le N\}$ can take any one of the R^{MN} possible realizations $\mathbf{h} = \{h_{ij}: 1 \le i \le M, 1 \le j \le N\} \in E^{MN}$, with probability mass function $Pr[\mathbf{H} = \mathbf{h}]$. We restrict $\mathbf{H}$ to be a GRF, whose probability mass function is given by the *Gibbs distribution*

$$\pi(\mathbf{h}) = \frac{1}{Z} A(\mathbf{h}), \tag{1a}$$

where

$$A(\mathbf{h}) = \exp(-\frac{1}{T} U(\mathbf{h})) \quad \text{and} \quad Z = \sum_{\mathbf{h}} A(\mathbf{h}). \tag{1b}$$

In (1), Z is a normalizing constant known as the *partition function*, T is a positive parameter known as the *temperature*, $U(\bullet)$ is the *energy function*, whereas the summation in (1b) is carried over all R^{MN} states. Calculating Z is, therefore, prohibitive, even for moderate grid sizes. Without loss of generality, we assume *second-order* and *homogeneous* GRF's [4], [8]. In this case,

$$A(\mathbf{h}) = \prod_{i=1}^{M} \prod_{j=1}^{N} \sigma(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1}), \tag{2}$$

where, $\sigma(x,y,z,\omega)$, $(x,y,z,\omega) \in E^4$, is the *local transfer function* (LTF) of the GRF $\mathbf{H}$. The LTF depends on the temperature T, and is positive and finite for $0 < T \le +\infty$.

A *fully observed GRF image* is a realization $\mathbf{h}_{obs}$ of the GRF $\mathbf{H}$, drawn with $Pr[\mathbf{H} = \mathbf{h}_{obs}] = \pi(\mathbf{h}_{obs})$. Its likelihood function is given by (see (1) and (2))

$$L(\mathbf{h}_{obs}) = \frac{1}{Z} \prod_{(x,y,z,\omega) \in E^4} \sigma(x,y,z,\omega)^{\nu_{\mathbf{h}_{obs}}(x,y,z,\omega)}, \tag{3}$$

where $\nu_{\mathbf{h}}(x,y,z,\omega)$ denotes the number of squares of the form $(x,y,z,\omega) = (h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1})$ in $\mathbf{h} \in E^{MN}$.

Assume now that the GRF $\mathbf{H}$ cannot be observed, but it is transformed into an observable random field $\mathbf{Y}$, defined on the same lattice Λ, by means of a stochastic degradation process described by $Pr[\mathbf{Y}=\mathbf{y} | \mathbf{H}=\mathbf{h}]$, for all $\mathbf{y}, \mathbf{h} \in E^{MN}$. A *partially observed GRF image* is then a realization $\mathbf{y}_{obs}$ of the random field $\mathbf{Y}$, with likelihood function

$$L_{noisy}(\mathbf{y}_{obs}) = \frac{\sum_{\mathbf{h}} A(\mathbf{h}) Pr[\mathbf{Y} = \mathbf{y}_{obs} | \mathbf{H} = \mathbf{h}]}{\sum_{\mathbf{h}} A(\mathbf{h})} = \frac{Z_{noisy}(\mathbf{y}_{obs})}{Z}. \tag{4}$$

Under natural assumptions on the degradation process

0-7803-0946-4/93 $3.00

[3], [5], the pair $(\mathbf{H},\mathbf{Y})$ is a second-order homogeneous GRF, and thus (4) is the ratio of two partition functions. It is now obvious that the problem of computing the likelihood (3), or (4), reduces to that of estimating suitably chosen partition functions. Therefore, we shall discuss Monte Carlo partition function estimation algorithms for the sum (1b).

Such algorithms are stochastic, and we are interested in the size of the Monte Carlo sample $K_{\min}$, sufficient to obtain Monte Carlo estimators of the partition function Z, namely $Z(K)$, which satisfy

$$Pr\left[1-\varepsilon \le \frac{Z(K)}{Z} \le 1+\varepsilon\right] \ge 1-\xi\,, \quad \text{for every} \quad K \ge K_{\min}\,, \quad (5)$$

for a given *accuracy* $0<\varepsilon<1$ and *confidence* $0<1-\xi<1$.

In Section 3 we use the concept of a *mutually compatible Gibbs random field* (MC-GRF) [8]. This is a special case of a GRF, with probability mass function

$$P(\mathbf{h}) = \prod_{i=1}^{M}\prod_{j=1}^{N} \tau(h_{ij},h_{i-1,j},h_{i-1,j-1},h_{i,j-1})\,, \quad \text{for all } \mathbf{h} \in E^{MN}\,, \quad (6a)$$

where the LTF's $\tau(\bullet,\bullet,\bullet,\bullet)$ are positive, and satisfy

$$\sum_{u\in E} \tau(u,y,z,\omega) = 1\,, \quad \text{for every } (y,z,\omega)\in E^3. \quad (6b)$$

The partition function in this case equals to one. Furthermore, samples of this random field can be drawn exactly in $O(RMN)$ time [8]. Both properties are essential in developing the importance sampling Monte Carlo partition function estimation procedure discussed in Section 3.

3. MONTE CARLO PARTITION FUNCTION ESTIMATION USING MC-GRF MODELS

Consider a MC-GRF $\mathbf{H}$, defined on Λ, with probability mass function $P(\mathbf{h})$, given by (6). We denote this class of probability mass functions by Δ_{MC}. From (1b) observe that

$$Z = \sum_{\mathbf{h}}\left[\frac{A(\mathbf{h})}{P(\mathbf{h})}\right]P(\mathbf{h}) = \sum_{\mathbf{h}} Q(\mathbf{h})\,P(\mathbf{h}) = \mathbf{E}_P[Q(\mathbf{H})]\,, \quad (7a)$$

where $\mathbf{E}_P[\bullet]$ denotes expectation, and (see (2) and (6a))

$$Q(\mathbf{h}) = \frac{A(\mathbf{h})}{P(\mathbf{h})} = \prod_{i=1}^{M}\prod_{j=1}^{N} \frac{\sigma(h_{ij},h_{i-1,j},h_{i-1,j-1},h_{i,j-1})}{\tau(h_{ij},h_{i-1,j},h_{i-1,j-1},h_{i,j-1})}\,, \quad (7b)$$

for all states $\mathbf{h}\in E^{MN}$. In this case,

$$Z_P(K) = \frac{1}{K}\sum_{k=1}^{K} Q(\mathbf{H}_k)\,, \quad (8)$$

is an *unbiased* and *consistent Monte Carlo estimator* of the partition function Z [4], [9]. In (8), $\{\mathbf{H}_k,\ k=1,2,\ldots,K\}$ is a collection of i.i.d. MC-GRF's, statistically equivalent to $\mathbf{H}$.

The main focus of our work in [4], [5] is to choose the appropriate probability mass function $P(\mathbf{h}) \in \Delta_{MC}$. We have concentrated our effort on finding a $P^{opt}(\mathbf{h}) \in \Delta_{MC}$ such that $\mathbf{Var}[Z_{p_{opt}}(K)] \le \mathbf{Var}[Z_P(K)]$, for every $K=1,2,\ldots$, and for every $P(\mathbf{h}) \in \Delta_{MC}$, in an effort to achieve *importance sampling* [9]. This problem is equivalent to minimizing, with respect to $P(\mathbf{h})$, an *Ali-Silvey* type *"distance"* of the probability mass function $P(\mathbf{h})$ from $\pi(\mathbf{h})$, given by

$$D_{IS}(\pi,P) = \frac{1}{MN}\ln\sum_{\mathbf{h}}\left[\frac{\pi(\mathbf{h})}{P(\mathbf{h})}\right]^2 P(\mathbf{h}) = \frac{1}{MN}\ln\left[1+\frac{\mathbf{Var}_P[Q^2(\mathbf{H})]}{Z^2}\right]. \quad (9)$$

This "distance" characterizes the convergence properties of estimator (8) (see Theorem 2), and is an indication of how well the MC-GRF probability mass function $P(\mathbf{h})$ approximates the original Gibbs distribution $\pi(\mathbf{h})$. The problem under consideration is, therefore, equivalent to approximating a general GRF by a MC-GRF [8].

A "naive" approximation of $\pi(\mathbf{h})$ by $P(\mathbf{h})$ can be obtained by using the iid MC-GRF with LTF $\tau^{iid}(x,y,z,\omega) = 1/R$, for every $(x,y,z,\omega)\in E^4$, and probability mass function $P^{iid}(\mathbf{h}) = 1/R^{MN}$. However, this MC-GRF fails to provide efficient estimators of Z by means of (8) (see Lemma 1).

Two alternative approximations to $\pi(\mathbf{h})$ have been proposed in [4] and [8]. Given the LTF $\sigma(\bullet,\bullet,\bullet,\bullet)$ of the original GRF, we choose two MC-GRF's, with probability mass functions $P^*(\mathbf{h})$, $P^{**}(\mathbf{h})$, given by (6a), and LTF given by

$$\tau^*(x,y,z,\omega) = \frac{\sum_{\mathbf{h}} \nu_{\mathbf{h}}(x,y,z,\omega)\,\pi(\mathbf{h})}{\sum_{\mathbf{h}}\sum_{u\in E} \nu_{\mathbf{h}}(u,y,z,\omega)\,\pi(\mathbf{h})}\,, \quad (10)$$

and

$$\tau^{**}(x,y,z,\omega) = \frac{\sigma(x,y,z,\omega)}{\sum_{u\in E}\sigma(u,y,z,\omega)}\,, \quad (11)$$

for every $(x,y,z,\omega)\in E^4$, respectively. These choices are justified by the following Theorem:

Theorem 1: The MC-GRF's with LTF's given by (10) and (11), are *optimal* with respect to the "distances"

$$d_{IL}(\pi,P) = \sum_{\mathbf{h}} \pi(\mathbf{h})\ln\frac{\pi(\mathbf{h})}{P(\mathbf{h})}\,,$$

and

$$d_{IU}(\pi,P) = \frac{1}{Z^2}\left[\frac{1}{R^3}\sum_{(x,y,z,\omega)\in E^4}\sigma^2(x,y,z,\omega)/\tau(x,y,z,\omega)\right]^{MN} - 1\,,$$

respectively; i.e., they minimize the corresponding distance within Δ_{MC}. Furthermore, $P^*(\mathbf{h}) = P^{**}(\mathbf{h})$, for all $\mathbf{h}\in E^{MN}$, iff

$$\frac{\sum_{\mathbf{h}} \nu_{\mathbf{h}}(x,y,z,\omega)\,\pi(\mathbf{h})}{\sum_{\mathbf{h}}[\sum_{u\in E}\nu_{\mathbf{h}}(u,y,z,\omega)]\,\pi(\mathbf{h})} = \frac{\sum_{\mathbf{h}} \nu_{\mathbf{h}}(x,y,z,\omega)\,P^{**}(\mathbf{h})}{\sum_{\mathbf{h}}[\sum_{u\in E}\nu_{\mathbf{h}}(u,y,z,\omega)]\,P^{**}(\mathbf{h})}\,, \quad (12)$$

for all $(x,y,z,\omega)\in E^4$.

Notice though, that both choices (10) and (11) are "suboptimal", when compared to the "optimal" choice

$$P^{opt}(\mathbf{h}) = arg\{\min_{P\in\Delta_{MC}} D_{IS}(\pi,P)\}\,, \quad (13)$$

which cannot be determined analytically, in general. However, very often, $D_{IS}(\pi,P^*) \simeq 0$, and, therefore, $P^*(\mathbf{h})$ is an approximate solution to (13). In general, condition (12) is approximately satisfied only at "high" temperatures [8].

In practice, a simple algorithm for estimating the partition function of a GRF is presented in [4]: Given the LTF of the GRF under consideration, we calculate the LTF of the approximating MC-GRF by using (10) or (11). We then draw K independent samples $\{\mathbf{h}_k,\ k=1,2,\ldots,K\}$ from $P^*(\mathbf{h})$, or $P^{**}(\mathbf{h})$. In the first case we call the algorithm *Method 1*, whereas in the second case, *Method 2*. We then calculate $Q(\mathbf{h}_k)$, for all K samples, by using (7b), and, finally, we compute the summation in (8), for a large enough K. In case that *Method 1* is used, the *Gibbs sampler* [1] has to be employed as a preliminary step, in order to obtain an estimate of the LTF (10) [4]. In practice, this step represents a small only fraction of the total computational cost of the method, and is, therefore, ignored.

The following Theorem shows that the computational complexity of the previously discussed algorithms is exponential with respect to the number MN of sites in Λ.

Theorem 2: For any Monte Carlo estimator $Z_P(K)$ of the partition function Z, given by (8), there exists an integer

$$K_{\min} = \lceil \frac{1}{\xi \varepsilon^2} \frac{\mathbf{Var}_P[Q(\mathbf{H})]}{Z^2} \rceil ,$$

where $\lceil x \rceil$ denotes the smallest integer $\geq x$, such that $Z_P(K)$ satisfies (5), for every $K \geq K_{\min}$. Furthermore,

$$0 \leq \lim_{M,N \to +\infty} \frac{1}{MN} \ln K_{\min} = \lim_{M,N \to +\infty} D_{IS}(\pi, P) = D_P(T) < +\infty , \quad (14)$$

for every $0 < T \leq +\infty$.

In practice, Theorem 2 shows that the Monte Carlo scheme discussed in this section will work well only if $D_P(T) \approx 0$. Extensive simulation experiments in [4] have shown that although both methods 1 and 2 work well for GRF images at "high" temperatures, only *Method 1* provides reliable estimates of Z at "low" temperatures (close or below the critical temperature). This is so, because

$$0 \approx D_{p^*}(T) \leq D_{p^{**}}(T) << D_{p^{iid}}(T) ,$$

for a variety of GRF models. The following interesting result is obtained for $D_{p^{iid}}(T)$:

Lemma 1: The constant $D_{p^{iid}}(T)$ is given by

$$D_{p^{iid}}(T) = \lim_{M,N \to +\infty} D_{IS}(\pi, P^{iid}) = \ln R + \lim_{M,N \to +\infty} \frac{1}{MN} \ln \frac{Z_{T/2}}{Z_T^2} ,$$

where Z_T denotes the partition function of a GRF at temperature T, and $0 \leq D_{p^{iid}}(T) \leq \ln R$. Furthermore,

$$D_{p^{iid}}(+\infty) \triangleq \lim_{M,N \to +\infty} \lim_{T \to +\infty} D_{IS}(\pi, P^{iid}) = 0 ,$$

whereas,

$$D_{p^{iid}}(0) \triangleq \lim_{M,N \to +\infty} \lim_{T \to 0^+} D_{IS}(\pi, P^{iid}) = \ln R ,$$

provided that $\ln(\eta_{\min}) = o(MN)$, as $M, N \to +\infty$, where $\eta_{\min}$ is the number of minimum energy states in the GRF model.

Lemma 1 shows that, the estimation of Z through (8), and by using samples drawn from $P^{iid}(\mathbf{h})$, fails at "low" temperatures, becoming computationally equivalent to a brute-force calculation of (1b), as the temperature approaches zero.

4. AN ALTERNATIVE MONTE CARLO SCHEME FOR PARTITION FUNCTION ESTIMATION

In this section, we discuss the properties of the partition function Monte Carlo estimation algorithm, proposed in [6]. This algorithm is based on drawing samples directly from the Gibbs distribution under consideration, which are used to form a certain Gibbs average. This is a nontrivial procedure, since, in general, such samples can only be approximately obtained [1], [7], [10].

Assume that we need to compute the expectation

$$G = \sum_{\mathbf{h}} g(\mathbf{h}) \pi(\mathbf{h}) = \mathbf{E}_\pi[g(\mathbf{H})] ,$$

where $g(\bullet)$ is a measurable transformation, from E^{MN} to $\mathbb{R}$. Here, in order to estimate G, we run the *Gibbs sampler* [1], and we create a homogeneous, irreducible, aperiodic and reversible Markov chain $\{\mathbf{H}_1, \mathbf{H}_2, \dots\}$. Random field $\mathbf{H}_{k+1}$ is obtained from $\mathbf{H}_k$ by randomly updating MN sites in Λ at once, whereas the initial probability mass function of $\mathbf{H}_1$ is chosen to be $P^{**}(\mathbf{h})$ [8]. If we denote $Pr[\mathbf{H}_k = \mathbf{h}] = \pi_k(\mathbf{h})$, then $\lim_{k \to +\infty} \pi_k(\mathbf{h}) = \pi(\mathbf{h})$, for all $\mathbf{h} \in E^{MN}$, with a geometric rate. Indeed [10],

$$\left| \frac{\pi_{k+1}(\mathbf{h}) - \pi(\mathbf{h})}{\pi(\mathbf{h})} \right| \leq \frac{\lambda_{\max}^k}{\pi_{\min}} , \text{ for all states } \mathbf{h},\ k = 1, 2, \dots ,$$

where $\pi_{\min} = \min\{\pi(\mathbf{h}), \mathbf{h} \in E^{MN}\}$, $0 \leq \lambda_{\max} = \max\{\lambda_2, |\lambda_r|\} < 1$, λ_i is the i^{th} (in descending order) eigenvalue of the transition probability matrix of the Markov chain, and $r = R^{MN}$. Thus, for large k we can obtain samples that are approximately drawn from the Gibbs distribution. Estimators for G can then be obtained by employing the *"Single Long Run" Approach* [7], [11].

In this case, a single Markov chain of length $n + (K-1)l + 1$ is generated. The first n samples of it are discarded, whereas, the following samples are subsampled at every l^{th} state, in order to reduce the correlations between subsequent samples. A Monte Carlo estimator for G is then given by

$$G_{slr}(K) = \frac{1}{K} \sum_{k=1}^{K} g(\mathbf{H}_{(k)}) , \quad (15)$$

where $\mathbf{H}_{(k)} = \mathbf{H}_{n+(k-1)l+1}$, $k = 1, 2, \dots, K$. By the *ergodic theorem*, estimator (15) is asymptotically unbiased and consistent, independently of n and l. Indeed, the bias and M.S.E. of estimator (15) satisfy

$$\left| \mathbf{E}[G_{slr}(K)] - G \right| \leq \frac{1}{K} \frac{\delta_n}{1 - \rho_l} \mathbf{E}_\pi[|g(\mathbf{H})|] , \quad (16a)$$

and

$$\mathbf{E}[(G_{slr}(K) - G)^2] \leq \frac{1}{K} \left[\mathbf{Var}_\pi[g(\mathbf{H})] + \frac{2\rho_l}{1 - \rho_l} \frac{1}{\pi_{\min}} \mathbf{E}_\pi^2[|g(\mathbf{H})|] \right]$$

$$+ \frac{1}{K^2} \frac{\delta_n}{1 - \rho_l} \left[\mathbf{Var}_\pi[g(\mathbf{H})] + \frac{2\rho_l}{1 - \rho_l} (1 + \frac{1}{\pi_{\min}}) \mathbf{E}_\pi^2[|g(\mathbf{H})|] \right] , \quad (16b)$$

respectively, where $\rho_l = \lambda_{\max}^l$, and $\delta_n = \lambda_{\max}^n / \pi_{\min}$. Furthermore,

$$\lim_{K \to +\infty} K\, \mathbf{E}[(G_{slr}(K) - G)^2] = \mathbf{Var}_\pi[g(\mathbf{H})] + \upsilon , \quad (17a)$$

where

$$|\upsilon| \leq \frac{1}{\pi_{\min}} \frac{2\rho_l}{1 - \rho_l} \mathbf{E}_\pi^2[|g(\mathbf{H})|] . \quad (17b)$$

Quantity υ reflects the effect of correlations, between random fields $\mathbf{H}_{(k)}$, $k = 1, 2, \dots, K$, on the rate of convergence of the M.S.E. to zero. From (16b) we see that for large K the effect of the bias is negligible, thus, we can assume burn-in $n = 0$, and ignore the second term of the bound (16b). Let L denote the length of the chain, i.e., $K = L/l$. Then, substituting K into (16b), and minimizing the M.S.E. with respect to l, we obtain the "optimal" subsampling spacing

$$l_{opt} = \arg\{ \min_{l \geq 1} \{ l (\frac{\mathbf{Var}_\pi[g(\mathbf{H})]}{\mathbf{E}_\pi^2[|g(\mathbf{H})|]} + \frac{2\rho_l}{1 - \rho_l} \frac{1}{\pi_{\min}}) \} \} = 1 .$$

These results agree with [7], and the discussion in [11].

The partition function estimation method that we would like to discuss is based on the identity (see also (1a))

$$\frac{R^{MN}}{Z} = \sum_{\mathbf{h}} A^{-1}(\mathbf{h})\pi(\mathbf{h}) = \mathbf{E}_\pi[A^{-1}(\mathbf{h})],$$

provided that $A(\mathbf{h}) > 0$, for all states $\mathbf{h}$. As a direct result of the previous discussion,

$$Z_{OT_1}(K) = \frac{R^{MN}}{\frac{1}{K}\sum_{k=1}^{K} A^{-1}(\mathbf{H}_{(k)})} = \frac{R^{MN}}{A_{inv}(K)}, \quad (18)$$

is a Monte Carlo estimator of Z. Estimator $A_{inv}(K)$ is *asymptotically normal* (as $K \to +\infty$) [10], and as a result of the *delta method* [10], we have

$$\lim_{K\to+\infty} \mathbf{E}[Z_{OT_1}(K)] = Z, \quad (19a)$$

whereas,

$$\lim_{K\to+\infty} K\,\mathbf{Var}[Z_{OT_1}(K)] = \frac{Z^4}{R^{2MN}}\left[\mathbf{Var}_\pi[A^{-1}(\mathbf{H})] + \upsilon\right], \quad (19b)$$

where υ is an asymptotic term satisfying (17), with $g(\bullet) \to A^{-1}(\bullet)$. Estimator $Z_{OT_1}(K)$ is, therefore, an asymptotically unbiased and consistent estimator of Z. However, we have the following result regarding its variance:

Theorem 3: At any temperature T, such that

$$T \le \frac{1}{(t+3)\ln R}\,\frac{U_{\max} - U_{\min}}{MN}, \quad t \ge 0,$$

we have that

$$\frac{\lim_{K\to+\infty} K\,\mathbf{Var}[Z_{OT_1}(K)]}{K\,\mathbf{Var}[Z_{piid}(K)]} \ge R^{tMN},$$

where $U_{\max} = \max\{U(\mathbf{h}), \mathbf{h} \in E^{MN}\}$, and $U_{\min} = \min\{U(\mathbf{h}), \mathbf{h} \in E^{MN}\}$, provided that $\upsilon \ge 0$ in (17).

Theorem 3 shows that estimator $Z_{OT_1}(K)$ is less efficient than $Z_{piid}(K)$, at low temperatures. By Lemma 1 we conclude that $Z_{OT_1}(K)$ is a non-efficient estimator for Z.

In analogy with Section 3, we examine the asymptotic (as $M,N \to +\infty$) computational complexity of the partition function estimation scheme (18). We assume the existence of an unbiased and consistent estimator $Z_{0,OT_1}(K)$ of Z that satisfies (19) with $\upsilon = 0$. If $\upsilon \ge 0$, the convergence properties of $Z_{OT_1}(K)$ will be worse than the ones of $Z_{0,OT_1}(K)$; therefore, studying $Z_{0,OT_1}(K)$ will be the "best case" scenario. We apply Theorem 2 with $P(\bullet) \to \pi(\bullet)$ and $Q(\bullet) \to (Z^2/R^{MN})A^{-1}(\bullet)$, and we define $D_{OT_1}(M,N,T)$ by (compare with (9))

$$D_{OT_1}(M,N,T) = \frac{1}{MN}\ln[1 + \frac{K}{Z^2}\mathbf{Var}[Z_{0,OT_1}(K)]].$$

Then (see (14)), we obtain

$$\lim_{M,N\to+\infty} \frac{1}{MN}\ln K_{\min} = \lim_{M,N\to+\infty} D_{OT_1}(M,N,T) = D_{OT_1}(T).$$

We now have the following lemma (compare to Lemma 1):

Lemma 2: The constant $D_{OT_1}(T)$ is given by

$$0 \le D_{OT_1}(T) = -2\ln R + \lim_{M,N\to\infty} \frac{1}{MN}\ln[Z(\sigma)Z(\sigma^{-1})],$$

where $Z(\sigma)$ denotes the partition function of a GRF with LTF σ. Furthermore, $D_{OT_1}(+\infty) = 0$, whereas, $D_{OT_1}(0) = +\infty$, provided that $U_{\max} > U_{\min}$, for all sufficiently large M, N.

According to the previous lemma, the method becomes computationally worse that a brute force summation of (1b), as the temperature approaches zero; therefore, it is inappropriate for partition function calculations.

5. CONCLUSIONS

Two different estimators, for the Monte Carlo estimation of a GRF likelihood function, have been studied here. We have determined a measure which characterizes the convergence properties of these Monte Carlo estimators, and we have showed that the second estimator is inappropriate for likelihood computations. By using a suitable MC-GRF model, we have obtained an efficient estimator which is robust at all temperatures.

ACKNOWLEDGMENTS

This work was supported by the office of Naval Research, Mathematical Sciences Division, under ONR Grant N00014-90-J-1345.

REFERENCES

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

[2] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, pp. 179-195, 1975.

[3] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," *Prob. Theory Rel. Fields,* vol. 82, pp. 625-645, 1989.

[4] G. Potamianos and J. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simulations," to Appear, *IEEE Trans. IT*, 1993.

[5] G. Potamianos and J. Goutsias, "On computing the likelihood function of partially observed Markov random field images using Monte Carlo simulations," *Proc. 26th CISS*, pp. 357-362, Princeton, NJ, 1992.

[6] Y. Ogata and M. Tanemura, "Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure," *Ann. Inst. Stat. Math.*, vol. 33, Part B, pp. 315-338, 1981.

[7] C.J. Geyer, "Markov chain Monte Carlo maximum likelihood." In *Computer Science and Statistics: Proc. 23rd Symp. Interface*, pp. 156-163, 1991.

[8] J. Goutsias, "Unilateral approximation of Gibbs random field images," *CVGIP: Graph. Models Image Proc.*, vol. 53, pp. 240-257, 1991.

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

[10] P. Billingsley, *Probability and Measure*, Second Edition. New York City, NY: John Wiley and Sons, 1985.

[11] J. Besag, J. York, and A. Mollié, "Bayesian image restoration, with two applications in spatial statistics," *Ann. Inst. Statist. Math.*, vol. 43, pp. 1-59, 1991.